State Data Platform - Data Integration Standard

Introduction
This standard outlines the data integration for the state data platform, allowing one tool for consumption of data from sources into the Google Cloud platform. This standard supports the integration of stakeholders, vendors and third-party data within the state data platform. This document and information are applicable to all State of Oklahoma agencies and partners.

Purpose
The purpose of this document is to ensure there is a standard tool for data integration into the state data platform outlining a uniform and sustainable system.

Definitions
Data governance – A set of processes that ensures that data assets are formally managed throughout the enterprise. A data governance model establishes authority and management and decision-making parameters related to the data produced or managed by the enterprise (NIST CNSSI 4409-2015 NAS/CSS Policy 11-1).

Data integration – The process of retrieving data from multiple source systems and combining it in such a way that it can yield consistent, comprehensive, current and correct information for business reporting and analysis (Gartner IT Glossary).

Data lake – A concept consisting of a collection of storage instances of various data assets. These assets are stored in a near-exact, or even exact, copy of the source format and are in addition to the originating data stores (Gartner IT Glossary).

Data warehouse – A storage architecture designed to hold data extracted from transaction systems, operational data stores and external sources. It then combines that data in an aggregate, summary form for data analysis and reporting for predefined business needs (Gartner IT Glossary).

Standard
OMES has selected MuleSoft as the standard for data integration with the state data platform for B2B applications and Cloud Data Fusion as the ETL/ELT (extract, transform and load vs. extract, load and transform) standard for the state data platform. Cloud Data Fusion is a part of the Google Cloud platform solution. This product provides a sustainable model for data management, specifically for data integration. It provides scalable, distributed data lakes and data warehouses for analytics resulting in the integration of siloed data platforms. The data integration involves ETL/ELT, data transformation for both real-time and batch processing and data quality and data validation reducing data redundancy, data discrepancy, data duplication and providing more quality data for analysis. MuleSoft API Management is an integration platform between systems and the state data platform.

OMES IS data driven services is responsible for maintaining the architecture standards for the Cloud Data Fusion and MuleSoft relevant to the state data platform.

Compliance

This standard shall take effect upon publication and is made pursuant to Title 62 O.S. §§ 34.11.1 and 34.12 and Title 62 O.S. § 35.8. OMES IS may amend and publish the amended standards policies and standards at any time. Compliance is expected with all published policies and standards, and any published amendments thereof. Employees found in violation of this standard may be subject to disciplinary action, up to and including termination.

Rationale

To coordinate and require central approval of state agency information technology purchases and projects to enable the chief information officer to assess the needs and capabilities of state agencies as well as streamline and consolidate systems to ensure that the state delivers essential public services to its citizens in the most efficient manner at the lowest possible cost to taxpayers.

Revision history

This standard is subject to periodic review to ensure relevancy.

Effective date: 04/12/2022	**Review cycle:** Quarterly
Last revised: 04/12/2022	**Last reviewed:** 10/18/2022
Approved by: Jerry Moore, Chief Information Officer	