Katharina Schüller

Towards a Framework for Data and AI Literacy

Katharina Schüller[1,2]

[1] STAT-UP Statistical Consulting & Data Science GmbH ; Munich/Germany ; katharina.schueller@stat-up.com

[2] German Statistical Society

Abstract

Statistical literacy, data literacy, information literacy, and AI literacy are terms that are often mentioned as essential competencies in relation to digitization. Obviously, the delimitation of the terms is by no means conclusively clarified. Furthermore, there is the danger of definitions being deliberately extended because “data” and “AI” are becoming more and more fashionable.

The framework for data and AI literacy proposed in this contribution aims to create a common understanding of data and AI literacy so that it can be systematically included into curricula and educational standards of schools, teacher training, higher education, and further education. The purpose of the framework is to facilitate that the respective set of skills and competencies are widely taught as a transdisciplinary competence across all subjects from three perspectives (application-oriented, technical-methodological, and socio-cultural) and can be acquired by all people so that every individual, and our society, will be enabled to deal with data and AI in a conscious and ethically sound manner.

Furthermore, the framework is intended to serve as a reference for data and AI literacy programs for extracurricular and vocational training, to enable lifelong learning of data and AI literacy. Finally, the standard should be interoperable with existing quality standards for statistics (e.g., the UN Fundamental Principles) and with the IEEE Standard for Digital Intelligence. Given that, such a standard can serve as a basis for eventually developing a convergent, coherent data literacy assessment, measurement and impact evaluation framework for the national, regional, and global levels that can help to fill the empirical evidence gap on outcomes from data literacy interventions and programs.

Keywords:

Statistical Literacy; Digital Literacy; Vocational Training; Standard; Education

1. Introduction

Data and AI literacy is often misunderstood as a set of technical skills, limited to data management and analysis and to the development and application of algorithms. However, data and AI literacy as a future skill of the 21st century serves to promote autonomy in a modern world shaped by data and its application as well as new technologies like AI and is therefore important for all people - not only for specialists. Data and AI literacy therefore encompasses a set of cognitive, meta-cognitive, affective, and socio-emotional competencies, which are grounded in universal moral values and enable individuals to face

the challenges of digital life and adapt to its demands. There is also a need for a standardized framework to capture at least a minimum set of foundational and cross-cutting data literacy competencies relevant for an individual, an organisation, or a system. This will help to identify clear data literacy needs, support the effective targeting of policies and programs to enable data literacy, and provide a benchmark to assess the impact of such efforts. Developing a suitable standard and measures that allow for a fit-for-purpose tracking of progress on data literacy will bolster the case for investing in data literacy beyond ad-hoc programming.

2. Background and Methodology

Data represents digital images of real phenomena, objects, and processes of a system, therefore abstracting the real world. By cleaning up and linking data, information is created. Organising, i.e., analysing data creates knowledge. Finally, applied knowledge that is meaningfully interpreted and used, constitutes wisdom or – as the French philosopher Michel Foucault calls it – power [9]. To obtain a framework for data literacy, this process of creating value from data needs to be described and the respective competencies that facilitate the process flow must be derived and structured.

Heidrich et al. [10] present a competence matrix based on Ridsdale [13] with five areas of competence, individual competences, and an evaluation according to degree of difficulty. The competence framework on digital competences of the EU Science Hub (DigComp 2.0), on the other hand, places data and information literacy as one of four sub-competences of digital literacy [7]. In this context, the terms *data and information literacy* are not differentiated from each other, and some skills that are defined as core competencies listed by Heidrich et al. [10] are missing.

In contrast, the ProCivicStat project developed a far more extensive conceptual model [11]. It relates data literacy or statistical literacy to contextual knowledge and differentiates more strongly between data forms that occur in the scientific environment and data forms, with which the citizen comes into contact. This creates the connection to the question of which educational purpose (imparting skills for the responsible educational citizen or for specialist disciplines) the framework should reflect.

Data literacy skills should enable their critical adopter to address and solve real problems, that means, context is important. Unlike problems in the technical environment, in which data is used, for example, to ensure the quality of production processes, problems with a societal dimension require additional skills and attitudes (data ethics, measurability of phenomena, classification of results in contextual knowledge, derivation of recommendations for action). How to collect or procure, evaluate, apply, and interpret data appropriately for the respective application must be systematically learned and practised.

