

# Data



*Some of the different types of data.*

**Data** are characteristics or information, usually numerical, that are collected through observation.<sup>[1]</sup> In a more technical sense, data is a set of values of qualitative

or quantitative variables about one or more persons or objects, while a **datum** (singular of data) is a single value of a single variable.<sup>[2]</sup>

Although the terms "data" and "information" are often used interchangeably, these terms have distinct meanings. In some popular publications, data is sometimes said to be transformed into information when it is viewed in context or in post-analysis.<sup>[3]</sup> In academic treatments of the subject, however, data are simply units of information. Data is employed in scientific research, businesses management (e.g., sales data,

revenue, profits, stock price), finance, governance (e.g., crime rates, unemployment rates, literacy rates), and in virtually every other form of human organizational activity (e.g., censuses of the number of homeless people by non-profit organizations).

Data is measured, collected and reported, and analyzed, whereupon it can be visualized using graphs, images or other analysis tools. Data as a general concept refers to the fact that some existing information or knowledge is represented or coded in some form suitable for better usage or processing. Raw data

("unprocessed data") is a collection of numbers or characters before it has been "cleaned" and corrected by researchers. Raw data needs to be corrected to remove outliers or obvious instrument or data entry errors (e.g., a thermometer reading from an outdoor Arctic location recording a tropical temperature). Data processing commonly occurs by stages, and the "processed data" from one stage may be considered the "raw data" of the next stage. Field data is raw data that is collected in an uncontrolled "in situ" environment. Experimental data is data that is generated within the context of a

scientific investigation by observation and recording.

Data has been described as the new oil of the digital economy.<sup>[4][5]</sup>

## Etymology and terminology

The first English use of the word "data" is from the 1640s. The word "data" was first used to mean "transmissible and storable computer information" in 1946. The expression "data processing" was first used in 1954.<sup>[6]</sup>

The Latin word *data* is the plural of *datum*, "(thing) given," neuter past participle of

*dare* "to give".<sup>[6]</sup> Data may be used as a plural noun in this sense, with some writers—usually scientific writers—in the 20th century using *datum* in the singular and *data* for plural. However, in everyday language, "data" is most commonly used in the singular, as a mass noun (like "sand" or "rain"). The APA manual of style requires "data" to be plural.<sup>[7]</sup>

## Meaning

the Reader, that he hath found, that the Apertures, which *Optick Glasses* can bear with distinctness, are in about a *subduplicate proportion* to their *Lengths*; whereof he tells us he intends to give the reason and demonstration in his *Dioptrick*, which he is now writing, and intends to finish, as soon as his Health will permit. In the mean time, he presents the Reader with a *Table* of such *Apertures*; which is here exhibited to the Consideration of the Ingenious, there being of this *French Book* but one Copy, that is known, in *England*.

A TABLE of the Apertures of Object-Glasses.

The Points put to some of these Numbers denote Fractions.

Length of Glasses.	For excellent use.		For ordinary use.		Length of Glasses.	For excellent use.		For ordinary use.		
	Foot.	Line.	Foot.	Line.		Foot.	Line.	Foot.	Line.	
4	4	4	2	15	2	4	2	10	2	4
6	7	5	4	20	2	3	3	2	2	7
9	7	6	5	25	4	0	2	4	2	10
12	8	7	6	40	4	3	3	7	2	10
15	9	8	7	45	4	6	2	10	1	2
20	11	10	8	50	4	9	4	0	1	4
24	11	11	9	55	5	0	4	3	2	6
30	11	11	10	60	5	2	4	6	2	8
36	11	11	11	65	5	4	4	8	2	10
42	11	11	11	70	5	7	4	10	4	10
48	11	11	11	75	5	9	5	0	4	2
54	11	11	11	80	5	11	5	2	4	5
60	11	11	11	85	6	4	6	6	4	7
66	11	11	11	90	6	8	6	9	4	10
72	11	11	11	95	7	3	6	1	5	3
78	11	11	11	100	7	0	7	0	5	11
84	11	11	11	100	7	6	8	0	6	9
90	11	11	11	100	8	2	9	2	7	8
96	11	11	11	100	8	8	9	10	8	5
102	11	11	11	100	9	4	10	9	9	0
108	11	11	11	100	9	10	11	6	9	8
114	11	11	11	100	10	6	11	6	9	2
120	11	11	11	100	10	12	11	6	9	2

Adrien Auzout's "A TABLE of the Apertures of Object-Glasses" from a 1665 article in Philosophical Transactions

Data, information, knowledge and wisdom are closely related concepts, but each has its own role in relation to the other, and each term has its own meaning. According to a common view, data is collected and analyzed; data only becomes information

suitable for making decisions once it has been analyzed in some fashion.<sup>[8]</sup> One can say that the extent to which a set of data is informative to someone depends on the extent to which it is unexpected by that person. The amount of information contained in a data stream may be characterized by its Shannon entropy.

