Enterprise Data Governance (EDG) Initiative Annual Report

As Required by Rider 164

Article II, HHSC, 2019,

General Appropriations Act

December 2019

Table of Contents

The Health and Human Services Commission (HHSC) submits this report to the Legislative Budget Board and the Office of the Governor, State of Texas, as stipulated in Rider 164 of the 2020-21 General Appropriations Act (Article II, HHSC, 86th Legislature) regarding the development of an enterprise data governance program for health information:

"**Enterprise Data Governance.** *Included in the amounts appropriated above in Strategy L.1.2, IT Oversight & Program Support, is $5,460,700 in All Funds ($792,175 in General Revenue) in fiscal year 2020 and $5,458,275 in All Funds ($791,569 in General Revenue) in fiscal year 2021 for the Enterprise Data Governance project.*

No later than December 1 of each year, the Health and Human Services Commission (HHSC) shall submit a report to the Legislative Budget Board and the Governor's Office reflecting actual expenditures, cost savings, and accomplishments implementing the Enterprise Data Governance project. The report shall include a detailed plan for the project, a proposed schedule of expenditures, and the status of implementation for a comprehensive Medicaid focused Enterprise Master Data Management system, Metadata Repository, and Information Management Program.

Additionally, the report shall include detailed strategies developed and implemented by HHSC to restrict the Enterprise Data Governance project to those items presented and approved by the Eighty-sixth Legislature, 2019."

2. **Plan and Strategy**

2.1. Detailed Plan

The Enterprise Data Governance (EDG) project is operating on an approved implementation advance planning (IAPD) document through March 2022, with annual gateway approvals from the Centers for Medicare and Medicaid Services (CMS). The EDG team identified, and CMS approved, five project tracks to implement the overall Medicaid-focused EDG solution. Each track is composed of a set of phases. All five tracks will continue beyond the 2021-2022 biennium.

Below are the project plans and approach details for each of the EDG project tracks, which are:

- Data and information management track (DIM),
- Data quality and standards track (DQS),
- Metadata and reference data management track (MRDM),
- Data architecture track, and
- Data and information controls track (DIC)

Data and Information Management Track

The purpose of the DIM track is to implement an enterprise-wide master data management (MDM) system.

Key functions of the MDM system are:

- **Data acquisition:** This capability acquires data from sources that contain similar data sets.
- **Data standardization:** This capability ensures that master data records meet the data standards adopted by the EDG Council.
- **Data mastering and data governance:** This includes data matching, merging, and retaining relevant information from various systems to form a single version of a matched record. Identity resolution requires data governance, which guides data alignment and the implementation of data-management policies.
- **Data publishing:** This capability allows for the integration of golden records (single, well-defined versions of data elements) throughout the enterprise.

Overall Solution

Implement a Medicaid-focused MDM system to manage all data for internal use like the HHS insights platform, which creates and manages mastered domain data.

Use the MDM layer as a hub to measure, manage, and increase data quality in critical HHS data systems.

Data Quality and Standards Track

Measuring data quality is essential to making improvements. The DQS track ensures that data quality can be measured and makes recommendations to improve quality through the application of data standards. Continuous monitoring of data quality within mastered data domains is an integral part of the EDG program.

Overall Solution

EDG's data quality process includes the following key components:

- Data profiling,
- Data remediation, and
- Data monitoring

The key components of the data quality process will be repeated when profiling and remediating quality issues.

DQS will include profiling data sets to identify anomalies and issues. The DQS track is divided into four phases. Systems that were prioritized on the Medicaid Information Technology and Architecture (MITA) roadmap will be added to each DQS phase, and each phase will deliver key capabilities for these systems.

Metadata and Reference Data Management Track

Metadata refers to data that provides information about data elements. For context, systems typically use reference data (such as county codes, region codes, and International Classification of Diseases codes).

Metadata and reference data are generally inconsistent and do not always meet HHS data management standards. This lack of consistent data standards often becomes evident during data integration or data consolidation efforts and can result in poor data quality and poor system alignment.

The MRDM track alleviates the issues that can come from different standards, definitions, and reference codes by:

- Collecting data from different systems;
- Storing that data in a centralized repository (i.e. a metadata repository and reference data management repository); and
- Governing the creation, use, and maintenance of the data.

Metadata management activities fall into three broad categories:

Collect: The collection of metadata/reference data from transactional systems includes the capture of changes to existing metadata/reference data. During the implementation phase of the project, HHS intends to collect relevant business, technical, and reference data.

