

## : *What is data?*

by Daniel J. Power

Data is important. Data is growing rapidly. Approximately 2.5 quintillion bytes of data are created each day or 2.5 billion gigabytes each day (as of May 2018). So approximately 1.7 megabytes of data is created every second for every person on earth. Trends indicate that there is a 40% growth in global data each year. Capturing, managing, storing and retrieving data are fundamental tasks that must be performed effectively in modern organizations. Data volumes and data velocity are increasing rapidly in organizations. Business managers want to capture and analyze all data as it is generated and store it for future analysis. Data is increasingly varied. Data is streaming over the global Internet to our computing devices and then back to online storage in the cloud. For these reasons, managing data is increasingly challenging and difficult. In a computing environment, data is processed and stored in a binary digital format.

Data is a plural of datum, which was originally a Latin noun meaning "something given." "Data" is used as both a singular and a plural noun in a computing context. Some people use "data" as a plural noun that precedes a plural verb --"Data are important in decision making." Most people use "data" as a general-purpose noun followed by a singular verb --"Data is important." "Data is accumulating quickly." Data exists in a variety of forms, including text on paper and bytes stored in electronic memory. A datum describes a single quality or quantity of some object or phenomenon.

Data is a broad concept referring to both digital data that is machine-readable and non-digital content. Data is any sequence of one or more symbols. The main types of data that can be input into a computer and processed are numeric, text, dates, graphics and sound. Some authors define data as facts from which information can be derived. Data, when viewed as an abstract concept, is the lowest or most granular and detailed level from which information and then knowledge can be derived. Raw or source data refers to any unprocessed collection of numbers, characters or images. Data processing commonly occurs by stages and the "processed data" from one stage may be considered the "raw data" of the next stage of processing. Although the terms "data", "information" and "knowledge" are often used interchangeably, each of these terms has a distinct meaning.

Data are the quantities, characters, or symbols on which operations are performed by a computer, being stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media. Data describes a single quality or quantity of some object or phenomenon. Raw data, also known as primary data, is data collected directly from a source. Data records what is perceived. Raw data refers to data that have not been changed since acquisition. Editing, cleaning or modifying the raw data results in lightly processed data. When data are processed, interpreted, organized, structured and presented to people in a meaningful or useful way, the data is transformed into information. Information is data presented and analyzed in context.

## : *What is data?*

Structured data is stored using clearly defined data types in a pattern that makes data easily searchable; while unstructured data – “everything else” – is “comprised of data that is usually not as easily searchable, including formats like audio, video, and social media postings”. There is also semi-structured data that “maintains internal tags and markings that identify separate data elements, which enables information grouping and hierarchies”, cf. Taylor (2018). There is much more unstructured data than structured data.

In general, data should be fine-grained, composed of small, distinguishable pieces, with high granularity. The greater the granularity, the deeper the level of detail. Granularity characterizes the scale or level of detail in a set of data. Data are a set of symbols or tokens that refer to a meaning or value (Mons, 2018). Digital data is what computers read, write and process, Power (2013). Numeric data may be discrete or continuous. Big data refers to a huge and growing digital data volume as well as the increasing data variety and the rapidly increasing velocity of both data collection and data movement.

Quality data has five characteristics: 1) accurate and precise, 2) complete and valid, 3) consistent and reliable, 4) relevant to the requirements of the user, and 5) timely. The 10 characteristics of data quality in the American Health Information Management Association (AHIMA) data quality model are Accuracy, Accessibility, Comprehensiveness, Consistency, Currency, Definition, Granularity, Precision, Relevance and Timeliness, cf. <http://bok.ahima.org/>. A collection of relevant data is called a database. High-quality data is easily accessible and can be processed further with relative ease, cf., Sharma (2019). Quality data are crucial for effective organization functioning.

"Data is pervasive and plentiful. And the things it can do for you are amazing. It can propel your business, improve your life and help people ... Data is the most plentiful natural resource of the 21st century", cf., Alexander (2016).

Data are **G**ranular **F**acts captured as **S**ymbols that are generally **M**achine-readable (GFSM). Data are raw, unprocessed facts. In summary, data are highly granular and detailed facts and figures captured as symbols that represent things and it is often stored as machine-readable digital data. Capturing data and using it in computing and information systems has been creating a data-driven global society for many years. Processed digital data can create value.

## References

AHIMA Data Quality Management Task Force. “Practice Brief: Data Quality Management Model.” Journal of AHIMA 69, no. 6 (1998): p. 2-7 of insert before p. 73.

## : *What is data?*

Alexander, F., "Data is everywhere and that's a good thing," IBM Business Analytics Blog, Oct. 07, 2016 at URL <https://www.ibm.com/blogs/business-analytics/data-is-everywhere/>

Mons, B., Data Stewardship for Open Science: Implementing FAIR Principles, CRC Press, 2018.

Power, D., "What is digital data?" DSS News, Vol. 14, No. 15, 07/21/2013 at URL <http://dssresources.com/faq/index.php?action=artikel&id=277>.

Sharma, J., "What Is Data And Its Characteristics?" Rebellion Rider, January 15, 2019 at URL <http://www.rebellionrider.com/data-definition-and-characteristics>

Taylor, C., "Structured vs. Unstructured Data," Datamation, March 28, 2018 at URL <https://www.datamation.com/big-data/structured-vs-unstructured-data.html>

Author: Daniel Power

Last update: 2020-01-06 03:27