Knowledge is the understanding based on extensive experience dealing with information on a subject. For example, the height of Mount Everest is generally considered data. The height can be measured precisely with an altimeter and entered into a database. This data may be



included in a book along with other data on Mount Everest to describe the mountain in a manner useful for those who wish to make a decision about the best method to climb it. An understanding based on experience climbing mountains that could advise persons on the way to reach Mount Everest's peak may be seen as "knowledge". The practical climbing of Mount Everest's peak based on this knowledge may be seen as "wisdom". In other words, wisdom refers to the practical application of a person's knowledge in those circumstances where good may result. Thus wisdom complements and completes the series "data", "information"

and "knowledge" of increasingly abstract concepts.

Data is often assumed to be the least abstract concept, information the next least, and knowledge the most abstract.<sup>[9]</sup>

In this view, data becomes information by interpretation; e.g., the height of Mount Everest is generally considered "data", a book on Mount Everest geological characteristics may be considered "information", and a climber's guidebook containing practical information on the best way to reach Mount Everest's peak may be considered "knowledge".

"Information" bears a diversity of

meanings that ranges from everyday usage to technical use. This view, however, has also been argued to reverse the way in which data emerges from information, and information from knowledge.<sup>[10]</sup> Generally speaking, the concept of information is closely related to notions of constraint, communication, control, data, form, instruction, knowledge, meaning, mental stimulus, pattern, perception, and representation. Beynon-Davies uses the concept of a sign to differentiate between data and information; data is a series of symbols, while information occurs when the symbols are used to refer to something.<sup>[11][12]</sup>

Before the development of computing devices and machines, people had to manually collect data and impose patterns on it. Since the development of computing devices and machines, these devices can also collect data. In the 2010s, computers are widely used in many fields to collect data and sort or process it, in disciplines ranging from marketing, analysis of social services usage by citizens to scientific research. These patterns in data are seen as information which can be used to enhance knowledge. These patterns may be interpreted as "truth" (though "truth" can be a subjective concept), and may be authorized as aesthetic and ethical criteria

in some disciplines or cultures. Events that leave behind perceivable physical or virtual remains can be traced back through data. Marks are no longer considered data once the link between the mark and observation is broken.<sup>[13]</sup>

Mechanical computing devices are classified according to the means by which they represent data. An analog computer represents a datum as a voltage, distance, position, or other physical quantity. A digital computer represents a piece of data as a sequence of symbols drawn from a fixed alphabet. The most common digital computers use a binary

alphabet, that is, an alphabet of two characters, typically denoted "0" and "1". More familiar representations, such as numbers or letters, are then constructed from the binary alphabet. Some special forms of data are distinguished. A computer program is a collection of data, which can be interpreted as instructions. Most computer languages make a distinction between programs and the other data on which programs operate, but in some languages, notably Lisp and similar languages, programs are essentially indistinguishable from other data. It is also useful to distinguish metadata, that is, a description of other

data. A similar yet earlier term for metadata is "ancillary data." The prototypical example of metadata is the library catalog, which is a description of the contents of books.

## **Data collection**

...

Gathering data can be accomplished through a primary source (the researcher is the first person to obtain the data) or a secondary source (the researcher obtains the data that has already been collected by other sources, such as data disseminated in a scientific journal). Data analysis methodologies vary and include data

triangulation and data percolation.<sup>[14]</sup> The latter offers an articulate method of collecting, classifying and analyzing data using five possible angles of analysis (at least three) in order to maximize the research's objectivity and permit an understanding of the phenomena under investigation as complete as possible: qualitative and quantitative methods, literature reviews (including scholarly articles), interviews with experts, and computer simulation. The data is thereafter "percolated" using a series of pre-determined steps so as to extract the most relevant information.



