Data Literacy

Valentin Unger, Michael Beck, and Vera Husfeldt

Definition

The competence to deal with data sovereignly, or "Data Literacy" is one of the fundamental skills of modern society. To define the term data literacy, its two semantic constituents data and literacy are examined in more detail.

1. Data are signs or symbols based on observations of the world and are constantly collected by or through us – both consciously and unconsciously (Cambridge University Press 2023b; Zins 2007, 482). Etymologically, the term data goes back to the plural of the Latin datum, which can be translated as "that which is given, a present" (Glare 2012, 532) and has been associated with computer processes since the mid-1940s (Harper 2021). Early on, data was associated with a prospective meaning: data have an impact on the future.
2. In 2000, the OECD defined literacy as follows: "The ability to understand and employ printed information in daily activities, at home, at work and in the community – to achieve one's goals and to develop one's knowledge and potential" (OECD 2000, x). This definition of literacy was criticized for several reasons, first, the focus on standardized testing. By the late 20th century, it was pointed out that a measurable continuum cannot reflect the diversity of ways in which people interact through spoken and written language (St. Clair 2012, 771). Second, the OECD's definition of literacy is also driven by the idea that the measured areas of reading, numeracy, and problem solving are directly related to economic and social progression in modern societies. By excluding other factors and focusing on one's own culturally shaped world, the model remains very simplistic and limited according to some critics (e.g. Guadalupe 2017, 334; Perry et al. 2020, 12; Sellar and Lingard 2014, 922). The increasing use of digital technologies over the past two decades has also made it necessary to refine the OECD's concept of literacy, which originally focused on printed information (OECD 2021, 5). Eventually, the original concept of

> literacy changed to a multiplicity of knowledge, skills, and values relevant to society's success (Vincent 2003, 342; Ware et al. 2016, 307).

The term data literacy emerged around the turn of the millennium and is often used in the context of digital competencies (Schüller 2020, 11–12). It is, therefore, socially relevant knowledge, skills, and attitudes necessary for dealing with and handling (digital) data sensitively, in a society characterized by permanent multifaceted information (Ridsdale et al. 2015, 4).

The model of the data–information–knowledge–wisdom pyramid states that data must be transformed into usable knowledge by sequencing its hierarchical ascent from data to information, knowledge, and wisdom. Each level can be seen as a precursor to the next, with unorganized and therefore useless data being the lowest. Data becomes usable information only by ascending the levels, which can then be used to build knowledge (Rowley 2007, 163–68). The highest level is the so-called "wisdom", which can be described as integrated, usable knowledge (Rowley 2007, 174). Appropriate use of data includes the insightful perception and production of data. Data literacy also includes knowledge, skills, and attitudes for insightful and ethical data collection, processing, and interpretation. It is also relevant when decisions are made based on the data that affect other people (Schüller 2020, 11). Schüller (2020, 23–40) has described six steps that need to be taken when encoding and decoding information: (1) "Establish a data culture", (2) "provide data", (3) "evaluate data", (4) "interpret results", (5) "interpret data", and (6) "derive actions".

Background

Data literacy is increasingly more relevant at work and in private life, which is why research is also increasingly addressing this phenomenon. "Data literacy is a hot topic" (Van Audenhove et al. 2020, 2) for various reasons:

As societies become more digital, people receive information from around the world – almost in real time (Hai et al. 2021, 25–28). The Covid-19 pandemic gave a significant boost to digitization. Many organizations had to switch to a remote mode of operation almost overnight and adapt their ways of communicating and working (Amankwah-Amoah et al. 2021, 605). Of course, this adaptation has not happened equally everywhere, but it is astonishing how quickly and widely digitalization has been adopted in order to avoid face-to-face contact. According to a study by Dingel and Neiman (2020, 1), 37 percent of jobs in the US can be done entirely from home. Schools around the world have also had to adapt to the new circumstances (Unger et al. 2022, 174).

