Federal Student Aid
An OFFICE of the U.S. DEPARTMENT of EDUCATION

PROUD SPONSOR of the AMERICAN MIND™

Data Standardization Policies and Procedures

Version 3.4 ● 10/29/2021

Document Version Control

VERSION	DATE	AUTHOR	DESCRIPTION
Draft	03/21/2007		Baseline Draft Document – B. Scott-White
02	03/29/2007		Revision of outline based on feedback from Holly and Lisa – B. Scott
03	04/05/2007		Incorporated feedback from Lisa and Steve – B. Scott
04	04/11/2007		Incorporated feedback from Holly and Lisa on first rough draft – B. Scott
05	04/25/2007		Incorporated feedback from CCSi peer review – B. Scott
06	04/27/2007		Incorporated changes from Holly: moved S&G section for data modeling to data model registration document; revised process flow – B. Scott
07	05/03/2007		Incorporated minor revisions requested by Holly to process flow; updated chart – B. Scott
08	05/07/2007		Updated Document by technical writer – S. Palmer
09	05/24/2007		Incorporated Lisa's feedback – U. Nasshan
10	05/29/2007		Did additional check – S. Palmer
1.0	05/30/2007		Final draft – U. Nasshan
1.1	09/22/2008		Minor revisions due to lessons learned during FY08 -- D. Franklin
2.0	06/20/2011		Reformatted Document to FSA Standards – Sue Lou
3.0	11/11/2011		Implemented changes resulting from EDS review
3.1	11/16/2011		Removed XML from the workflow and, updated internal workflow
3.2	2/27/2018	EITASME	Converted document into new FSA Master Template format, removed remaining XML references, and improved TOC
3.3	5/23/2017	EITASME	Made document 508 compliant, updated content based on Data Standards Assessment results
3.4	10/29/2021	Rick Armwood	Minor updates to document content. Replaced ER Studio with Erwin as the Enterprise data modeling tool

Table of Contents

List of Figures

List of Tables

Section 1. Overview

1.1. Introduction

Data is a vital resource for Federal Student Aid (FSA), and like any other resource, data must be managed effectively to deliver its full benefit. Data standardization is the process of reviewing and documenting the names, meaning, and characteristics of data elements so that all users of the data have a common, shared understanding of it. Whereas, a data element is the basic unit of identifiable information and definable information. A data element occupies the space provided by fields in a record or a block on a form. It has an identifying name and value or values that express a specific fact.

This document describes data element standardization procedures that when followed, benefit Federal Student Aid.

1.2. Purpose

This document describes policies and procedures for data standardization in Federal Student Aid.

1.2.1. Background

Federal Student Aid supports the following program-wide business objectives:

- Reduce redundant data storage.
- Improve customer service.
- Increase accuracy of analytics.
- Increase efficiency in data handling.
- Reduce costs.

1.3. Scope

The scope of this document is to provide overviews of Data Standardization Procedures and of Conceptual, Logical and Physical data models. The standards described in this document apply to the data managed by Federal Student Aid in conjunction with its business partners.

1.4. Intended Audience

This document is intended for business managers, implementers, and system developers. The business manager audience includes purchasing officers (POs), sponsors, champions of business cases, the Chief Information Officer, subject matter experts, and in-house consultants. The implementer audience includes business project managers, project managers, business case preparers, and IT project staff.

1.5. Organization of the Document

This document is divided into the following sections:

Section 1: Overview — describes the document's high-level purpose, scope, intended audience, objectives, and benefits.

Section 2: Data Standardization Process — describes the process and procedures of data standardization to support FSA.

Section 3: Data Standardization Standards and Guidelines — gives an overview of standards and guidelines for data element, data models, entities and attributes.

Appendices — includes list of Abbreviations and Acronyms and a Glossary.