# In other fields

Although data is also increasingly used in other fields, it has been suggested that the highly interpretive nature of them might be at odds with the ethos of data as "given".

Peter Checkland introduced the term *capta* (from the Latin *capere*, "to take") to distinguish between an immense number of possible data and a sub-set of them, to which attention is oriented.<sup>[15]</sup> Johanna

Drucker has argued that since the humanities affirm knowledge production as "situated, partial, and constitutive," using *data* may introduce assumptions that are counterproductive, for example

that phenomena are discrete or are observer-independent.<sup>[16]</sup> The term *capta*, which emphasizes the act of observation as constitutive, is offered as an alternative to *data* for visual representations in the humanities.

## See also

- Biological data
- Computer memory.
- Data (Star Trek).
- Data acquisition
- Data analysis
- Data cable

- Data curation
- Dark data
- Data domain
- Data element
- Data farming
- Data governance
- Data integrity
- Data maintenance
- Data management
- Data mining
- Data modeling
- Data point
- Data visualization
- Computer data processing

- Data preservation
- Data publication
- Data protection
- Data remanence
- Data science
- Data set
- Data structure
- Data warehouse
- Database
- Datasheet
- Environmental data rescue
- Fieldwork
- Information engineering
- Machine learning

- Open data
- Scientific data archiving
- Statistics
- Secondary Data

## References

---

This article is based on material taken from the *Free On-line Dictionary of Computing* prior to 1 November 2008 and incorporated under the "relicensing" terms of the GFDL, version 1.3 or later.

1. *OECD Glossary of Statistical Terms*.  
OECD. 2008. p. 119. ISBN 978-92-64-025561.

2. *"Statistical Language - What are Data?" . Australian Bureau of Statistics. 2013-07-13. Archived from the original on 2019-04-19. Retrieved 2020-03-09.*
3. *"Data vs Information - Difference and Comparison | Diffen" .  
www.diffen.com. Retrieved 2018-12-11.*
4. *Yonego, Joris Toonders (July 23, 2014). "Data Is the New Oil of the Digital Economy" – via  
www.wired.com.*

5. *"Data is the new oil" . July 16, 2018. Archived from the original on 2018-07-16.*
6. *"data | Origin and meaning of data by Online Etymology Dictionary" .  
www.etymonline.com.*
7. *"APA Style 6th Edition Blog: Data Is, or Data Are?" . blog.apastyle.org.*
8. *"Joint Publication 2-0, Joint Intelligence" (PDF). Joint Chiefs of Staff, Joint Doctrine Publications. Department of Defense. 23 October 2013. pp. I-1. Retrieved July 17, 2018.*
9. *Akash Mitra (2011). "Classifying data for successful modeling" .*

10. *Tuomi, Ilkka (2000). "Data is more than knowledge". Journal of Management Information Systems. 6 (3): 103–117. doi:10.1080/07421222.1999.11518258 .*
11. *P. Beynon-Davies (2002). Information Systems: An introduction to informatics in organisations. Basingstoke, UK: Palgrave Macmillan. ISBN 0-333-96390-3.*
12. *P. Beynon-Davies (2009). Business information systems. Basingstoke, UK: Palgrave. ISBN 978-0-230-20368-6.*
13. *Sharon Daniel. The Database: An Aesthetics of Dignity.*



14. *Mesly, Olivier (2015). Creating Models in Psychological Research. États-Unis : Springer Psychology : 126 pages. ISBN 978-3-319-15752-8*
15. *P. Checkland and S. Holwell (1998). Information, Systems, and Information Systems: Making Sense of the Field. Chichester, West Sussex: John Wiley & Sons. pp. 86–89. ISBN 0-471-95820-4.*
16. *Johanna Drucker (2011). "Humanities Approaches to Graphical Display" .*

## External links

---

Look up **data** in Wiktionary, the free dictionary.

Wikimedia Commons has media related to **Data**.

- Data is a singular noun (a detailed assessment)

Retrieved from

["https://en.wikipedia.org/w/index.php?title=Data&oldid=966279285"](https://en.wikipedia.org/w/index.php?title=Data&oldid=966279285)

---

Last edited 3 days ago by Bruce1ee

Content is available under CC BY-SA 3.0 unless otherwise noted.