

What is Critical Data?



Critical data is any of the following:

- Critical to business operations – Data that is essential to the operation and warrants a focus for data management.
- High adverse impact – Data that requires a higher level of data management or security controls and governance as the business will be adversely impacted if incorrect, including personal data.
- Derive critical measures – Data that may be used extensively and repeatedly because it is included in the calculation of critical measures that drive the success of business. The understanding of information lineage is critical.
- Compliance / regulatory – Data that supports compliance / regulatory obligations.
- Business priority – There is a commitment to investing time to govern this data.

Critical data is not:

- Used for data analysis processes that are not essential to the business – Data required for deriving insights, that is not necessary to support enterprise analysis and reporting.
- Used in operational or reporting processes with limited value add to business operations – Detailed workflow information where validity expires and/or is used for non-critical processes.
- Where the business priority is low.

Data Quality Dimensions



- Data quality dimensions enable the measurement of data quality
- A dataset may be assessed for quality across one or more dimensions, with the goal of determining whether the data is fit for its intended use
- Fitness for use is the degree to which the data is relevant, appropriate for the intended purpose and meets business specifications.
- Below are some examples but there are many others such as reasonableness, integrity and others

Dimension	Definition
Validity	The extent to which the data conforms to the defined domain of values.
Completeness	The extent to which the data is not missing and is of sufficient breadth and depth for the intended purpose.
Consistency	The extent to which there is a uniform representation of an entity record across all operational and analytical systems. This also applies to the alignment to external data sets which are deemed trustworthy.
Accuracy	The extent to which the data is error free and represents the correct information for IPEA to conduct its business. Accuracy is context sensitive and therefore needs to be confirmed by all stakeholders which use or require data.
Timeliness	The degree to which the data represents reality at the required point in time.
Uniqueness	The degree to which no entity exists more than once within the data set.

Examples of Data Quality Standards

Dimension & Rule	Measure	Calculation	Example Standard/Result
Validity: Travel expenses can only be claimed if they are for a valid business purpose	Count the number of travel claim records where purpose is set to a valid expense code	Count of invalid records as a percentage of the total number of records	Unacceptable is above 0%
Completeness: Population of a field is mandatory	Count of populated and unpopulated fields	Count of populated fields divided by total count	Acceptable is 95% or above populated
Consistency: Phone numbers for offices are consistent with the office location	Count of phone numbers that are not consistent the location of the office.	Count of inconsistent phone records as a percentage of total number of records	Acceptable is 10% or below
Accuracy: Expense values on receipts match values entered in systems	Count receipts or reports values that do not match system values	Count of non-matching values as a percentage of total number of records	Acceptable is 0.01% or below
Timeliness: Records arrive within the schedule time frame	Count of processed and available records Total count of records (processed and unprocessed)	Count of processed and available records as a percentage of total number of records	Acceptable is 95% or above processed and available
Uniqueness: No entity exists more than once within the dataset	Count the number of duplicate records in the dataset.	Count of duplicates as a percentage of the total number of records	Unacceptable is above 2%