1.6. Reference Documents

The following sources contributed to the content included in this document:

DOCUMENT TITLE	DOCUMENT VERSION
Department of Education Data Standards and Guidelines	
Briefing on the Assessment of Documentation Standards	
Data Warehousing > Concepts > Data Modeling - Conceptual, Logical, And Physical Data Models	
LearnDataModeling.com	

Table 1-1: Reference Documents

1.7. Standardization Goals and Objectives

The main goals and objectives of Federal Student Aid data standardization are to:

- Implement data administration in ways that provide clear, consistent, unambiguous, and easily accessible data throughout FSA.
- Standardize and register data elements to meet the requirements for data sharing and interoperability among information systems throughout FSA.
- Promote standardization in FSA, consistent with requirements for sharing data among business owners, with other Federal agencies, and with the educational community at large.
- Minimize or eliminate the cost and time required to transform, translate, or research redundant data elements across various systems.
- Incorporate applicable Federal standards before creating FSA standards or using common commercial best practices.
- Provide steps for the development and/or modification of newly-identified data and, provide towards the required level for correctness, consistency, and completeness of these data elements as represented in their respective data models.
- Provide workflows and instructions for the implementation of data standards.
- Reuse of data found in FSA data dictionaries and data models.

1.8. Standardization Benefits

The full benefits of data standardization will only be achieved if organizations use the same data element definitions and if those definitions are available for all business partners to search, retrieve, and use for file transfer specification development. At Federal Student Aid, data standardization is supported through the **Enterprise Data Dictionary (EDDI)**, maintained in Business Glossary and Erwin (Federal Student Aid's data modeling tool). These tools support seamless integration of data between Federal Student Aid systems.

These tools provide benefits and support such as, but not limited to:

- **Data Sharing:** Facilitates data sharing within and between Federal Student Aid business owners and its partners where allowed, and where mutually agreed upon.
- **Reliable Data:** Enterprise Data Standardization will allow Federal Student Aid personnel to be more productive and self-sufficient in their jobs by knowing where information is located, what it looks like, and its meaning. This will facilitate the reduction in occurrences of multiple systems capturing the same data, and displaying different answers to the same question (multiple versions of the truth).
- **Commonly Defined Data:** Definitions from individual systems will be replaced with Enterprise-wide data definitions. Application interoperability depends on standardization, and standardization of data is only possible when data is defined and understood.
- **Integration of Operations:** The support of integrated operations among lines of business, communities of practice, and the facilitation of decision-making using standard data.

1.9. Standardization Policies and Procedures

This section gives an overview of procedures for establishing and adopting data standardization and exchange standards at Federal Student Aid. By using a structured process (outlined in Federal Student Aid Data Model Standards and Guidelines, Registration Policies and Procedures and Manual), Federal Student Aid gains the authority to set and maintain standards.

Section 2. Data Standardization Process

The procedures detailed in this document will facilitate the development and standardization of new and updated data elements though use of the Federal Student Aid's data element standardization approval process.

