

Definitions of Data Categories

Data categories are groupings of data with common characteristics or features. The table that starts here and continues on the next page includes definitions and examples for major

data categories. These definitions were jointly created by Danette McGilvray and Gwen Thomas, president of the Data Governance Institute.

• Definitions of Data Categories

DATA CATEGORY	DEFINITION
Master Data	<p>Master data describe the people, places, and things that are involved in an organization's business.</p> <p><i>Examples</i> include people (e.g., customers, employees, vendors, suppliers), places (e.g., locations, sales territories, offices), and things (e.g., accounts, products, assets, document sets).</p> <p>Because these data tend to be used by multiple business processes and IT systems, standardizing master data formats and synchronizing values are critical for successful system integration.</p> <p>Master data tend to be grouped into master records, which may include associated reference data. An example of associated reference data is a state field within an address in a customer master record.</p>
Transactional Data	<p>Transactional data describe an internal or external event or transaction that takes place as an organization conducts its business.</p> <p><i>Examples</i> include sales orders, invoices, purchase orders, shipping documents, passport applications, credit card payments, and insurance claims.</p> <p>These data are typically grouped into transactional records, which include associated master and reference data.</p>
Reference Data	<p>Reference data are sets of values or classification schemas that are referred to by systems, applications, data stores, processes, and reports, as well as by transactional and master records.</p> <p><i>Examples</i> include lists of valid values, code lists, status codes, state abbreviations, demographic fields, flags, product types, gender, chart of accounts, and product hierarchy.</p> <p>Standardized reference data are key to data integration and interoperability and facilitate the sharing and reporting of information. Reference data may be used to differentiate one type of record from another for categorization and analysis, or they may be a significant fact such as country, which appears within a larger information set such as address.</p> <p>Organizations often create internal reference data to characterize or standardize their own information. Reference data sets are also defined by external groups, such as government or regulatory bodies, to be used by multiple organizations. For example, currency codes are defined and maintained by ISO.</p>

(Continued)

• Definitions of Data Categories (Continued)

DATA CATEGORY	DEFINITION
Metadata	<p>Metadata literally means “data about data.” Metadata label, describe, or characterize other data and make it easier to retrieve, interpret, or use information.</p> <p>Technical metadata are metadata used to describe technology and data structures. Examples of technical metadata are field names, length, type, lineage, and database table layouts.</p> <p>Business metadata describe the nontechnical aspects of data and their usage. Examples are field definitions, report names, headings in reports and on Web pages, application screen names, data quality statistics, and the parties accountable for data quality for a particular field. Some organizations would classify ETL (Extract–Transform–Load) transformations as business metadata.</p> <p>Audit trail metadata are a specific type of metadata, typically stored in a record and protected from alteration, that capture how, when, and by whom the data were created, accessed, updated, or deleted. Audit trail metadata are used for security, compliance, or forensic purposes. Examples include timestamp, creator, create date, and update date. Although audit trail metadata are typically stored in a record, technical metadata and business metadata are usually stored separately from the data they describe.</p> <p>These are the most common types of metadata, but it could be argued that there are other types of metadata that make it easier to retrieve, interpret, or use information. The label for any metadata may not be as important as the fact that it is being deliberately used to support data goals. Any discipline or activity that uses data is likely to have associated metadata.</p>
<i>Additional data categories that impact how systems and databases are designed and data are used:</i>	
Historical Data	<p>Historical data contain significant facts, as of a certain point in time, that should not be altered except to correct an error. They are important to security and compliance. Operational systems can also contain history tables for reporting or analysis purposes. Examples include point-in-time reports, database snapshots, and version information.</p>
Temporary Data	<p>Temporary data are kept in memory to speed up processing. They are not viewed by humans and are used for technical purposes. Examples include a copy of a table that is created during a processing session to speed up lookups.</p>