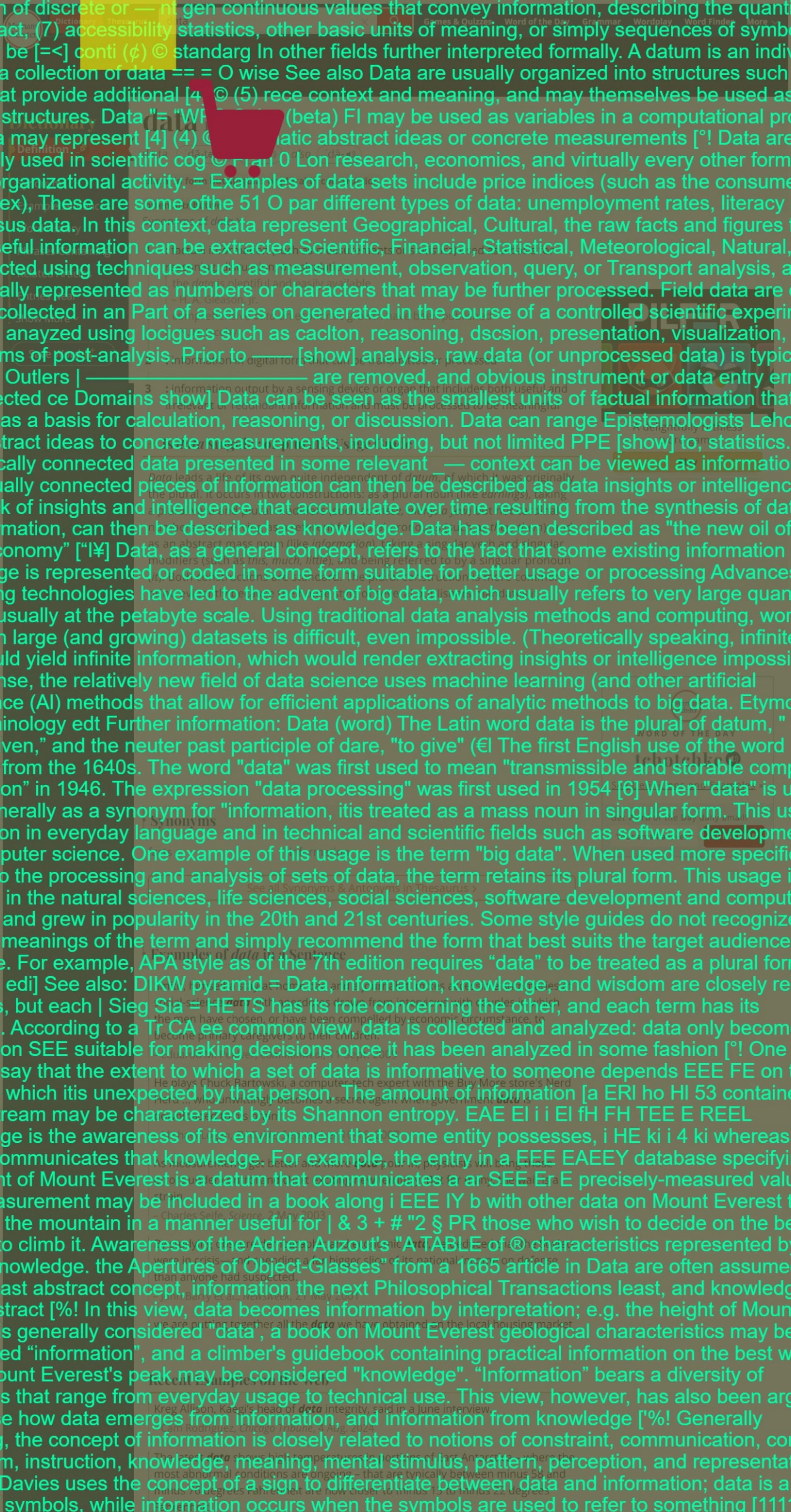

aData MA 97 languages Contents hide Article Talk Read Edit View history Tools Vv Appearance hide (Top) From Wikipedia, the free encyclopedia Text among "Scientific data" redirects here. Not to be confused with Scientific Data (journal) ©) smal — For data in computer science, see Data (computer science). For other uses, see Data (disambiguation) and Datum @) standara ™ Data documents (disambiguation). O Large Data collection In common usage, data (/ derts/, aiso Us: / deste) is a collection of discrete or — nt gen continuous values that convey information, describing the quantity, quality, fact, (7) accessibility statistics, other basic units of meaning, or simply sequences of symbols that may be [=<] conti (¢) © standarg In other fields further interpreted formally. A datum is an individual value in a collection of data == = O wise See also Data are usually organized into structures such as tables that provide additional [4 © (5) rece context and meaning, and may themselves be used as data in larger structures. Data "= "WP (beta) FI may be used as variables in a computational process [12] Data may represent [4] (4) © atic abstract ideas or concrete measurements [°! Data are commonly used in scientific cog © Fran 0 Lon research, economics, and virtually every other form of human organizational activity. = Examples of data sets include price indices (such as the consumer price index), These are some ofthe 51 O par different types of data: unemployment rates, literacy rates, and census data. In this context, data represent Geographical, Cultural, the raw facts and figures from which useful information can be extracted Scientific, Financial, Statistical, Meteorological, Natural, Data are collected using techniques such as measurement, observation, query, or Transport analysis, and are typically represented as numbers or characters that may be further processed. Field data are data that are collected in an Part of a series on generated in the course of a controlled scientific experiment. Data are anayzed using locigues such as caclton, reasoning, dscsion, presentation, visualization, or other forms of post-analysis. Prior to —— [show] analysis, raw data (or unprocessed data) is typically cleaned: Outlers | ——————————— are removed, and obvious instrument or data entry errors are corrected ce Domains show] Data can be seen as the smallest units of factual information that can be used as a basis for calculation, reasoning, or discussion. Data can range Epistemologists Lehow] from abstract ideas to concrete measurements, including, but not limited PPE [show] to, statistics. Thematically connected data presented in some relevant — context can be viewed as information. Contextually connected pieces of information can then be described as data insights or intelligence. The stock of insights and intelligence that accumulate over time resulting from the synthesis of data into information, can then be described as knowledge. Data has been described as "the new oil of the digital economy" ["I¥] 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 Advances in computing technologies have led to the advent of big data, which usually refers to very large quantities of data, usually at the petabyte scale. Using traditional data analysis methods and computing, working with such large (and growing) datasets is difficult, even impossible. (Theoretically speaking, infinite data would yield infinite information, which would render extracting insights or intelligence impossible.) In response, the relatively new field of data science uses machine learning (and other artificial intelligence (AI) methods that allow for efficient applications of analytic methods to big data. Etymology and terminology edt Further information: Data (word) The Latin word data is the plural of datum, " (thing) given," and