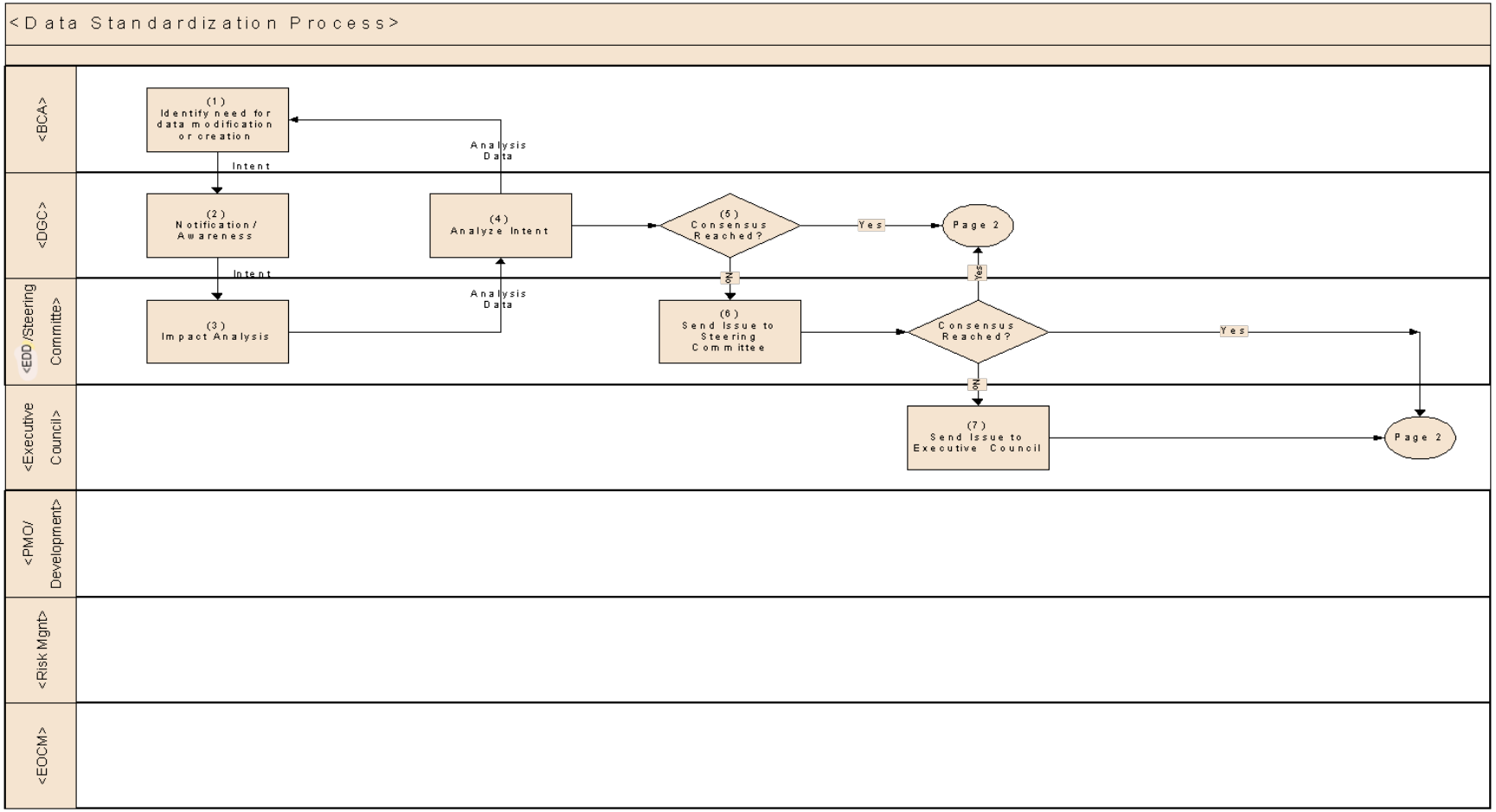

Figure 2-1: Data Standardization Process- Page 1.

