

Approaching Data Literacy Education

Tibor Koltay

*Retired college professor from Eszterházy Károly Catholic University
Jászberény Campus
Institute of Learning Technology*

Introduction

This paper considers data literacy (DL) from various perspectives, some of which are defined by different literacies. To support related pedagogical tasks, we present a selection of contributions from various disciplines that outline the fundamental characteristics of DL and we conclude by mentioning some of the potential applications of AI literacy based on it.

Presenting educational approaches and programmes requires careful and detailed theoretical underpinning. However, given the diversity of views and approaches, it is important to emphasise that the chronological order in which we present our topics is secondary.

We also consider DL from various perspectives, some of which are influenced by different literacies. To support related pedagogical tasks, we present a selection of contributions from various disciplines that outline the fundamental characteristics of DL. Finally, we conclude by mentioning some of the potential applications of AI literacy based on DL.

Background

Technological advances today have created both opportunities and threats to the free flow of information in society. We are witnessing a similar evolution to that experienced with data and DL. Due to the applicability of different types of data, and the need to use them properly, is similar to what we have seen with information literacy (IL) and other literacies. The increasing applicability of different types of data and the need to use them properly bear many similarities to IL and other types of literacy.

It is well-known that presenting educational approaches and programmes requires careful and detailed theoretical underpinning. However, it is important to emphasise that, given the diversity of views and approaches, the chronological order in which we present our topics is less important. Taking this into account

to discover and identify relevant literature, a heuristic review was conducted to identify and describe DL, because we believe that there is not always a definitive, uncontroversial, and consistent set of criteria for identifying and using the best literature. As Boell and Cecez-Kecmanovic (2010) emphasise it, such reviews constantly iterate and reinterpret the literature to achieve a “deeper and more comprehensive understanding of the relevant publications” (p. 130). This decision was made because, unlike systematic reviews, we believe that they do not aim to be exhaustive, but rather to be transparent, replicable and unbiased. This is because they operate within an iterative cycle that may involve re-evaluating some of the articles identified.

The nature of DL

Data consists of declarative statements of facts that reflect reality. These statements describe physical things with attributes such as identity, location, dimensions, current status and colour. We experience this in the physical world. Any dataset has context because a string of numbers is meaningless in isolation. Moreover, we can collect and store data, displayed in raw or processed forms unless we decide that it is irrelevant (Grieves, 2024).

Data can be used in a variety of ways, so one of the most important aspects of DL is its role in supporting data citizenship (Carmi et al. 2020).

Nevertheless, this paper on research DL, which is particularly important in higher education and is often treated in a relatively well-defined sense, i.e. limited to its relation to scientific research, in particular to research data and its management.

Existing datasets are vital for producing and communicating new knowledge, as they enable experiments to be both robust and reproducible. With this in mind, Corrall (2019) identifies components of data management that includes proper professional conduct, data-driven decision-making, ethics, and data visualisation.

Due to the novelty, breadth and depth of investigation DL, this paper places a strong emphasis on theoretical issues, while also considering possible approaches to teaching related subjects. However, we must heed the warning that “datasets do not speak for themselves, as they require context for analysis and interpretation” (Gebre & Morales, 2020, p. 1).

We should be aware of the fact that presenting educational approaches and programmes requires careful and detailed theoretical underpinning. However, it is important to emphasise that, given the diversity of views and approaches, the chronological order in which we present our topics is secondary.

The complexity of DL

Today technological advances have created both opportunities and threats to the free flow of information in society. We are witnessing a similar evolution to that experienced with data, which is leading to the emergence of DL.

Unsurprisingly, it is almost impossible to discuss digital literacy (DL) without first defining literacy as a concept that originally encompassed reading and writing textual messages. However, its meaning has constantly evolved, primarily due to the increasing significance of digital technologies and the resulting socio-technological changes and challenges arising from the convergence of media, telecommunications, and information and communication technologies (Livingstone et al., 2008).