the neuter past participle of dare, "to give" (€l 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 [6] When "data" is used more generally as a synonym for "information, itis treated as a mass noun in singular form. This usage is common in everyday language and in technical and scientific fields such as software development and computer science. One example of this usage is the term "big data". When used more specifically to refer to the processing and analysis of sets of data, the term retains its plural form. This usage is common in the natural sciences, life sciences, social sciences, software development and computer science, and grew in popularity in the 20th and 21st centuries. Some style guides do not recognize the different meanings of the term and simply recommend the form that best suits the target audience of the guide. For example, APA style as of the 7th edition requires "data" to be treated as a plural form [7] Meaning edi] See also: DIKW pyramid = Data, information, knowledge, and wisdom are closely related concepts, but each | Sieg Sia = HE TS has its role concerning the other, and each term has its meaning. According to a Tr CA ee common view, data is collected and analyzed: data only becomes information SEE suitable for making decisions once it has been analyzed in some fashion [°! One BT fers can say that the extent to which a set of data is informative to someone depends EEE FE on the extent to which itis unexpected by that person. The amount of information [a ERI ho HI 53 contained in a data stream may be characterized by its Shannon entropy. EAE El i i El fH FH TEE E REEL Knowledge is the awareness of its environment that some entity possesses, i HE ki i 4 ki whereas data merely communicates that knowledge. For example, the entry in a EEE EAEEY database specifying the height of Mount Everest is a datum that communicates a ar SEE Er E precisely-measured value. This measurement may be included in a book along i EEE IY b with other data on Mount Everest to describe the mountain in a manner useful for | & 3 + # "2 § PR those who wish to decide on the best method to climb it. Awareness of the Adrien Auzout's "A TABLE of ® characteristics represented by this data is knowledge. the Apertures of Object-Glasses" from a 1665 article in Data are often assumed to be the least abstract concept, information the next Philosophical Transactions least, and knowledge the most abstract [%! 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 range from everyday usage to technical use. This view, however, has also been argued to reverse how data emerges from information, and information from knowledge ['%! 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 [11112]

Before the development of computing devices and machines, people had to manually collect data and impose patterns on it. With the development of computing devices and machines, these devices can also collect data. In the 2010s, computers were widely used in many fields to collect data and sort or process it, in disciplines ranging from marketing, analysis of social service usage by citizens to scientific research. These patterns in the data are seen as information that 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 ['*] Mechanical computing devices are classified according to how 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, that 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 documents edi) Whenever data needs to be registered, data exists in the form of a CRESS data document. Kinds of data documents include Library and information science . hoy crests = « data study by Ere + data set pip 3 FE ke « software Es « data paper Y Y/ Ve « database \% Z pei oom | + data journal Histories [show] Some of these data documents (data repositories, data studies, data Focus [show] sets, and software) are