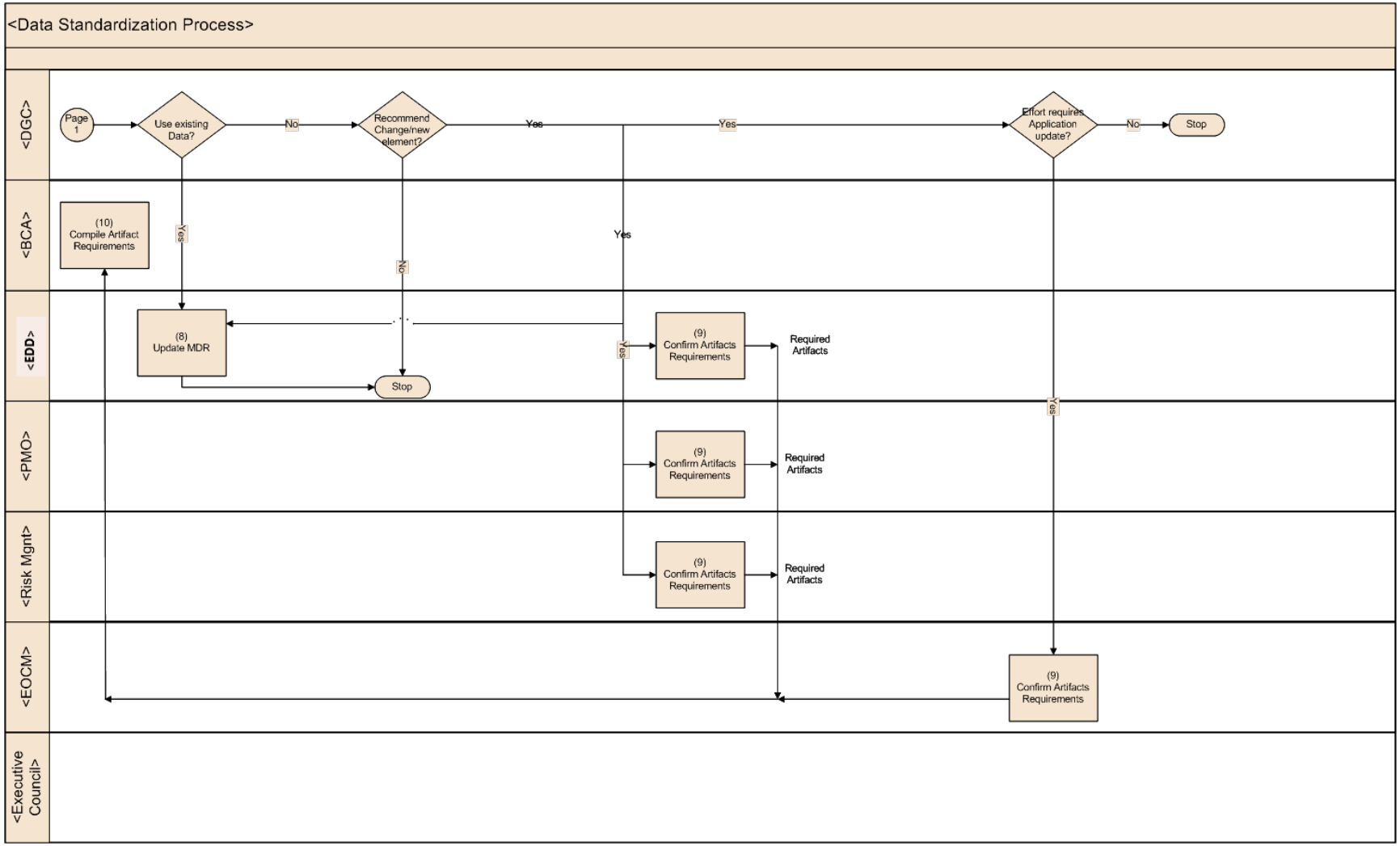

Figure 2-2: Data Standardization Process- Page 2.

2.1. Federal Student Aid Data Element Standardization Procedures

2.1.1. Step 1: Identify need for Data Modification or Creation

The Business Capability Area (BCA) determines it has a need to create or modify an existing data element or elements. The BCA's data steward brings to the next Data Governance Committee meeting, details of the data action intent.

2.1.2. Step 2: Notification/Awareness

In this step, the BCA members (of the Data Governance Committee (DGC)) are made aware of the intent to create or modify existing data elements. Preliminary discussions begin with regard to the nature of the change, if the requirement can be fulfilled without a change and any FSA standards and concerns associated with the type of change.

Wherever possible, workgroups should re-use existing data elements without modification in order to minimize the number of data elements used across applications.

2.1.3. Step 3: Impact Analysis

In this step, the Enterprise Data Directorate (EDD) Metadata manager performs an initial impact analysis in effort to:

- Identify/confirm that the change requirement can or, cannot be satisfied with existing data and,
- Identify all the potential BCAs that would be affected should the change be recommended.

The full benefit of the data standardization is achieved, when the same data element definitions are available to, and used by Federal Student Aid, organizations, and all partners in the education community. The re-use of approved data elements can save considerable time, effort and associated costs.

2.1.4. Step 4: Analyze Intent

In this step, the DGC uses the results of the EDD impact analysis to:

- Review existing data elements proposed to satisfy the change intent requirement and,
- Confirm the list of affected BCAs.

The Data Stewards bring the analysis information to their respective BCAs for further due diligences such as, but not limited to:

- In this step, the DGC uses the results of the EDD impact analysis to:
- Review existing data elements proposed to satisfy the change intent requirement and,
- Confirm the list of affected BCAs.

The Data Stewards bring the analysis information to their respective BCAs for further due diligences such as, but not limited to:

- Cost of change
- Preliminary Level Of Effort (LOE) and
- BCA's official position on the change

2.1.5. Step 5: Consensus Reached?

In this step, the DGC members vote on whether to utilize existing data, recommend the change go forward or suggest it not be performed. The decision is documented and forwarded to the EDD team:

- Should a consensus not be reached, the analysis results and positions are forwarded to the Steering Committee (EDD) for further review and recommendation.

- Should a decision be reached, notification is sent to EDD, Project Management Office (PMO) /Development and Risk management.

2.1.6. Step 6: Send Issue to Steering Committee

When the DGC cannot reach a consensus on the data issue, the Steering Committee (EDD) will be apprised of the issue and, requested to work with the DGC in effort to marshal a recommendation. Should the Steering Committee's efforts be unsuccessful, the Executive Council is requested to make a decision.

2.1.7. Step 7: Send Issue to Executive Council

When the Steering Committee cannot reach a consensus on the data issue, the Executive Council will be requested to make a decision.