Govern: During governance activities, data stewards analyze collected metadata/reference data to identify discrepancies and recommend standards. The EDG Council determines the priority of metadata/reference data domains of Medicaid systems and Medicaid projects on the MITA roadmap.

Store/Publish: These activities allow the storage and publication of metadata/reference data sets. Storage creates a "hub" for metadata/reference data domains. This project will help ensure the EDG Council can govern reference data using a single interface.

Overall Solution

Implement a data asset repository (DAR) to collect comprehensive data-asset information across the enterprise (as mandated by Sunset Advisory Commission report, July 2015, Issue 7). A data asset is anything that contains valuable records, which can be a system, database, or simply a document. Large systems usually contain multiple types of data assets, which are migrated as a set when systems are modernized or replaced. Ensure access to the DAR for data subject matter experts. Ensure data assets, technical, and business metadata are widely available and searchable. Implement a technical metadata repository to allow access to data models and provide search capabilities.

Data Architecture Track

Data architecture is a key EDG competency. The primary goal of data architecture is to ensure key Medicaid data domains are identified, defined, and managed appropriately. Additionally, this track is responsible for creating and maintaining a logical data architecture that aligns with the business architecture referenced in the MITA framework. This track aligns closely with MITA to identify improvements in data architecture that can improve business processes.

Overall Solution

Develop and implement data governance processes to increase the maturity of enterprise data management. Integrate reference data and data quality metrics into MDM tools. Implement reference data standards that data stewards, who must explain the utility of data beyond a single program area, can use. Data analysts and business interface designers can use reference data mapped to the HHS system standard to standardize analytic output. Operational system architects and portfolio managers can use the same reference standards to increase system alignment and help meet interoperability goals.

Data and Information Controls Track (DIC)

The DIC track serves two basic purposes. First, it increases efficiencies associated with identification, definition, creation, and implementation of various controls and metrics. Secondly, it helps identify and monitor various data controls, like data security and data access, for key data domains.

Overall Solution

Leverage the EDG Council to help prioritize the creation of data quality metrics. Utilize existing workgroups to adopt access and publication standards.

2.2. Detailed Strategy

This section details the strategies of the five project tracks.

Data and Information Management Track (DIM)

- Use measurement tools and defined processes to ensure that data stewards can manage master data.
- Improve the quality of Medicaid provider and member data and measure the cost of non-conformance to data quality standards within the business processes that depend on data attributes.
- Increase the visibility of provider and member data between agency programs.

Data Quality and Standards Track (DQS)

- Adopt data quality metrics for provider attribution, member attribution, and operational source data quality, which could include:
  - Percentage of data assets with no identified business owner by departmental area,
  - Trends in data quality (such as the percentage of members with unresolvable physical addresses), and
  - Trends in managed care encounters with known defects.
- In conjunction with data stewards (embedded in the Center for Analytics and Decision Support, Medicaid/Children's Health Insurance Program, and Access and Eligibility Services), design dashboards within the insights platform that can track and trend critical data quality metrics.

Metadata and Reference Data Management Track (MRDM)

- High rate of fully described data assets as measured and tracked by the DAR scorecards on SharePoint.
  (*Note:* Sunset Report Issue 7 specifically mandates this data asset management.)
- Acquire technical system metadata and updated data models for all major HHS systems that will be held at a secure location.
- Implement change control processes for data collection, and update or replace system metadata at appropriate times in system development.
- Refresh and enhance the technical metadata from all major Medicaid source systems.

Data Architecture Track

- Re-charter a cross-functional EDG Council and create data governance workgroups. The EDG Steering Committee may invite state agency partners that share major data exchange processes as ex officio participants (e.g. Department of Family and Protective Services).
- Create data governance policies for reference data and master data.
- Publish reference values within the HHS insights platform.
- Establish data stewardship as a practice.
- Identify low utility, duplicate, or irrelevant data interfaces and exchanges.
- Identify high-value, trusted interfaces and exchanges and promote their re-use.
- Create and maintain a data architecture that aligns with the architecture referenced through the MITA framework.
- Coordinate with the IT governance model to determine critical areas, which are related to previous MITA information architecture deliverables, including the 2015 state self-assessment.
- Create enterprise models to identify and reference data.
- EDG Council working groups will use additional standards for technical architecture documentation to further align data management across the enterprise. These include data modeling and data dictionary standards, change management notification, and data stewardship reporting processes.
- The EDG Council will make recommendations for the use of member and provider "best version of the truth" records published to an Oracle repository. This repository will allow data analysts and subject matter experts to access mastered records. This data structure will serve as the foundation for future system analytics.

