

What is Data?

Assembled by Laurie Gries for DA4ALL

Creative Commons License: CC BY-NC-SA



This license enables reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator.

Before anything else, there is always data. Data is the foundation of data science; it is the material on which all the analyses are based. In the context of data science, [there are two types of data: traditional, and big data.](#)

Traditional data is data that is structured and stored in databases which analysts can manage from one computer; it is in table format, containing numeric or text values. Actually, the term “traditional” is something we are introducing for clarity. It helps emphasize the distinction between big data and other types of data.

Big data, on the other hand, is... bigger than traditional data, and not in the trivial sense. From variety (numbers, text, but also images, audio, mobile data, etc.), to velocity (retrieved and computed in real time), to volume (measured in tera-, peta-, exa-bytes), big data is usually distributed across a network of computers.

– Excerpt from “The What, Where and How of Data for Data Science” by Iliya Valchanov



Data means information, more specifically facts, figures, measurements and amounts that we gather for analysis or reference. The term's meaning also includes descriptive information about things, plants, animals, and people. We collect and store data typically through observation.

We gather facts and numbers, which we exam and consider when trying to make, for example, a business decision. These facts and figures are data, as is electronic information that we store, and our computers or smartphone use.


Data is the plural of *datum*. However, most people tend to use the word *data* for both the plural and singular forms.

The *Cambridge Dictionary* [has the following definition of the term](#):

“Information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer.”

— Excerpt from “What is data? Definition and examples” by [Market Business News](#)







There is a naive assumption that if you see numbers in a spreadsheet, they are real somehow. But data is never this raw, truthful input, and it is never neutral. It is information that has been collected in certain ways by certain actors and institutions for certain reasons. For example, there is a comprehensive database at the US federal level of sexual assaults on college campuses – colleges are required to report it. But whether students come forward to make those reports will depend on whether the college has a climate that will support survivors. Most colleges are not doing enough, and so we have vast underreporting of those crimes. It is not that data is evil or never useful, but the numbers should never be allowed to “speak for themselves” because they don’t tell the whole story when there are power imbalances in the collection environment.

–Excerpt from Interview with Catherine D’Ignazio by [Zoë Corbyn](#), published in [The Guardian](#)







Data is a communicative action that is always embroiled in and constitutive of social relations of power. As a communicative action, data is both a form of representation and a rhetorical act with mediating power to persuade, stir emotions, incite action, and effect diverse kinds of socio-material change.

—Excerpt from “A Rhetorical Data Studies Approach to Data Advocacy” by Laurie Gries







...from the beginning, data was a rhetorical concept. Data means—and has meant for a very long time—that which is given prior to argument. As a consequence, the meaning of data must always shift with argumentative strategy and context—and with the history of both.

—Excerpt from “The History of Data-as-Rhetoric“ by Mark Carrigan







At first glance, data are apparently before the fact: they are the starting point for what we know, who we are, and how we communicate. This shared sense of starting with data often leads to an unnoticed assumption that data are transparent, that information is self-evident, the fundamental stuff of truth itself. If we're not careful, in other words, our zeal for more and more data can become a faith in their neutrality and autonomy, their objectivity. Think of the ways people talk and write about data. Data are familiarly “collected,” “entered,” “compiled,” “stored,” “processed,” “mined,” and “interpreted.” Less obvious are the ways in which the final term in this sequence — interpretation — haunts its predecessors. At a certain level the collection and management of data may be said to presuppose interpretation. “Data [do] not just exist,” Lev Manovich explains, they have to be “generated.” Data need to be imagined as data to exist and function as such, and the imagination of data entails an interpretive base.

–Excerpt from the Introduction to *“Raw Data” is an Oxymoron* by Lisa Gitelman and Virginia Jackson





Data are a form of power. Organizations own vast quantities of user information and hold lucrative data capital (Yousif, 2015), wield algorithms and data processing tools with the ability to influence emotions and culture (Gillespie, 2014; Kramer et al., 2016; Striphos, 2015), and researchers invoke data in the name of scientific objectivity while often ignoring that data are never raw but always “cooked” (Gitelman, 2013).

–Excerpt from “Critical Data Studies: An Introduction” by Andrew Iliadis and Federica Russo.



Bibliography

Carrigan, Mark. "The History of Data-as-Rhetoric." *Markcarrigan.net*, 12 September 2016.
<https://markcarrigan.net/2016/09/12/the-history-of-data-as-rhetoric/>. Accessed 6 June 2023.

Corbyn, Zoë. 'Interview with Catherine D'Ignazio: Data is Never a Raw, Truthful Input – and It is Never Neutral.' *The Guardian-U.S. Edition*. 21 March 2020. <https://www.theguardian.com/technology/2020/mar/21/catherine-dignazio-data-is-never-a-raw-truthful-input-and-it-is-never-neutral>. Accessed 6 June 2023.

Gitelman, Lisa, editor. "Raw Data" Is an Oxymoron. *The MIT Press*, 2013.

Gries, Laurie, "A Rhetorical Data Studies Approach to Data Advocacy," 2024, CC BY-NC-ND 4.0.

Iliadis, Andrew and Federica Russo. 2016. "Critical Data Studies: An Introduction." *Big Data & Society*, vol. 3, no. 2, 2016. (Creative Commons CC-BY).

Market Business News. 'What is Data? Definition and Examples.' *Marketbusinessnews.com*, n.d.
<https://marketbusinessnews.com/financial-glossary/data-definitio/>. Accessed 6 June 2023.

Vaichanoy, Iliya. "The What, Where and How of Data for Data Science." *KDnuggets.com*. 12 June 2018.
<https://www.kdnuggets.com/2018/06/what-where-how-data-science.html>. Accessed 6 June 2023.