2.1.8. Step 8: Update MDR

In this step, the Business Glossary and Erwin repositories are updated to recognize the new or modified data elements and, the systems consuming said data elements.

2.1.9. Step 9: Confirm Artifact Requirements

In this step, if the decision was reached to recommend the addition or modification of an element, the EDD, PMO/Development and Risk Management teams are notified. The EOCM team is also notified if the change or addition will impact an existing application. These teams identify the specific artifacts required to precede (or accompany) the effort (as detailed in the Lifecycle Management methodology) and submit the requirement(s) to the requesting BCA's Data Steward.

2.1.10. Step 10: Compile Artifact Requirements

The BCA's Data Steward compiles a list of the artifact requirements as dictated by the EOCM, EDD, PMO/Development and Risk Management SMEs. These requirements are brought back to the BCA for further consideration.

Section 3. Data Standardization Standards and Guidelines

Data standards are rules that govern the development and modification of the names, definitions, and other metadata for entities, tables, attributes, columns, and the data models in which they appear. Additionally, data standards also establish a required level for the correctness, consistency, and completeness of these data elements and data models.

Data modeling standards specify the level of detail required for particular types of data models. Logical data models are more detailed and complete than conceptual data models, and have more applicable data standards and guidelines. The goal is to encourage the development of one design for a single enterprise repository residing on a shared platform.

Reference information about all standards and guidelines for data modeling can be found in the document titled "**Data Model Standards and Guidelines, Registration Policies and Procedures**."

3.1. Conceptual, Logical, and Physical Data Model Overview

There are three levels of data models: conceptual, logical, and physical. This section explains the differences between them, the order in which they are created, and how to transition from one level to another.

3.2. Conceptual Data Model (CDM)

The conceptual data model includes all major entities and relationships. It does not contain detailed information about attributes, and is often used in the initial planning phase.

Business requirements from various sources such as business documents, discussions with business area stakeholders, functional teams, business analysts, subject matter experts, and end users who do the reporting on the database, provide input to the CDM development. Data modelers create a CDM and give that model to the functional team and business area stakeholders for review.

3.2.1. Conceptual Data Model - Highlights

The CDM is the first step in constructing a data model in a top-down approach. It is a clear and accurate visual representation of the business of an organization. The CDM shows the overall structures and provides high-level information about the business subject areas or data structures of an organization. The CDM discussion starts with a main subject area of an organization and then all the major entities of each subject area are reviewed in detail.

In the CDM the relationships between the subject areas and between entities in subject areas are drawn using symbolic notation (IDEF1X or IE). In a data model, entities can have different types of relations with each other: one-to-one, one-to-many, many-to-one, or many-to-many. The CDM contains data structures that ultimately, may not be implemented in the database.

In CDM discussions, technical as well as non-technical teams project their ideas for building a sound logical data model. Conceptual data modeling helps the functional and technical teams to understand how business requirements would be described in the logical data model (LDM).

3.2.2. Conceptual Data Model Summary

Features of the conceptual data model:

- The important entities and the relationships between them, grouped by subject area.
- No attribute is specified.
- No primary key information is specified.
- Many-to-Many relationships allowed.
- At this level, the data modeler attempts to identify the highest-level relationships among the different entities.

3.3. Logical Data Model (LDM)

After acceptance of the conceptual data model, the functional team develops the LDM. The logical data model implements and extends the conceptual data model by fully-attributing all entities. This data model version wholly, or partially, represents the business requirements of an organization, and is developed before the physical data model. The LDM does not consider any technical requirements or restrictions introduced by the physical implementation.

The steps for designing the LDM are:

- Identify all entities.
- Specify primary keys for all entities.
- Find the relationships between different entities.
- Find all attributes for each entity.
- Find the properties and valid domain ranges for each attribute.
- Resolve many-to-many relationships.
- Normalize.

Once the LDM is completed, it is forwarded to the business area stakeholders and functional teams for review. A sound logical design clearly defines data structures and the relationships between them. When creating the LDM consider current and future business requirements. This thinking will allow LDM to support a growing business without a major redesign. The LDM includes all required entities, attributes, key groups, and relationships that represent business information and define business rules.

3.3.1. Logical Data Model – Summary

The LDM:

- Includes all entities and relationships between them.
- Specifies all attributes for each entity, including properties and domain ranges.
- Specifies the primary key for each entity.
- Specifies foreign keys (keys identifying the relationship between different entities).
- Normalizes data to third normal form.