Data and Information Controls Track

- The EDG Council will engage in a partnership-planning model with IT Governance, the Office of the Chief Technology Officer, and others within the Office of the Chief Information Officer, to define the metrics of success for the EDG program.

Additionally, the EDG Council will adopt policies and procedures that govern:

- Access to the HHS master data repository (MDR) and associated metadata repositories.
- Adoption of reference code standards for use with analytics.
- Authorization of standard reference values for use in system analytics. (This will specify fundamental requirements for an HHS system analytic strategy.)

3. **Accomplishments**

- The EDG program adopted a maturity framework to score the metadata maturity of data assets. This framework allows for a dashboard related to metadata maturity, and maturity goals. In future years, this dashboard will help the EDG Council target the resources needed to improve metadata linked to strategic priorities.
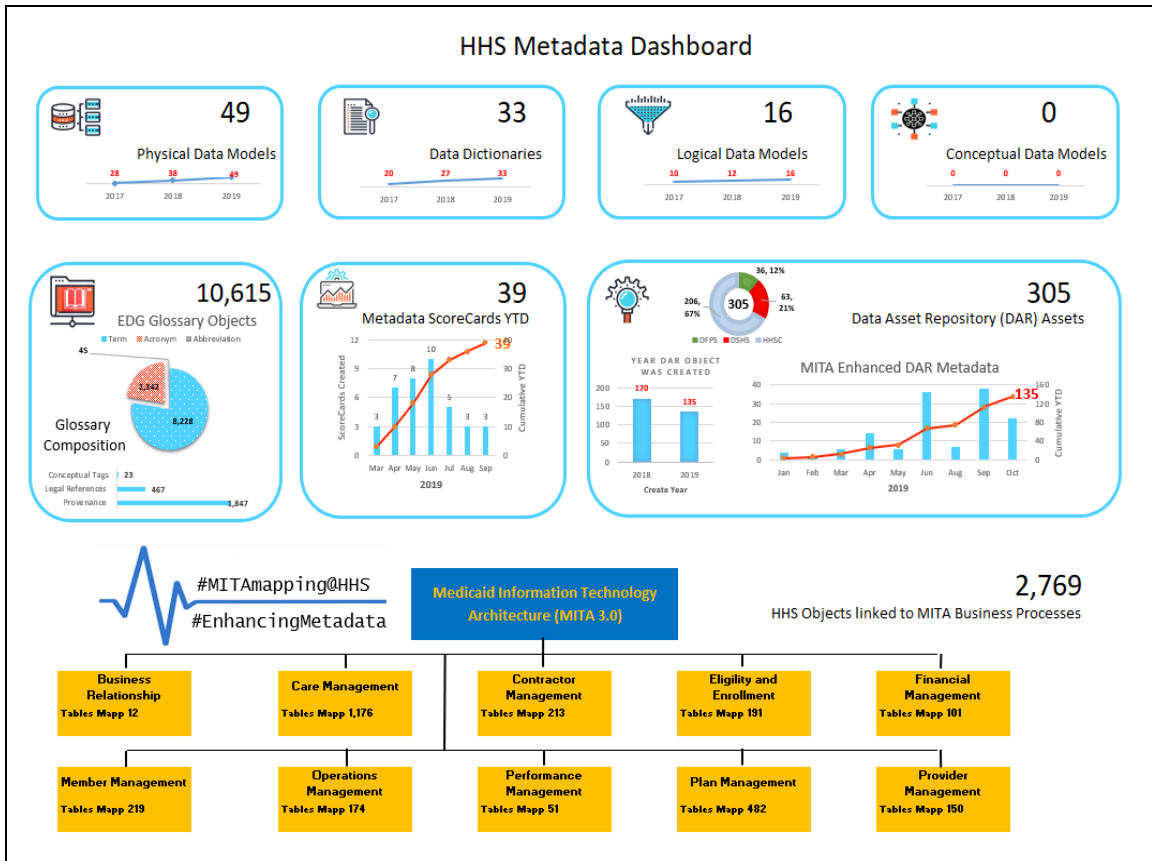
- The Metadata Repository supported discovery and user-directed browsing of HHS metadata and data models in a single interface. In fiscal year 2020 Q1, the EDG program added or enhanced technical and business metadata from the following major systems:
  - ‣ Encounters Operations Data Store,
  - ‣ Consolidated Services Interest List,
  - ‣ Service Authorization System Online, and
  - ‣ Long Term Care Services Interest List

- The DAR supported discovery and user-directed browsing of HHS data assets for users via SharePoint. In fiscal year 2020 Q1, the EDG program added the following three major assets to the DAR:
  - ‣ Service Authorization System Online
  - ‣ Consolidated Services Interest List
  - ‣ Combined Medicaid/Children's Health Insurance Program Datamart
  - ‣ Regional Intake System
  - ‣ Quality Assurance and Improvement System

- Completed eleven sprints focused on infrastructure development for data quality measurements, application development in MDM, and resolution of quality issues in provider attribution data.