Working with data

Working with data starts with reading and understanding datasets, and recognising what aspects of the world that they represent. Working with data involves creating, acquiring, cleaning and managing it. In harmony with this, we are obliged to accept standard data management workflows. Beyond this, we need to pay attention to distinguishing between efficacy and its perceived level of the knowledge and skills necessary for performing related tasks correctly and successfully (Bandura, 1997). Moreover, it is technology agnostic and based on abstraction with the aim of improving logical-mathematical skills and reasoning.

One of the goals of fostering contemporary DL demands training for it is part of human behaviour that solves problems at multiple levels of abstraction using both human and computer algorithms (Wing, 2006). It enables solving problems, designing systems, and understanding human behaviour (Gretter & Yadav, 2016).

In line with the above-mentioned thinking skills, computation can be seen as a core skill for managing data, therefore we need include it in educational models, as it is beneficial for achieving better academic performance (Federer et al., 2020).

The need for teaching DL

Higher-order knowledge is characterised by coherent and explanatory potential, which is not only a feature of DL, but also of IL. As Bawden and Robinson (2016) state it, this relationship becomes apparent when we consider a widely

accepted definition of IL, which encompasses information in all its forms, including data.

In this context, raising awareness of IL is also necessary, but it is often unclear how and to what extent this occurs. However, it is clear that replacing IL with DL alone is insufficient for best practice in DL education (Hunt, 2005).

According to Papamitsiou et al. (2021), the process of making datasets usable involves reading and interpretation through inclusion and exclusion, which are essentially subjective judgements based on norms and standards regarding data in different contexts (Neff et al., 2017). This highlights the paramount importance of educational DL, which forms the backbone of DL education and involves the skills of collecting, managing, analysing, understanding and interpreting educational data. This process should also adhere to ethical, meaningful and critical behaviour. To accomplish this, we need to determine the value of a particular piece of data, not forgetting about the presence of societal factors that might influence the value of that data (Abner, 2020).

Approaches for teaching

Data can be a force for managing and improving education. This demands recognising the quality of the data, organising it and making decisions. It is an activity that helps capturing and organising ideas and translating the information into meaningful action (Earl, & Katz, 2002). This also means that the starting point is informing our students about the fact data is not evidence in itself, but rather a multifaceted entity that can substantiate valid arguments (Owens, 2011).

As Van Audenhove et al. (2020) underline it, we need to recognise DL as a transversal competence. On the other hand, we can also agree with Munasinghe and Svirsky (2021), who remind us that part of our students' education includes not only data analysis, but the recognition and application of quantitative algorithms.

The curriculum framework developed by Atenas et al. (2023) presents a curriculum framework for collaborative practices according to their level (beginner, intermediate and advanced) and proficiency. They have built a curriculum framework for collaborative practices at initial, intermediate, advanced levels. It provides tools for differentiating between activities according to their suitability for all students, undergraduates and graduates according to their suitability for these students. An example of an activity for advanced undergraduates is engaging them in political and legal considerations and debates at local and global levels, when asking them to analyse datasets.

In addition to the above Deja et al. (2021) have found some notable DL variables:

- Knowing, how to use spreadsheets and software, as well as validating graphical and tabular representations of data.
- Understanding, how to carry out mathematical calculations and how to ensure the instrumentation and technology to collect and store data.

Teachers, who may have different backgrounds and levels of knowledge, must act as both producers and consumers of data products (Carlson et al., 2011). For instance, information literacy and DL are similar not only in their definitions, but also in the way they approach and solve related problems.

In parallel with IL, students should become data literate and familiar with the basics of statistical literacy, even if its application seems to be overemphasised in some cases. Nevertheless, the latter literacies still rely on the ability “to read and interpret summary statistics in everyday media: in graphs, tables, statements, surveys and studies” (Schield, 2010, p. 135). When we look at it, the teaching of DL is different because the latter focuses less on literature-based attributes, but aims to pay attention to their functional ability in collecting, processing, managing, evaluating and using data (Qin & D'Ignazio, 2010).

