Data Definitions

Glossary[1]

Aggregate data: Data combined or collected together in summary or other form such that the data cannot be identified with any individual.

Big data: Huge volumes of data passively emitted and/or collected by digital devices that are not possible to process with traditional methods.

Closed data: Data that requires, if shared, a user-specific custom license or contract for use.

Data attribute: Any general characteristic that can be associated with a specific collection of data. Examples include data type, temporality, duration, sensitivity, etc.

Data classification: Data classification is the process of separating and organizing data into relevant groups ("classes") based on their shared characteristics, such as their level of sensitivity, the risks they present, and the compliance regulations that protect them.

Data harmonization: Process of combining data from different sources in a way that provides a comparable view of the data and reorganizing into a single scheme

Data integration: Process of translating data across systems to bring data together into one location. This is like a conversation through an interpreter.

Data interoperability: Ability for systems to speak to each other. This is like speaking the same language.

Data linkages: Method of bringing information from different sources together about the same person or entity to create a new, richer dataset.

Data set: A collection of discrete items of related data that may be accessed individually or in combination or managed as a whole entity. A data set is organized into some type of data structure.

Data source: Location or system that stores and manages data, and it can take on many different forms, from traditional databases and spreadsheets to cloud-based platforms and APIs.

Data triage: Data triage is the process of identifying, prioritizing, and addressing data quality issue

Data triangulation: Use of multiple methods or data sources (across time, space, people) to develop a comprehensive more reliable/valid understanding.

[1] This glossary only intends to offer a range of commonly used business terms and definitions and does not include official definitions.

Data valuation: Process that encompasses determining the value of data that is stored, collected, analyzed, and exchanged by companies or organizations

Macro data: Large-scale data that can provide insights into broad trends.

Meta data: Data that describes other data, providing a structured reference that helps to sort and identify attributes of the information it describes.

Micro data: Microdata are sets of records containing information on individuals, households or businesses.

Open data: Data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share-alike.

Qualitative data: Qualitative data is non-numeric information, such as in-depth interview transcripts, diaries, anthropological field notes, answers to open-ended survey questions, audio-visual recordings and images.

Quantitative data: Quantitative data that be counted or measured in numerical values.

Raw data: Data in its original form, not yet formatted or processed.

Semi structured data: Semi-structured data is a combination of structured and unstructured data and shares characteristics of both. It also follows certain schema, consistency and exists to ease space.

Structured data: Structured data is when data is in a standardized format, has a well-defined structure, complies to a data model, follows a persistent order, and is easily accessed by humans and programs. This data type is generally stored in a database.

Unstructured data: Data not stored in a structured database format. Unstructured data has an internal structure, but it's not predefined through data models.

Virtuous cycle of data: A virtuous cycle of data occurs when the value of the data is good and processes are transparent so appropriate resources are allocated to standardise and improve the data collection process. This results in coherent, trusted data with clear validation paths to the sources.

Sources: Alpha Serve, TechTarget, Leanscape, Datamation, Legal Information Institute, European Commission, DataScientest, Data-Pop Alliance, ScienceDirect, Spotfire, NetApp, UK Data Service, Australia Bureau of Statistics, The Open Data Handbook, IcebreakerOne, US National Institutes of Health, Spirion, Menzies Institute for Medical Research, Revelate, Overseas Development Institute