4. Expenditures

4.1. Schedule of Expenditures and Actual Expenditures

The table below provides the schedule of expenditures (forecast), actual expenditures (expended), and remaining balance for the fiscal year 2020-21 biennium.

Rider 164 - Annual Reporting - EDG

For FY2020

Capital Expenditures as of October 31, 2019

Table 1 - Enterprise Data Governance

Rider 164 Reporting Categories	AY 2020 Forecast	AY 2020 Expended GR	AY 2020 Expended FED	AY 2020 Expended AF	AY 2020 Balance	AY 2021 Forecast	AY 2021 Expended GR	AY 2021 Expended FED	AY 2021 Expended AF	AY 2021 Balance
EDG Contracted Services	$3,820,000	$16,189	$145,703	$161,892	$3,658,108	$3,820,000	$0	$0	$0	$3,820,000
Software Licensing / Maintenance	$883,037	$13,183	$39,549	$52,732	$830,305	$883,740	$0	$0	$0	$883,740
Hardware	$77,400	$0	$0	$0	$77,400	$77,400	$0	$0	$0	$77,400
Data Center Services	$680,263	$11,270	$33,810	$45,080	$635,183	$677,135	$0	$0	$0	$677,135
Enterprise Data Governance	$5,460,700	$40,642	$219,062	$259,704	$5,200,996	$5,458,275	$0	$0	$0	$5,458,275

Table 2 - Enterprise Data Governance Appropriated Funds

	AY 2020 Forecast	AY 2020 Expended GR	AY 2020 Expended FED	AY 2020 Expended AF	AY 2020 Balance	AY 2021 Forecast	AY 2021 Expended GR	AY 2021 Expended FED	AY 2021 Expended AF	AY 2021 Balance
Total	$5,460,700	$40,642	$219,062	$259,704	$5,200,996	$5,458,275	$0	$0	$0	$5,458,275
Appropriated Funds	$5,460,700	$792,175	$4,668,525	$5,460,700	N/A	$5,458,275	$791,569	$4,666,706	$5,458,275	N/A
Difference	$0	$751,533	$4,449,463	$5,200,996	N/A	$0	$0	$0	$0	N/A

4.2. Year-to-Date Expenditures

The table below provides year-to-date budget and expenditures for EDG initiatives.

Year-to-Date Budget and Expenditures

Budget and Expenditures as of **October 31, 2019**

Table 3 - Adjusted Appropriated Funds with MOF

	AY2012	AY2013	AY2014	AY2015	AY2016	AY2017	AY2018	AY2019	AY2020	AY2021
General Revenue	$ 318,257	$2,601,088	$207,894	$10,352,835	$4,263,565	$6,297,166	$920,175	$888,175	$792,175	$791,569
Federal Funds	$2,853,399	$16,928,199	$1,719,440	$44,351,851	$31,247,878	$36,224,116	$5,820,525	$5,532,525	$4,668,525	$4,666,706
Total	$3,171,656	$19,529,287	$1,927,334	$54,704,686	$35,511,443	$42,521,282	$6,740,700	$6,420,700	$5,460,700	$5,458,275

Table 4 - Expenditures by LBB OOE

Object of Expense	AY2012	AY2013	AY2014	AY2015	AY2016	AY2017	AY2018	AY2019	AY2020	AY2021
2001 - Professional Fees and Services	$1,098,798	$574,463	$1,054,886	$3,064,902	$4,213,606	$2,619,899	$1,797,032	$3,086,534	$256,204	-
2003 - Consumable Supplies	-	-	$68	$53	-	-	-	-	-	-
2004 - Utilities	-	-	-	-	-	$34,890	$35,262	$35,724	-	
2007 - Rent - Machine and Other	-	-	-	-	-	-	-	-	-	-
2009 - Other Operating Expense	-	$788	$88,917	$678,415	$53,476	$2,545,405	$832,773	$780,473	$3,500	-
5000 - Capital Expenditures	-	-	-	$5,500,000	-	-	-	-	-	-
Grand Total	**$1,098,798**	**$575,251**	**$1,143,871**	**$9,243,370**	**$4,267,082**	**$5,200,193**	**$2,665,067**	**$3,902,731**	**$259,704**	**-**