"To collect, manage, evaluate and apply data in a critical way", in Ridsdale's words, defines a process that requires data literacy to master [13]. This process is often represented as a cycle. Specific tasks can be arranged along the different steps of the process cycle. Data literacy can then be seen as a set of competences or, in other words, a cluster of effective behaviours and attitudes towards the fulfilment of a certain task. Therefore, Schüller defines the term data literacy as follows [15]: *“Data Literacy is the cluster of all efficient behaviours and attitudes for the effective execution of all process steps for creating value or making decisions from data.”*

“Effective execution” is one that is target-oriented with respect to the target of the whole process, which is creating value or knowledge from data, i.e., it answers the question of what needs to be done to achieve this goal ("doing the right things"). “Efficient behaviour and attitudes” allow to achieve this target with the least possible use of resources, answering the question of how the target can be achieved in the best possible way ("doing things right").

The explicit highlighting of attitudes is in accordance with Seidl et al. who emphasize the importance of attitudes, values, and ethics for key competences [16].

Competence frameworks often follow a thematic order; for example, the OECD Competence Framework groups various key competences into competence categories [12]. The European eCompetence Framework (e-CF) [6] arranges competences in fields of competence, defining levels in each case in accordance with the European Qualifications Framework for Lifelong Learning (EQR) [5] and gives examples of knowledge and abilities or skills. In addition to knowledge (theoretical and factual knowledge) and skills (cognitive and practical skills or abilities), the EQR distinguishes between actual competences (responsibility and autonomy). The German Qualifications Framework, on the other hand, distinguishes between technical competences (knowledge and skills) and personal competences (social competence and independence).

This already reveals the problem of differentiation: responsibility and autonomy in the sense of motivational, volitional, and social skills are represented in the EQR and DQR as a competence category, but in the e-CF they are represented by the description of competence levels. In the field of competence "planning", for example, performance level 5 is shown in that someone that "uses far-reaching leadership strength to achieve consensus and commitment of the company's management to the ITK strategy". Such a level description does not distinguish precisely between ability and willingness, but rather assumes both. The ESCO (European Skills and Competencies of Occupations) project of the European Commission, on the other hand, defines "attitudes" or "attitudes and values" as independent, general competencies [8].

While EQR and DQR propose "knowledge", "skills" and "values" as different categories of a competence, the KSAVE model (Binkley et al., 2012) proposes them as different dimensions. The dimension perspective is also the one of this paper.The proposed competence framework is very detailed and complex in relation to the examples found in the literature, but the ambition was to create a framework comparable to the standardised European competence framework. The procedure for developing the competence framework was therefore based on the methodology used in comparable projects of the European Commission, such as e-CF or ESCO.

3. Result

The Data Literacy Framework is structured in four outline levels. They reflect the different levels of the process and its process steps as follows:

Outline level 1 determines 6 fields of competence, derived from the process steps: (A) Establish data culture - (B) Provide data - (C) Evaluate data - (D) Interpret results - (E) Interpret data - (F) Derive actions. The competence fields (A) to (C) correspond to the production process (“coding” from the real world to data and then to data products, the competence fields (D) to (F) correspond to the reception process (“decoding”) from data products to data and then to the real world.

Outline level 2 describes essential competencies for each field, each with a generic description. A further subdivision is made in competence field (B).

Outline level 3 gives examples of knowledge, skills/abilities, and motivation and (value) attitudes. The dimension "knowledge" refers to the knowledge needed to master the respective process step. This dimension deals with the (complex technical) knowledge. The dimension "skills" describes the corresponding abilities. It refers to the application of knowledge, i.e., the abstraction of what has been learned. The dimension "Attitudes, Values, Ethics" (AVE) describes motivations and values that an individual should possess. It covers ethical requirements, e.g., to enable objectivity and to avoid misuse of data and analyses. In addition, it is about motivation, openness, and the willingness to learn from mistakes. The AVE dimension is often neglected in competence frameworks. Although values and ethics

are sometimes formulated as separate competences or as competence bundles, the personal attitude of the individual is rarely the focus of attention.

Outline level 4, finally, specifies competence levels that represent a simplification of the EQR levels. They provide European references for the levels of complexity to which a competence applies. Here, the competences have been specified at three levels of complexity of the underlying requirements, which are to roughly outline a "basic level", an "advanced level" and an "expert level". They are defined via requirement levels that build on each other. These specifications should be worked out in detail in a follow-up study to enable the derivation of concrete learning objectives.

