

Data Ingestion – Data Quality Standard

Introduction

Data loses value if it has poor quality as incorrect insights can be derived from incorrect data. As such, data quality checks must be put in place to ensure that the data is valuable.

Purpose

The purpose of this standard is to ensure that all data stored on the state data platform has the appropriate data quality checks in place. These checks should take the form of automated alerting to ensure rapid remediation.

Standard

To the greatest extent possible, the following automated alerts should be created for every ingestion pipeline on the state data platform:

Type of Alert	Metric Being Alerted On
Observability	<ul style="list-style-type: none"> Errors – Any issues that occur when data is being moved.
Audit	<ul style="list-style-type: none"> Completeness – The expected number of rows and columns are populated on the target system. Timeliness – The expected volume of data lands within the desired amount of time (could be combined with completeness, or separate if the entire dataset cannot be verified on a regular basis). Consistency – The data contains the expected content (i.e. a hash or sum of a primary field). Duplicate Entry – Ensure data is not being duplicated in transit.

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: 05/11/2022	Review cycle: Annual
Last revised: 05/11/2022	Last reviewed: 12/10/2024
Approved by: Joe McIntosh, Chief Information Officer	