New ways of using information and data rely on our students’ previous experiences. This is inevitable because learning about a subject motivates students to be aware of new ways of using information in parallel to developing new understandings of the subjects they are studying (Maybee, & Zilinski, 2015).

Due to the relative newness and diversity of DL concepts, coupled to a rather small amount of teaching material available, it is not easy to develop appropriate curricula. One of the challenges is that translating formal data into personal contexts when applying any concept or approach to teaching DL can be extremely complex (Bowler et al., 2017).

The concept of data infrastructure literacy is also a starting point for the argument that learners should be allowed to “account for participating in the wider socio-technical infrastructures through which data is created, stored, and analysed” (Gray et al., 2018, p. 8).

Obviously, this requires “adherence to data standards, disciplinary knowledge and practices, curriculum knowledge, pedagogical content knowledge, and an understanding of how children learn.” (Gummer & Mandinach, 2015, p. 2). They identify problems and frame questions, as well as providing data sources for design and implementation. This enables them to understand how to analyse

data and draw inferences and conclusions from it. As Mandinach and Gummer (2016) emphasise there are three types of data user, i.e. data literate, broadly literate, and those who lack basic training.

As Raffaghelli and Stewart (2020) warn us, teacher education for fostering DL may be lacking, so it remains important to equip it with the necessary knowledge by being aware of the possibility of poor interpretation or ill-informed decisions. The problem is that educators' DL tends to cover fragmented sets of skills, mostly focusing on data management and addressing technical skills, with less emphasis on critical, ethical and personal approaches to data in education. It has also been argued that teachers are inundated with data, and their ability to use data productively and responsibly is a salient and complex skill.

Sánchez-Cruzado et al. (2021) underline that it is “essential to focus on the teachers themselves, to identify and address their main weaknesses, and to help them achieve an adequate level of DL in order to successfully face the new educational paradigm (p. 26)”.

As Henderson and Corry (2020) point out, conducting targeted inquiry is beneficial for informing training practices for teachers and educational leaders. Unfortunately, less emphasis has been placed on addressing the issues of datafication. However, they also acknowledge that many of the related DL concepts have become more concrete, albeit with ongoing disagreements.

A brief outlook for the future

As Olari and Romeike (2021) point out, there is a new development that offers the possibility of linking DL to artificial intelligence. This fusion is paving the way for a new type of literacy: AI literacy.

AI literacy is on its way to become a widely discussed topic, among others, aiming at empowering the players of education by being able to critically evaluate AI and collaborate effectively with AI systems, and utilize as tools across diverse contexts, even when we know that practice depends on age, level of education, and individual background. This is reality thanks to the increased focus on machine learning, fuelled by rule-based systems or autonomous decision-making, without the need for direct programming. All these may influence K–12 education (Mannila et al., 2025).

Conclusion

DL and AI literacy open up new ways, particularly in terms of their educational applications. Understanding and applying DL is relatively widespread, while incorporating AI literacy into teaching is still in its infancy. Both are promising new approaches to teaching, but we need to realise that their thorough investigation and systematic application may require additional research in a number of areas of expertise.

References

Abner, K. (2020). Data literacy as digital humanities literacy: Exploration of threshold concepts. In M. Brooks, M. Hubbard, J. Perkins, & J. Russel (Eds.). *Literacies in a Digital Humanities Context: A dh+ lib Special Issue* (pp. 19–22). dh+ lib.

Atenas, J., Havemann, L., Rodés, V., & Podetti, M. (2023). Critical data literacy in praxis: An open education approach for academic development. *Edutec. Revista Electrónica de Tecnología Educativa*, (85), 49–67.
<https://doi.org/10.21556/edutec.2023.85.2851>

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman and Company.