Outline levels 1 and 2 of the data literacy framework are summarised in the following Table 1.

Process	**Competence field**	**Competencies**	**Description**
Coding	(A) Establish data culture	(A1) Identify data applications	Identifies knowledge gaps and background information, identifies on this basis a concrete task that can be solved with the help of data, has an idea of the possible value contribution of the data
		(A2) Specify data applications	Defines minimum and optional requirements, defines delimitations to other tasks, structures the process flow into objects and their relationships, derives measurable objects and hypotheses about their relationships, communicates the requirements to an expert
		(A3) Coordinate data applications	Plans and coordinates a data project, if necessary, with participation of further individuals (from interdisciplinary areas)
Coding	(B) Provide data	(B1) Plan – (B1.1) Model data applications	Maps the measurable objects into variables with definable properties and their relationships in a model structure
		(B1) Plan – (B1.2) Comply with data protection/security	Observes guidelines for secure and ethically sound data processing and implements it accordingly where no clear guidelines are defined
		(B2) Obtain – (B2.1) Identify data sources	Identifies various common and novel data sources (internal, external) and evaluates their accessibility, relevance, and usability
		(B2) Obtain – (B2.2) Integrate data	Automatically reads data in various formats, integrates it and documents the integration
		(B3) Prepare – (B3.1) Verify data	Checks the data quality regarding various criteria (correctness, relevance, representativeness, completeness) and documents the audit systematically
		(B3) Prepare – (B3.1) Pre-process data	Cleans data, corrects errors, imputes missing values, standardises, and transforms data, filters relevant data for a given question, links data
Coding	(C) Evaluate data	(C1) Analyse data	Uses analysis methods from various disciplines (statistics, analytics, machine learning), with the help of suitable tools in a factual and purpose-oriented manner
		(C2) Visualise data	Utilises static and dynamic visualisations with the help of suitable tools in an appropriate and purpose-oriented manner
		(C3) Verbalise data	Verbalises the results of data analyses in various text forms in a factual and purpose-oriented manner
Decoding	(D) Interpret results	(D1) Interpret verbalisations	Interprets verbal descriptions of data (analyses) to draw conclusions about

			underlying data and results, and critically examines interpretations given explicitly or implicitly
		(D2) Interpret visualisations	Interprets visual descriptions of data (analyses) to draw conclusions about underlying data and results, and critically examines interpretations given explicitly or implicitly
		(D3) Interpret data analyses	Interprets results of data analyses, e.g., statistical parameters, to draw conclusions about underlying data points and relationships or to make forecasts
Decoding	(E) Interpret data	(E1) Decode transformations	Recognises and assesses the (statistical) methods used; recognises and interprets transformations of the data
		(E2) Trace back data sources	Can trace back, based on the analyses and information provided, how the data was obtained, from which source it originated and how much confidence can be placed in the data
		(E3) Reconstruct data concept	Can draw conclusions about the data basis and potential fallacies resulting from the process of mapping objects into data
Decoding	(F) Derive actions	(F1) Identify potential actions	Identifies concrete options for action and for the assessment of which can be evaluated with data; has an idea of the possible value contribution of the data when deriving possibilities for action
		(F2) Act data-driven	Integrates results into the decision-making process and acts based on these results
		(F3) Evaluate impact	Evaluates data-based actions based on their effectiveness for solving the task, identifies positive and negative outcomes, both anticipated or not

Table 1: Overview of the data literacy framework ([15], shortened and slightly modified)

4. Discussion and Conclusion

Since its first release in German [14], the data literacy framework has been implemented in over two dozen curricula of data literacy programmes at German universities. Furthermore, it has been adapted to derive the curriculum for an app for adult education in Germany. Creating data literacy for all, raising awareness of data handling issues and establishing a confident approach to new technological developments such as AI, IoT or Big Data and one's own data – these are some of the goals that the German Adult Education Association ("Deutscher Volkshochschul-Verband") is pursuing with the app "Stadt|Land|DatenFluss" [3]. The app is under the patronage of the German Chancellor. In a playful way, the topics of work/economy, smart city/mobility and health are addressed in a virtual city. To this date, the app has been awarded "App of the month July 2021" by the Deutsche Akademie für Kinder- und Jugendliteratur (German Academy for Children's and Youth Literature) and has been put on the shortlist of the Deutsche Preis für Online-Kommunikation 2021 (German Award for Online-Communication) in the category "innovation of the year".