Aside from the advantages of the increase in constantly updated information (accessibility of general education to almost everyone, worldwide information can help people form opinions and generate knowledge, the world is moving closer together, etc.), there are also problematic aspects. For example, digitalization increasingly leads to poor quality data, i.e. deliberately misinterpreted or misleadingly prepared information that can be disseminated unfiltered – especially via the internet (Schüller 2020, 11). New ethical challenges also arise, such as the increased need for awareness about privacy and data protection (Faraj et al. 2021, 4). Furthermore, the ability to detect misinformation has become more important (Nguyen 2021, 212).

"The world's most valuable resource is no longer oil, but data" (The Economist 2017). Referencing this, the slogan "data is the new oil" became a popular refrain, claiming data, especially "big data", was becoming increasingly valuable. Big data can be defined as "very large sets of data … that can only be stored, understood, and used with the help of special tools and methods" (Cambridge University Press 2023a). Data is collected and used commercially everywhere and at all times – in smartphones, online shopping, mobility services, etc. Based on big data, numerous decisions are made that directly or indirectly affect citizens. While there are positive effects attributed to big data such as improved public safety or cancer diagnostics optimized by machine learning (Bhagespur 2019; Chang 2021, 1–10), the conclusions drawn from big data are often criticized – especially in the context of screening people for economic reasons under questionable data protection rules (Brayne 2017, 980; Zuboff 2019). Nonetheless, it is relevant that people are familiar with the basics of big data mining and how big data is interpreted and acted upon.

Understanding decisions on the basis of data is becoming more complex (Van Audenhove et al. 2020, 3), while data analysis is automated via algorithms and artificial intelligence. Data literacy can enable people to understand and challenge these automated data analyses and the decisions based on them through relevant knowledge, skills, and values. Furthermore, people should also be able to recognize and evaluate algorithmic bias (Baker and Hawn 2021, 1083).

Another argument for data literacy is "misunderstandings, misperceptions, mistrust and misgivings" (Wallmann 1993, 1) regarding statistics: Data is often aggregated in order to generalize information as far as possible. However, this aggregation inevitably means that information is lost, making individual inferences very difficult (Holderness 2016, 9). One must be aware of this to interpret the omnipresent data in an informed way. Furthermore, there are often misconceptions in society, which can be traced back to a simplified or simply incorrect view of statistical terms, statistical evaluation procedures, and how they can be interpreted. For example, some people misunderstand the mean value as the highest expression of the values in a population (Yilmaz 2013, 22). Problems can also arise from improperly drawing causal conclusions from correlations (Pearl 2000, 1) due to a lack of knowledge about (1) how data are generated, (2) what conclusions are

even possible, (3) and by what process of data generation (observational or experimental) or (4) statistical analysis procedures (traditional statistical inference or modern causal inference).

This shows that data literacy subsumes competencies that help to address challenges systematically by enabling the confident handling of data. It is a transdisciplinary key competence for 21st-century society (Ludwig and Thiemann 2020, 436; Schüller 2020, 44), which is relevant in several fields. Data literacy can be an important skill when assessing the nature, quality, and credibility of different sources of information, which can vary greatly between different disciplines. Accordingly, data literacy can be adapted anywhere in a specific subject context with corresponding learning cultures. For example, data literacy is relevant for dealing with "big data" in large companies as it provides the basis for making future decisions based on this data, but also for understanding major societal issues such as climate change, mass migration, or pandemics.