indexed in Data Citation Indexes, while data papers are indexed in traditional bibliographic databases, e.g., ELISE [Show] Science Citation Index. Interdisciplinary fields [show] Areas [show] Data collection [edi] 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 ['#] The latter offers an articulate method of collecting, classifying, and analyzing data using five possible angles of analysis (at least three) 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 Data longevity and accessibility edi An important field in computer science, technology, and library science is the longevity of data. Scientific research generates huge amounts of data, especially in genomics and astronomy, but also in the medical sciences, e.g. in medical imaging. In the past, scientific data has been published in papers and books, stored in libraries, but more recently practically all data is stored on hard drives or optical discs. However, in contrast to paper, these storage devices may become unreadable after a few decades. Scientific publishers and libraries have been struggling with this problem for a few decades, and there is stil no satisfactory solution for the long-term storage of data over centuries or even for eternity. Data accessibility. Another problem is that much scientific data is never published or deposited in data repositories such as databases. In a recent survey, data was requested from 516 studies that were published between 2 and 22 years earlier, but less than one out of five of these studies were able or willing to provide the requested data. Overall, the likelihood of retrieving data dropped by 17% each year after publication [5] Similarly, a survey of 100 datasets in Dryad found that more than half lacked the details to reproduce the research results from these studies [6] This shows the dire situation of access to scientific data that is not published or does not have enough details to be reproduced Asolution to the problem of reproducibility is the attempt to require FAIR data, that is, data that is Findable, Accessible, Interoperable, and Reusable. Data that fulfils these requirements can be used in subsequent research and thus advances science and technology [7] In other fields (eat) 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 [8 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 [1%] The term capta, which emphasizes the act of observation as constitutive, is offered as an alternative to data for visual representations in the humanities. The term data-driven is a neologism applied to an activity which is primarily compelled by data over all other factors [citation needed] Data-driven applications include data-driven programming and data-driven journalism See also [edit] « Biological data « Data mining « Data-driven learning « Computer data processing « Data modeling « Data-driven science « Computer memory « Data point « Data-driven control system « Dark data « Data preservation « Data-driven marketing « Data (computer science) « Data protection « Digital privacy « Data acquisition « Data publication « Environmental data rescue « Data analysis « Data remanence « Fieldwork « Data bank « Data science « Information engineering « Data cable « Dataset « Machine learning « Data curation « Data structure « Open data « Data domain « Data visualization « Scientific data archiving « Data element « Data warehouse « Secondary Data « Data farming « Database « Statistics « Data governance « Datasheet « Digital data « Data integrity « Data-driven programming « Data aggregation « Data maintenance « Data-driven journalism « Data management « Data-driven testing References [edi] 1.» OECD Glossary of Statistical Terms. OECD. 2008. p. 119. ISBN 978-92-64-025561 2. A "Statistical Language - What are Data?" 2. 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" 2. www.diffen.com. Retrieved 2018-12-11.

