Source Data Extract Controls Standard

Section 1 - Purpose

(1) This document provides requirements on how data is to be extracted from an operational source system. It outlines the need for data to be taken from the system without adjustment and outlines the controls that need to be in place to ensure down-stream data is replicated correctly, in a timely manner, and remains synchronised with the master source.

Section 2 - Authority

(2) Authority for this document is established by the Information Governance Policy.

Section 3 - Scope

(3) This standard applies:

    a. to all staff requesting or providing data from source systems

    b. regardless of the integration style used, such as batch files or APIs.

Section 4 - Standard

Introduction

(4) RMIT has many systems capturing and generating a significant amount of data. This data is often needed by systems or processes across the organisation and externally, outside of the system from which it originates.

(5) Requests for the data have created numerous extracts from the sources systems, most bespoke to meet the consumers' needs. This leads to significant processing load, as well as entanglement of systems, inhibiting updates of both producers and consumers.

(6) This set of standards and guidelines for source system owners intends to point towards providing a reduced, base set of 'raw' data extracts, allowing focus on the original primary intent of the system.

Data Extraction Standards

Data Extract Set

(7) Data extract set design must:

    a. be based on the source system's data structure, not a consumer's data requirements (accurate);

    b. ensure all of the system's business data is included (complete);

c. ensure there are no overlaps between extracts (efficient); and

d. not extract data that did not originate in the system (original).

Data Extract

(8) Data extract design must:

a. be based on the source system's data structure, not a consumer's data requirements (accurate);

b. include all data that naturally fits in the extract at hand, and does not omit data if it adds value (complete);

c. include all records (unfiltered); and

d. be natural, not enriched using business rules foreign to the system (not enriched).

Data Movement Controls

(9) All batch data extracts must have the following data movement controls:

a. extraction process end date-time

b. data batch record count

c. file checksum

d. data batch sequence number or business scheduled date

e. source system unique identifier

f. data set time zone identifier.

(10) Delta batch data extracts must also have the following data movement controls:

a. delta batch data point start date-time

b. delta batch data point end date-time.

Status and Details

Status	Current
Effective Date	11th November 2020
Review Date	3rd September 2022
Approval Authority	Chief Financial Officer
Approval Date	2nd October 2020
Expiry Date	Not Applicable
Policy Owner	Nicola Burgess Interim Executive Director, Technology
Policy Author	Chloe Sanford Director, Engagement & Enablement
Enquiries Contact	Data Management and Governance