Debate and criticism

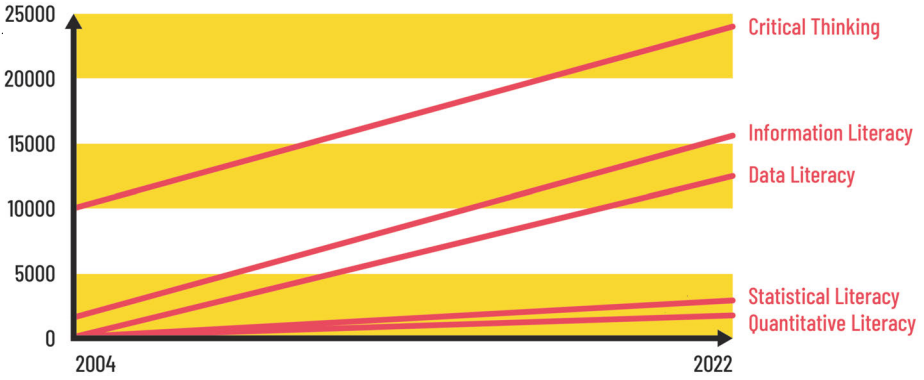
The evolution of the term data literacy itself as distinct from other closely related and sometimes synonymous terms, such as statistical literacy or information literacy, is worth highlighting. As early as 2004, Schield (2004, 8) presented an integrative model in which data literacy is seen as one of the core competencies for critical thinking. He emphasizes the technical component of data literacy, especially the use of suitable software solutions (e.g. databases, analysis, and presentation software); the term is closely related to practical aspects of data collection and analysis. In his model, data literacy can be seen as an essential sub-competence of statistical literacy, which is an essential sub-competence of information literacy.

Given this, it is surprising that Schüller (2020, 11) notes that even today, definitions and competency frameworks do not distinguish between data, statistical, and information literacy. The delineation of the term, as well as its classification as a super- and sub-category of other forms of literacy, turns out to be fundamentally problematic, as Gould (2017) demonstrates. Partly under the influence of a workshop (Oceans of Data Institute 2016), he calls for broadening the concept of statistical literacy to include aspects of data literacy (and thus follows Schield's 2004 model), although he immediately notes that definitions of data literacy include statistical literacy but also go far beyond it. According to his observations, "this is because the notions of SL [statistical literacy] that have arisen from the statistics education community are perceived by those who work in data science as falling short of what is required" (Gould 2017, 23). It remains unclear whether data literacy is a subset of statistical literacy, or vice versa. The quote suggests that data literacy may also be statistical literacy in a new (more digital) guise. This is supported

by earlier definitions of statistical literacy that are indistinguishable from current definitions of data literacy (e.g. Gal 2002, 3–4; Wallmann 1993, 1). A closer look at the levels of associated skills and competencies often reveals little difference (apart from a stronger focus on the creation and collection of data and the inclusion of buzzwords such as "big data" and "machine learning" in the term data literacy).

Schield's (2004) article gives an opportunity to explore the evolution of the term data literacy and its relative importance in the context of critical thinking, information literacy, and statistical literacy. His search of the ERIC database (Institute of Education Sciences n.d.) at the time yielded over 10,000 hits for "critical thinking" and just under 1,500 for "information literacy", whereas "quantitative literacy", "statistical literacy", and "data literacy" only received a maximum of 65 hits each (Shield does not provide individual figures for the last three aspects of competence). In 2022, on the other hand, "critical thinking", "information literacy", and "data literacy" have between 12,000 and 15,000 more entries, whereas "statistical literacy" (just under 3,000 more entries) and "quantitative literacy" (just over 1,600 more entries) have become much less established as terms (see Figure 1).

Figure 1. Occurrence of different terms in Education Resources Information Center 2022 (Institute of Education Sciences n.d.)



thinking, has grown substantially, whereas statistical literacy has seen a comparatively small increase in hits. It seems that statistical literacy is now seen as part of data literacy. Information literacy, on the other hand, continues to be considered quite separately. However, it would undoubtedly be instructive to break down the frequency of use of the terms by discipline to consider the extent to which the different terminology is simply a consequence of subject-specific differences in terminology.

However, based on the definitions and competence frameworks in use today, it remains unclear to what extent competencies in data literacy differ from sta-

tistical literacy, information literacy, or problem-solving competencies in general, especially in comparison to Schield (2004). It seems problematic that most definitions tend not to focus explicitly on information technologies, but data literacy without reference to digitalization seems to have no added value.