Table 5 - MOF on Expenditures

	AY2012	AY2013	AY2014	AY2015	AY2016	AY2017	AY2018	AY2019	AY2020	AY2021
EDG -General Revenue	$109,880	$57,643	$127,769	$1,851,120	$456,845	$789,440	$463,240	$581,219	$40,642	-
EDG - Federal Funds	$988,918	$517,608	$1,016,102	$7,392,249	$3,810,236	$4,410,753	$2,201,827	$3,321,512	$219,062	-
sub-total EDG	**$1,098,798**	**$575,251**	**$1,143,871**	**$9,243,370**	**$4,267,082**	**$5,200,193**	**$2,665,067**	**$3,902,731**	**$259,704**	**-**

Table 6 - Outstanding Encumbrances

	AY2012	AY2013	AY2014	AY2015	AY2016	AY2017	AY2018	AY2019	AY2020	AY2021
General Revenue	-	-	-	-	-	-	-	$26,963	$651,573	-
Federal Funds	-	-	-	-	-	-	-	$242,664	$4,084,557	-
Total	-	-	-	-	-	-	-	$269,626	$4,736,129	-

Table 7 - Informational Dollars (non-capital)

	AY2012	AY2013	AY2014	AY2015	AY2016	AY2017	AY2018	AY2019	AY2020	AY2021
General Revenue	-	-	-	$113	-	$411	$2,303	-	-	-
Federal Funds	-	-	-	$1,021	-	$411	$13,125	-	-	-
Total	**-**	**-**	**-**	**$1,135**	**-**	**$822**	**$15,428**	**-**	**-**	**-**

5. Status

Table 8 - EDG Implementation Phase

Project Milestones	Planned Start Date mm/dd/yyyy	Actual Start Date mm/dd/yyyy	Planned Finish Date mm/dd/yyyy	Actual Finish Date mm/dd/yyyy	Percentage Complete
EDG Implementation Phase:	04/06/2015	08/20/2015	03/31/2022	-	65%

Table 9 - Medicaid Focused Enterprise Master Data Management System

Project Milestones	Planned Start Date mm/dd/yyyy	Actual Start Date mm/dd/yyyy	Planned Finish Date mm/dd/yyyy	Actual Finish Date mm/dd/yyyy	Percentage Complete
Data and information management Phase 1 – EMDM member/provider select systems with limited capabilities	04/06/2015	08/20/2015	08/31/2016	08/31/2016	100%
Data quality and data standards Phase 1 – metric definition, implement solution, and profiling	04/06/2015	08/20/2015	08/31/2016	08/31/2016	100%
Data and information management Phases 2-4	05/01/2016	09/22/2016	03/31/2022	-	62%
Data quality and data standards Phases 2-4	07/01/2017	09/22/2016	03/31/2022	-	59%

Table 10 - Metadata Repository

Project Milestones	Planned Start Date mm/dd/yyyy	Actual Start Date mm/dd/yyyy	Planned Finish Date mm/dd/yyyy	Actual Finish Date mm/dd/yyyy	Percentage Complete
Metadata management phase 1 – collect and publish metadata for various systems	04/06/2015	08/20/2015	02/28/2017	02/28/2017	100%
Metadata management phase 2 – implement change control process	05/04/2015	10/15/2015	02/28/2017	02/28/2017	100%
Metadata management phase 3	04/01/2017	03/01/2017	11/30/2018	11/30/2018	100%
Reference data management phase 1	01/01/2018	10/01/2018	09/29/2020	-	54%

Table 11 - Information Management Program

Project Milestones	Planned Start Date mm/dd/yyyy	Actual Start Date mm/dd/yyyy	Planned Finish Date mm/dd/yyyy	Actual Finish Date mm/dd/yyyy	Percentage Complete
Data architecture phase 1	01/01/2017	07/16/2018	07/13/2020	-	65%
Data and information control phase 1	03/04/2019	03/04/2019	03/31/2020	-	62%
Data architecture phase 2	04/02/2020	-	03/31/2022	-	0%
Data and information control phase 2	04/04/2019	04/04/2019	03/31/2022	-	20%