"Data vs Information - Difference and Comparison | Diffen" 2. www.diffen.com. Retrieved 2018-12-11 4.» Yonego, Joris Toonders (July 23, 2014). "Data Is the New Oil of the Digital Economy" 2. Wired — via www.wired.com. 5. ~ "Data is the new oil". July 16, 2018. Archived from the original 2 on 2018-07-16. 6. 8D rgata | Origin and meaning of data by Online Etymology Dictionary" 2. www. etymonline.com. 7. American Psychological Association (2020). "6.11". Publication Manual of the American Psychological Association: the official guide to APA style. American Psychological Association. ISBN 9781433832161 8. ~ "Joint Publication 2-0, Joint Intelligence" @ (PDF). Joint Chiefs of Staff, Joint Doctrine Publications. Department of Defense. 23 October 2013. pp. I-1. Archived from the original{@ (PDF) on 18 July 2018. Retrieved July 17, 2018. 9. ~ Akash Mitra (2011). "Classifying data for successful modeling". Archived from the original on 2017-11-07. Retrieved 2017-11-05. 10. ~ Tuomi, Ilkka (2000). "Data is more than knowledge". Journal of Management Information Systems. 6 (3): 103-117. doi:10.1080/07421222.1999.11518258 2. 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, Springer Psychology : 126 pages. ISBN 978-3-319- 15752-8 15. ~ Vines, Timothy H.; Albert, Arianne Y. K.; Andrew, Rose L.; Débarre, Florence; Bock, Dan G.; Franklin, Michelle T; Gilbert, Kimberly J.; Moore, Jean-Sébastien; Renaut, Sébastien; Rennison, Diana J. (2014-01-06). "The availability of research data declines rapidly with article age" 2. Current Biology. 24 (1): 94-97. arXiv:1312.5670 @. doi:10.1016/j.cub.2013.11.014 3. ISSN 1879-04452. PMID 2436106512". S2CID 77996622. 16. » Roche, Dominique G.; Kruuk, Loeske E. B.; Lanfear, Robert: Binning, Sandra A. (2015). "Public Data Archiving in Ecology and Evolution: How Well Are We Doing?" 2. PLOS Biology. 13 (11): €1002295. doi:10.1371/journal.pbio. 1002295 3. ISSN 1545-7885 2. PMC 4640582 @. PMID 26556502 C2. 17. Eisenstein, Michael (April 2022). "In pursuit of data immortality" 2. Nature. 604 (7904): 207-208. Bibcode:2022Natur.604..207E 2. doi:10.1038/d41586-022-00929-33. ISSN 1476-4687 2. PMID 353799892. S2CID 247954952 2. 18.» 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. 19. ~ Johanna Drucker (2011). "Humanities Approaches to Graphical Display" 2. Digital Humanities Quarterly. 005 (1). External links feat « Data is a singular noun (a detailed assessment) Alps Wiktionary, the free +®Y dictionary. media related to Data. This page was last edited on 28 July 2024, at 10:26 (UTC). Texts available under the Creative Commons Aftribution-ShareAlike License 4.0; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc. a non-profit organization. Privacy policy About Wikipedia Disclaimers Contact Wikipedia Code of Conduct Developers Statistics Cookie statement Mobile view ["y wiateoms | Poveredty @ ven | £3 TsData MA 97 languages Contents hide Article Talk Read Edit View history Tools Vv Appearance hide (Top) From Wikipedia, the free encyclopedia Text among "Scientific data" redirects here. Not to be confused with Scientific Data (journal) ©) smal — For data in computer science, see Data (computer science). For other uses, see Data (disambiguation) and Datum @) standara ™ Data documents (disambiguation). O Large Data collection In common usage, data (/ derts/, aiso Us: / deste) is a collection of discrete or — nt gen continuous values that convey information, describing the quantity, quality, fact, (7) accessibility statistics, other basic units of meaning, or simply sequences of symbols that may be [=<] conti (¢) © standarg In other fields further interpreted formally. A datum is an individual value in a collection of data == = O wise See also Data are usually organized into structures such as tables that provide additional [4] © (5) rece context and meaning, and may themselves be used as data in larger structures. Data "= "WPS Color (beta) FI may be used as variables in a computational process [12] Data may represent [4] (4) © Automatic abstract ideas or concrete measurements [°! Data are commonly used in scientific cog © Fran 0 Lon research, economics, and virtually every other form of human organizational activity. = Examples of data sets include price indices (such as the consumer price index), These are some ofthe 51 O par different types of data: unemployment rates, literacy rates, and census data. In this context, data represent Geographical, Cultural, the raw facts and figures from which useful information can be extracted Scientific, Financial, Statistical, Meteorological, Natural, Data are collected using techniques such as measurement, observation, query, or