A critical problem in defining the term seems to be that in the course of digitalization, the distinction between "data" and "information" now seems obsolete (Schüller 2020, 11), as information that cannot be stored in the form of data has become almost inconceivable. Schüller (2020, 11) makes a claim for the "triumph" of data literacy: she assumes that it is a fashionable term resulting from the fact that jobs with the title "data engineer" or "data scientist" pay higher salaries than comparable titles such as "statistician" or "IT specialist" (unfortunately, she does not support this assumption with actual salary analyses).

Current forms of implementation in higher education

Data literacy has significant role in today's educational landscape. When it comes to curricula for schools or universities, the significance of digital literacy and data literacy, in particular, is undeniable (Bandtel et al. 2021, 396).

As part of a European Erasmus+-funded program, an international research group was able to show that it is essential to support educators in acquiring data literacy skills in order to enable them to deal with data and base their decisions on data (Papamitsiou et al. 2021, 21). Using an extensive literature review of international databases from academic publishers, the authors identified different courses that include educational data literacy training in higher education or professional development (Papamitsiou et al. 2021, 8). Some of them are briefly presented here. The Norwegian University of Science and Technology offers a course on digital literacy and smart learning that aims to develop a thoughtful relation to the use of digital services in various teaching and learning processes (Norwegian University of Science and Technology 2020). A 15-week interactive hands-on course on data literacy is given online by the RETAIN Center of Excellence at Newberry College for educators from South Carolina (n.d.). Edx, a consortium of universities from all over the world that offer online courses, provides various courses for teachers. One of them is a six-week course on learning analytics that addresses schoolteachers who want to improve their teaching through valuable data-driven insights during three to four hours a week (Edx 2023a). Another course focuses on big data in education (Edx 2023b), and again another one focuses on wise action in connection with data (Edx 2023c).

In addition, there are other examples from around the world. In Tanzania, the Tanzania data lab (dLab), is fostering data literacy to engage communities to address issues like the HIV/AIDS epidemic but also gender inequality and economic

growth by using data. Together with the University of Dar es Salaam, College of Information and Communication Technology (UDSM CoICT) they helped to start the first East African Masters in Data Science (dLab 2021).

As part of the "Swiss Digital Skills Academy", a project funded by Swissuniversities, the association of higher education institutions in Switzerland, led by the École polytechnique fédérale de Lausanne, the "Develop Data Literacy" project was launched in 2021, with four universities of applied sciences participating as a core group. At the University of Applied Sciences of the Grisons, two courses were held where information science students first studied the theoretical foundations of data literacy and subsequently developed Open Educational Resources on various topics of data literacy. Open Educational Resources are freely accessible learning opportunities that are mostly made available digitally (UNESCO & Commonwealth of Learning 2011, v). At the St.Gallen University of Teacher Education, an Open Educational Resource is currently being developed in cooperation with educational researchers and in-service teachers as part of the same project, which will enable in-service and prospective teachers to receive further training regarding computer-based educational technologies as an example.

The contribution shows that data literacy – despite existing terminological imprecision – is a transdisciplinary key competence of our time, which should be fostered more actively in university teaching.

References

Amankwah-Amoah, Joseph, Zaheer Khan, Geoffrey Wood, and Gary Knight. 2021. COVID-19 and digitalization: The great acceleration. Journal of Business Research 136: 602–11.

Baker, Ryan S., and Aaron Hawn. 2021. Algorithmic bias in education. International Journal of Artificial Intelligence in Education 32: 1052–92.

Bandtel, Matthias, Leonie Kauz, and Natalia Weißker. 2021. Data Literacy Education für Studierende aller Fächer. Kompetenzziele, curriculare Integration und didaktische Ausgestaltung interdisziplinärer Lehr-Lern-Angebote. In Digitalisierung in Studium und Lehre gemeinsam gestalten, eds. Geschäftsstelle beim Stifterverband, 395–412. Wiesbaden: Springer.

Bhagespur, Kiran. 2019. Data is the new oil – And that's a good thing. Forbes. Available from https://www.forbes.com/sites/forbestechcouncil/2019/11/15/data-is-the-new-oil-and-thats-a-good-thing/?sh=796640cb7304.