At this level, the data model describes the data in as much detail as possible, without regard to how it will be physically implemented in the database.

3.4. Physical Data Model (PDM)

Once the functional team approves the logical data model, work starts on development of the PDM. The transformation from a logical to a physical model includes imposing database rules, implementation of referential integrity, super types, and sub types.

Therefore, the PDM includes all required tables, columns, relationships, and database properties for the physical implementation of databases. Database performance, indexing strategy, data access path, physical storage, and denormalization are important parameters of a PDM. Once completed, it is then forwarded to technical teams (developer, group lead, DBA) for review and validation.

The steps for PDM design are:

- Convert entities into tables.
- Convert relationships into foreign keys.
- Convert attributes into columns.
- Add objects to the schema that did not exist at the LDM level such as, but not limited to: Indexing for performance and, Audit Log Columns.

3.4.1. Physical Data Model - Summary

The physical data model:

- Specifies all tables and columns.
- Uses foreign keys to identify relationships between tables.
- May de-normalize data depending on user access and reporting requirements.
- May be quite different from the logical data model.

The PDM specifies how the LDM will be realized in a specific physical database management system (i.e., DB2, Oracle, SQLServer, etc.).

Appendix A - Acronyms and Abbreviations

The following abbreviations and acronyms are used in this document or are pertinent to its content:

ACRONYM	DEFINITION
BCA	Business Capability Area
CCSi	Creative Computing Solutions, Inc.
CDM	Conceptual Data Model
DGC	Data Governance Committee
ECDM	Enterprise Conceptual Data Model
ED	Department of Education
EDDI	Enterprise Data Dictionary
EDD	Enterprise Data Directorate
EDM	Enterprise Data Management
EDS	Enterprise Data Services
EDSG	Enterprise Data Standards and Guidelines
ELDM	Enterprise Logical Data Model
EOCM	Enterprise Operational Change Management
FEA	Federal Enterprise Architecture
IDEF1X	Integration DEFinition for Information Modeling
IE	Information Engineering
IT	Information Technology
ITSS	Information Technology System Services
LDM	Logical Data Model
LOE	Level Of Effort
MDR	Meta Data Repository
PDM	Physical Data Model
PESC	Postsecondary Electronic Standards Council
PMO	Project Management Office

Table A-1: Acronyms and Abbreviations

Appendix B - Glossary

The following terms are used in this document or are pertinent to its content:

TERM	DEFINITION
Column	A set of data values of the same type collected and stored in the rows of a table.
Database	A set of table spaces and index spaces.
Data Element	A generic term for an entity/class, table, attribute, or column in a conceptual, logical, and/or physical data model.
Enterprise Conceptual Data Model (ECDM)	One of the initial components of a maturing Enterprise Data Architecture and the first enterprise level data model developed. The ECDM identifies groupings of data important to lines of business, conceptual entities, and defines their general relationships. The ECDM provides a picture of the data the enterprise needs to conduct its business.
Enterprise Data Dictionary (EDDI)	One of the initial components of a maturing Enterprise Data Architecture. The EDDI lists metadata objects and a complete description of the object at a sufficient level of detail to ensure that they are discrete and clearly understood. Such descriptions shall include, at a minimum, labels (names, titles, etc.) and definitions (or text descriptions), but may include additional descriptive metadata such as object type, classifications, content data type, rules (business, validation, etc.), valid and default values, etc. The EDDI is the definitive source for the meaning of metadata objects.
Enterprise Logical Data Model (ELDM)	A component of a maturing Enterprise Data Architecture. The second enterprise level data model developed. It is the result of merging application level data model information into the existing Enterprise Conceptual Data Model (ECDM). The ELDM extends the ECDM level of detail.
Enterprise Data Standards and Guidelines (EDSG)	A component of a maturing Enterprise Data Architecture. Rules and recommendations for the creation and updating of metadata objects and structures, as well as for creating conceptual and physical models and schemas at both the enterprise and application level.
Schema (Data)	Any diagram or textual description of a structure for representing data.
Table	A set of related columns and rows in a relational database.
Table Space	A portion of a database reserved a table. Table structure is the mapping of tables into spaces.
World Wide Web Consortium (W3C)	An international industry consortium founded in 1994 to develop standards for the Web. The W3C standardized many of the fundamental technologies of the Web, including HTML, XML, and others.

Table B-1: Glossary