Transport analysis, and are typically represented as numbers or characters that may be further processed. Field data are data that are collected in an Part of a series on generated in the course of a controlled scientific experiment. Data are anayzed using locigues such as caclton, reasoning, dscsion, presentation, visualization, or other forms of post-analysis. Prior to —— [show] analysis, raw data (or unprocessed data) is typically cleaned: Outlers | ———————————— are removed, and obvious instrument or data entry errors are corrected ce Domains show] Data can be seen as the smallest units of factual information that can be used as a basis for calculation, reasoning, or discussion. Data can range Epistemologists Lehow] from abstract ideas to concrete measurements, including, but not limited PPE [show] to, statistics. Thematically connected data presented in some relevant _— context can be viewed as information. Contextually connected pieces of information can then be described as data insights or intelligence. The stock of insights and intelligence that accumulate over time resulting from the synthesis of data into information, can then be described as knowledge. Data has been described as "the new oil of the digital economy" ["I¥] 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 Advances in computing technologies have led to the advent of big data, which usually refers to very large quantities of data, usually at the petabyte scale. Using traditional data analysis methods and computing, working with such large (and growing) datasets is difficult, even impossible. (Theoretically speaking, infinite data would yield infinite information, which would render extracting insights or intelligence impossible.) In response, the relatively new field of data science uses machine learning (and other artificial intelligence (AI) methods that allow for efficient applications of analytic methods to big data. Etymology and terminology edt Further information: Data (word) The Latin word data is the plural of datum, "(thing) given," and the neuter past participle of dare, "to give" (€I 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" mean "transmissible and storable computer information" in 1946. The expression "data processing" was first used in 1954 [6] When "data" is used more generally as a synonym for "information, itis treated as a mass noun in singular form. This usage is common in everyday language and in technical and scientific fields such as software development and computer science. One example of this usage is the term "big data". When used more specifically to refer to the processing and analysis of sets of data, the term retains its plural form. This usage is common in the natural sciences, life sciences, social sciences, software development and computer science, and grew in popularity in the 20th and 21st centuries. Some style guides do not recognize the different meanings of the term and simply recommend the form that best suits the target audience of the guide. For example, APA style as of the 7th edition requires “data” to be treated as a plural form [7] Meaning edi] See also: DIKW pyramid = Data, information, knowledge, and wisdom are closely related concepts, but each | Sieg Sia = HE TS has its role concerning the other, and each term has its meaning. According to a Tr CA ee common view, data is collected and analyzed: data only becomes information SEE suitable for making decisions once it has been analyzed in some fashion [°! One BT fers can say that the extent to which a set of data is informative to someone depends EEE FE on the extent to which itis unexpected by that person. The amount of information [a ERI ho HI 53 contained in a data stream may be characterized by its Shannon entropy. EAE El i i El fH FH TEE E REEL Knowledge is the awareness of its environment that some entity possesses, i HE ki i 4 ki whereas data merely communicates that knowledge. For example, the entry in a EEE EAEEY database specifying the height of Mount Everest is a datum that communicates a ar SEE Er E precisely-measured value. This measurement may be included in a book along i EEE IY b with other data on Mount Everest to describe the mountain in a manner useful for | & 3 + # "2 § PR those who wish to decide on the best method to climb it. Awareness of the Adrien Auzout's "A TABLE of ® characteristics represented by this data is knowledge. the Apertures of Object-Glasses” from a 1665 article in Data are often assumed to be the least abstract concept, information the next Philosophical Transactions least, and knowledge the most abstract [%! 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