Brayne, Sarah. 2017. Big data surveillance: The case of policing. American Sociological Review 82 (5): 977–1008.

Cambridge University Press, ed. 2023a. Big data. Available from https://dictionary.cambridge.org/dictionary/english/big-data.

Cambridge University Press, ed. 2023b. Data. Available from https://dictionary.cambridge.org/dictionary/english/data.

Chang, Victor. 2021. An ethical framework for big data and smart cities. Technological Forecasting & Social Change 165 (120559): 1–11.

Dingel, Jonathan, and Brent Neiman. 2020. How many jobs can be done at home? Journal of Public Economics 189 (104235): 1–8.

dLab. 2021. The Tanzania data lab (dLab) and the UDSM. Available from https://dlab.or.tz/the-tanzania-data-lab-dlab-and-the-university-of-dar-es-salaam-launch-the-first-masters-of-science-in-data-science.

Edx. 2023a. Analytics for the classroom teacher. Available from https://www.edx.org/course/analytics-for-the-classroom-teacher.

Edx. 2023b. Big data and education. Available from https://www.edx.org/course/big-data-and-education.

Edx. 2023c. Introduction to data wise: A collaborative process to improve learning & teaching. Available from https://www.edx.org/course/introduction-to-data-wise-a-collaborative-process.

Faraj, Samer, Wadih Renno, and Anand Bhardwaj. 2021. Unto the breach: What the COVID-19 pandemic exposes about digitalization. Information and Organization 31 (1): 1–7.

Gal, Iddo. 2002. Adults' statistical literacy: Meanings, components, responsibilities. International Statistical Review 70 (1): 1–51.

Glare, Peter G. W., ed. 2012. Oxford Latin dictionary. 2nd edition. Vol. A–L. Oxford: Oxford University Press.

Gould, Robert. 2017. Data literacy is statistical literacy. Statistics Education Research Journal 16 (1): 22–25.

Guadalupe, César. 2017. Standardisation and diversity in international assessments: Barking up the wrong tree? Critical Studies in Education 58 (3): 326–40.

Hai, Thanh Nguyen, Quang Nguyen Van, and Mai Nguyen Thi Tuyet. 2021. Digital transformation: Opportunities and challenges for leaders in the emerging countries in response to Covid-19 pandemic. Emerging Science Journal 5: 21–36.

Harper, Douglas. 2021. Data. Available from https://www.etymonline.com/word/data.

Holderness, Clifford G. 2016. Problems using aggregate data to infer individual behavior: Evidence from law, finance, and ownership concentration. Critical Finance Review 5 (1): 1–40.

Institute of Education Sciences. n.d. ERIC. Available from https://eric.ed.gov.

Ludwig, Thomas, and Hannes Thiemann. 2020. Datenkompetenz – Data Literacy. Informatik Spektrum 43 (6): 436–39.

Nguyen, Dennis. 2021. Mediatisation and datafication in the global COVID-19 pandemic: On the urgency of data literacy. Media International Australia 178 (1): 210–14.

Norwegian University of Science and Technology. 2020. Digital literacy – Smart learning. Available from https://www.ntnu.edu/studies/courses/SOS2020/2017/1#tab=omEmnet.

Oceans of Data Institute, ed. 2016. Building global interest in data literacy: A dialogue. Workshop report. Available from http://oceansofdata.org/our-work/building-global-interest-data-literacy-dialogue-workshop-report.

OECD [Organisation for Economic Cooperation and Development], ed. 2000. Literacy in the information age. Final report of the international adult literacy survey. Paris: OECD Publications Service. Available from https://www.oecd.org/education/skills-beyond-school/41529765.pdf.

OECD [Organisation for Economic Cooperation and Development], ed. 2021. 21st-century readers: Developing literacy skills in a digital world. Paris: OECD Publishing. Available from https://www.oecd-ilibrary.org/education/21st-century-readers_a83d84cb-en.