Bawden, D., & Robinson, L. (2016). Information and the gaining of understanding. *Journal of Information Science*, 42(3), 294–299.
<https://doi.org/10.1177/0165551515621691>

Boell, S. K., & Cecez-Kecmanovic, D. (2010). Literature reviews and the hermeneutic circle. *Australian Academic & Research Libraries*, 41(2), 129–144.
<https://doi.org/10.1080/00048623.2010.10721450>

Bowler, L., A., Acker, W. Jeng, & Chi, Y. (2017). ‘It lives all around us’: Aspects of data literacy in Teen’s Lives. *Proceedings of the Association for Information Science and Technology*, 54(1), 27–35.
<https://doi.org/10.1002/pra2.2017.14505401004>

Carlson, J., Fosmire, M., Miller, C. C. & Nelson, M. S. (2011). Determining data information literacy needs: A study of students and research faculty. portal: *Libraries and the Academy*, 11(2), 629–657.
<https://doi.org/10.1353/pla.2011.0022>

Carmi, E., Yates, S.J., Lockley, E. & Pawluczuk, A. (2020). Data citizenship: rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review* 9(2).
<https://doi.org/10.14763/2020.2.1481>

Corrall, S. Repositioning data literacy as a mission-critical competence. 2019. <http://d-scholarship.pitt.edu/id/eprint/36975>

Davies, A., Fidler, D., & Gorbis, M. (2011). Future work skills 2020. Institute for the Future for University of Phoenix Research Institute, 540. <http://hdl.voced.edu.au/10707/194830>

Deja, M., Januszko-Szakiel, A., Korycińska, P., & Deja, P. (2021). The impact of basic data literacy skills on work-related empowerment: The alumni perspective. *College & Research Libraries*, 82(5), 708–728. <https://doi.org/10.5860/crl.82.5.708>

D'Ignazio, C. (2017). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, 23(1), 6–18. <https://doi.org/10.1075/ijd.23.1.03dig>

Earl, L., & Katz, S. (2002). Leading schools in a data-rich world. In K. Leithwood, & P. Hallinger (Eds.), *Second international handbook of leadership and administration* (pp. 1003–1022). Dordrecht: Kluwer Academic. https://doi.org/10.1007/978-94-010-0375-9_34

Federer, L., Clarke, S. C., & Zaringhalam, M. (2020). Developing the librarian workforce for Data Science and Open Science. Bethesda, MD.: National Library of Medicine <https://osf.io/uycax/> <https://doi.org/10.31219/osf.io/uycax>

Gebre, E. H., & Morales, E. (2020). How ‘accessible’ is open data? Analysis of context-related information and users’ comments in open datasets. *Information and Learning Sciences*, 121(1/2), 19–36. <https://doi.org/10.1108/ILS-08-2019-0086>

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, 5(2), 1–13. <https://doi.org/10.1177%2F2053951718786316>

Gretter, S., & Yadav, A. (2016). Computational thinking and media & information literacy: An integrated approach to teaching twenty-first century skills. *TechTrends*, 60(5), 510–516. <https://doi.org/10.1007/s11528-016-0098-4>

Grieves, M. (2024). DIKW as a general and digital twin action framework. *Data, Information, Knowledge, and Wisdom. Knowledge*, 4, 120–140. <https://doi.org/10.3390/knowledge4020007>

Gummer, E. S., & Mandinach, E. B. (2015). Building a conceptual framework for data literacy. *Teachers College Record*, 117(4), 1–22. <https://doi.org/10.1177/016146811511700401>

Henderson, J., & Corry, M. (2020). Data literacy training and use for educational professionals. *Journal of Research in Innovative Teaching & Learning*, 14(2), 232–244. <https://doi.org/10.1108/JRIT-11-2019-0074>

Hunt, K. (2005). The challenges of integrating data literacy into the curriculum in an undergraduate institution. *IASSIST Quarterly*, 28(2-3), 12–15.
<https://doi.org/10.29173/ij791>