This demonstrates the flexibility and practical usefulness of the data literacy framework, both in academic and non-formal education. It also supports the thesis that the basic concept of data literacy and its sub-areas applies to all educational fields and levels, even if the desired level of competence may differ in each case.

On the other hand, the curriculum of "Stadt | Land | Datenfluss" extends the framework presented here by examples of knowledge, skills, and attitudes, that could also be associated with AI literacy – a concept not systematically described in detail so far, even though it is mentioned in the DQ standard/ IEEE 3527.1™ Standard for Digital Intelligence (DQ) [4]. This is subject to further research.

It should be beyond question that data (and AI) literacy must be anchored in all formal and non-formal education sectors and thus established as part of general education. In concrete terms, this requires the inclusion of data literacy in the curricula and educational standards of schools, teacher training and higher education. To make lifelong learning of data literacy possible, data literacy programmes for extracurricular and vocational training are also needed.

References:

1. Binkley M, Erstad O, Herman J, Raizen S, Ripley M, Miller-Ricci M, Rumble M (2012). Defining Twenty-First Century Skills. In: Griffin P, Care E (Eds.): Assessment and teaching of 21st century skills. Methods and approach. Dordrecht: Springer (17:66).
2. Deutscher Qualifikationsrahmen (n.d.). DQR-Niveaus. https://www.dqr.de/content/2315.php
3. Deutscher Volkshochschul-Verband (2021). Stadt | Land | Datenfluss. Die App für mehr Data Literacy. https://stadt-land-datenfluss.de/
4. DQ Institute (2020). Global Standards for Digital Literacy, Skills, and Readiness. https://www.dqinstitute.org/global-standards/
5. European Commission (2008). Der Europäische Qualifikationsrahmen für lebenslanges Lernen. Luxemburg: Amt für amtliche Veröffentlichungen der Europäischen Gemeinschaften.
6. European Commission (2016). European e-Competence Framework 3.0. Ein gemeinsamer europäischer Rahmen für ITK-Fach- und Führungskräfte in allen Branchen. CWA 16234:2014 Teil 1.
7. European Commission (2018). The Digital Competence Framework 2.0. https://ec.europa.eu/jrc/en/digcomp/digital-competence-framework
8. European Commission (n.d.). ESCO – Skills/Competences – European Commission. https://ec.europa.eu/esco/portal/skill
9. Foucault M. (1980). Power/knowledge: Selected Interviews and Other Writings, 1972-1977. New York: Pantheon Books.
10. Heidrich J, Bauer P, Krupka D (2018). Future Skills: Ansätze zur Vermittlung von Data Literacy in der Hochschulbildung. Arbeitspapier Nr. 37. Kaiserslautern: Hochschulforum Digitalisierung.
11. Nicholson J, Gal I, Ridgway J (2018). Understanding Civic Statistics: A Conceptual Framework. A product of the ProCivicStat Project.
12. Organisation for Economic Co-operation and Development (2002). Definition and selection of competencies (DeSeCo): Theoretical and conceptual foundations: Strategy paper.
13. Ridsdale C, Rothwell J, Smit M, Ali-Hassan H, Bliemel M, Irvine D, Kelley D, Matwin S, Wuetherick B (2015). Strategies and Best Practices – Knowledge Synthesis Report.
14. Schüller K, Busch P, Hindinger C (2019). Future Skills: Ein Framework für Data Literacy. Arbeitspapier Nr. 47. Berlin: Hochschulforum Digitalisierung.
15. Schüller, K (2020). Future Skills: A Framework for Data Literacy. Working Paper No. 53. Berlin: Hochschulforum Digitalisierung.
16. Seidl T, Baumgartner P, Brei C, Lohse A, Kuhn S, Michel A, Pohlenz P, Quade S, Spinath B (2018). (Wert-)Haltung als wichtiger Bestandteil der Entwicklung von 21st Century Skills an Hochschulen (AG Curriculum 4.0). Diskussionspapier Nr. 3. Berlin: Hochschulforum Digitalisierung