Papamitsiou, Zacharoula, Michail E. Filippakis, Marilena Poulou, Demetrios Sampson, Dirk Ifenthaler, and Michail Giannakos. 2021. Towards an educational data literacy framework: Enhancing the profiles of instructional designers and e-tutors of online and blended courses with new competences. Smart Learning Environments 8 (18): 1–26.

Pearl, Judea. 2000. Causality: Models, reasoning, and inference. Cambridge: Cambridge University Press.

Perry, Kristen, Donita Shaw, and Sara Saberimoghaddam. 2020. Literacy practices and the Programme for the International Assessment of Adult Competencies (PIAAC): A conceptual critique. International Review of Education 66 (1): 9–28.

RETAIN Center of Excellence at Newberry College. n.d. Data literacy 01. Available from http://artofeducating.com/online-courses.

Ridsdale, Chantel, James Rothwell, Mike Smit, Hossam Ali-Hassan, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin, and Brad Wuetherick. 2015. Strategies and best practices for data literacy education knowledge synthesis report. Halifax: Dalhousie University. Available from http://rgdoi.net/10.13140/RG.2.1.1922.5044.

Rowley, Jennifer. 2007. The wisdom hierarchy: Representations of the DIKW hierarchy. Journal of Information Science 33 (2): 163–80.

Schield, Milo. 2004. Information literacy, statistical literacy and data literacy. IASSIST Quarterly 28 (2/3): 6–11.

Schüller, Katharina. 2020. Future skills: A framework for data literacy. Competence framework and research report. Working Paper No. 53. Berlin: Hochschulforum Digitalisierung.

Sellar, Sam, and Bob Lingard. 2014. The OECD and the expansion of PISA: New global modes of governance in education. British Educational Research Journal 40 (6): 917–36.

St. Clair, Ralf. 2012. The limits of levels: Understanding the International Adult Literacy Surveys (IALS). International Review of Education 58 (6): 759–76.

The Economist, ed. 2017. The world's most valuable resource is no longer oil, but data. The Economist. Available from https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

UNESCO and Commonwealth of Learning, ed. 2011. Guidelines for open educational resources (OER) in higher education. Paris: UNSECO.

Unger, Valentin, Laura Schmidberger, and Albrecht Wacker. 2022. "Ja, der Fernunterricht hat sich deutlich verbessert…!" Befunde einer Schüler*innenbefragung zum schulisch angeleiteten Lernen zu Hause während der zweiten Schulschliessungen in Deutschland. In Schule in Distanz – Kindheit in Krise. Auswirkungen der Covid-19 Pandemie auf Wohlbefinden und Lebensbedingungen von Kindern und Jugendlichen, eds. Jürgen Budde, Drorit Lengyel, Caroline Böning, Carolina Claus, Nora Weuster, Katharina Doden, and Tobias Schroedler, 173–96. Wiesbaden: Springer VS.

Van Audenhove, Leo, Wendy Van den Broeck, and Ilse Mariën. 2020. Data literacy and education: Introduction and the challenges for our field. Journal of Media Literacy Education 12 (3): 1–5.

Vincent, David. 2003. Literacy literacy. Interchange 34 (2): 341–57.

Wallmann, Katherine K. 1993. Enhancing statistical literacy: Enriching our society. Journal of the American Statistical Association 88 (421): 1–8.

Ware, Paige, Richard Kern, and Mark Warschauer. 2016. The development of digital literacies. In Handbook of second and foreign language writing, eds. Rosa M. Manchón and Paul Kei Matsuda. Boston: De Gruyter, 307–28.

Yilmaz, Zuhal. 2013. Usage of tinker plots to address and remediate 6th grade students' misconceptions about mean and median. The Anthropologist 16 (1–2): 21–29.

Zins, Chaim. 2007. Conceptual approaches for defining data, information, and knowledge. Journal of the American Society for Information Science and Technology 58 (4): 479–93.

Zuboff, Shoshana. 2019. The age of surveillance capitalism: The fight for a human future at the new frontier of power. London: Profile Books.