Livingstone, S., Van Couvering, E., & Thumin, N. (2008) Converging traditions of research on media and information literacies: disciplinary and methodological issues. In Coiro J, Mandinach, E. B. & Gummer, E. S. (2016). Every teacher should succeed with data literacy. *The Phi Delta Kappan*, 97(8), 43–46.
<https://doi.org/10.1177/0031721716647018>

Mannila, L., Hallström, J., Nordlöf, C., Heintz, F., Sperling, K., & Stenliden, L. (2025). Framing AI Literacy for K-12 Education: Insights from Multi-Perspective and International Stakeholders. In Proceedings of the 27th Australasian Computing Education Conference (pp. 85–94).
<https://doi.org/10.1145/3716640.3716650>

Maybee, C. & Zilinski, L. (2015). Data informed learning: A next phase data literacy framework for higher education. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community. American Society for Information Science. pp. 108–111.

Munasinghe, T., & Svirsky, A. (2021). Engaging Students in Data Literacy: Lessons Learned from Data Intensive Classrooms. In Companion Publication of the 13th ACM Web Science Conference 2021 (pp. 40–43)
<https://doi.org/10.1145/3462741.3466665>

Neff, G., Tanweer, A., Fiore-Gartland, B., Osburn, L. (2017). Critique and contribute: a practice-based framework for improving critical data studies and data science. *Big Data*, 5(2), 85–97.
<https://doi.org/10.1089/big.2016.0050>

Olari, V., & Romeike, R. (2021). Addressing AI and Data Literacy in Teacher Education: A Review of Existing Educational Frameworks. ACM International Conference Proceeding Series. ACM International Conference *Proceeding Series*. (17), 1–2.
<https://doi.org/10.1145/3481312.3481351>

Owens, T. (2011). Defining data for humanists: Text, artefact, information or evidence. *Journal of Digital Humanities*, 1(1), 6–8.

Papamitsiou, Z., Filippakis, M. E., Poulou, M., Sampson, D., Ifenthaler, D., & Giannakos, M. (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.
<https://doi.org/10.1186/s40561-021-00163-w>

Qin, J. & D'Ignazio, J. (2010). Lessons Learned from a Two-year Experience in Science Data Literacy Education, Proceedings of the 31st Annual IATUL

Conference, June 20–24, 2010. <http://docs.lib.psu.edu/iatul2010/conf/day2/5>

Raffaghelli, J. E., & Stewart, B. (2020). Centering complexity in ‘educators’ data literacy’ to support future practices in faculty development: A systematic review of the literature. *Teaching in Higher Education*, 25(4), 435–455. <https://doi.org/10.1080/13562517.2019.1696301>

Sánchez-Cruzado, C., Santiago Campión, R., & Sánchez-Compañía, M. (2021). Teacher digital literacy: The indisputable challenge after COVID-19. *Sustainability*, 13(4), 1858. <https://doi.org/10.3390/su13041858>

Schield, M. (2010). Assessing statistical literacy: Take CARE. Assessment methods in statistical education: An international perspective, 133–152. <https://doi.org/10.1002/9780470710470.ch11>

Van Audenhove, L., Van den Broeck, W., & Mariën, I. (2020). Data literacy and education: Introduction and the challenges for our field. *Journal of Media Literacy Education*, 12(3), —5. <https://doi.org/10.23860/JMLE-2020-12-3-1>

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35. <https://doi.org/10.1145/1118178.1118215>

Abstract

With the emergence of information literacy (IL) and media literacy, we need to learn that there are other, useful literacies, such as data literacy (DL) and artificial intelligence (AI) literacy. Both offer new directions that are on their way to becoming an important part of the skills, abilities and competencies required for work with data. Our review shows that the idea of a “data-driven world” is becoming more than a utopia, because it is increasingly becoming a reality in secondary and tertiary education, providing insight into the attitudes and competencies, and is now increasingly becoming part of research data education.