



# IO 1: DaLiCo Map

## Intermediate Report – Part 1

Author/s:

Prof. Dr. Ulrike Spree – HAW Hamburg

Prof. Christine Gläser – HAW Hamburg



# IO 1: DaLiCo Map – Intermediate Report

## PROJECT INFORMATION

**PROJECT TITLE:**

Data Literacy in Context

**PROJECT ACRONYM:**

DaLiCo

**PROJECT NUMBER:**

2019-1-DE01-KA203-005066

**ERASMUS+ PROGRAMME:**

KA2 Partnerships for Innovation and the Exchange of Good Practices

**WEBSITE:**

[www.dalico.info](http://www.dalico.info)

**PARTNERS:**



Co-funded by the  
Erasmus+ Programme  
of the European Union





# IO 1: DaLiCo Map – Intermediate Report

## AUTHORS

Prof. Dr. Ulrike Spree (HAW Hamburg, Faculty of Design, Media, Information, Dept. Information)

Prof. Christine Gläser (HAW Hamburg, Faculty of Design, Media, Information, Dept. Information)

**DATE:** 2020-06-19

**START DATE:**

2020-02-20<sup>1</sup>

This document was written by partners of the "DaLiCo" project and is published as open educational resource. The project was funded through the European Commission. The content of this document exclusively reflects the opinions of the authors. The EU Commission and the National Agency at the DAAD assume no liability arising from the use of the information presented here.

## LICENCE

2020. This work is licensed under a

[Creative Commons License Attribution 4.0 International CC-BY-SA.](https://creativecommons.org/licenses/by-sa/4.0/)

---

<sup>1</sup> Start of task in IO1



## Contents

1	INTRODUCTION .....	1
2	THE CONCEPT OF DATA LITERACY – DEFINITIONS AND STAKEHOLDERS .....	1
3	THEORETICAL CONTEXTUALISATION – THEORIES OF KNOWLEDGE AND POLITICAL ENGAGEMENT .....	9
	RESOURCES .....	13

## 1 Introduction

The vivid discussion within the DaLiCo project indicated that the widely used Ridsdale matrix (Ridsdale et al. 2015) is a suitable starting point to arrive at a shared understanding of data literacy education (DL). On the other hand, it became obvious during the last 6 months that the participants' views on data literacy largely differ.

Based on the appraisal of research reviews in contributions in academic journals and theses, current programmatic publications on data literacy education, lighthouse projects like nationally or internationally awarded data literacy projects this working paper

- summarizes the most recent discussion (during the last three years based on existing literature reviews) on the concept of data literacy
- carves out common ground on data literacy education
- sheds light onto ongoing debates and controversies and identifies contentious issues

This reflection is necessary in order to determine which position the DaLiCo project will take within the ongoing, dynamic, and partly controversial discussion on data literacy and to circumscribe common ground for the intellectual outcomes, especially the train-the-trainer concept, the learning space and the data literacy standard<sup>22</sup>.

The results serve as basis for the data literacy knowledge map which will increase visibility of and grant access to existing competencies and personal expertise at the partner institutions including resources, projects and good practice examples.

## 2 The concept of data literacy – definitions and stakeholders

*"If social scientists do not understand people's definition of a situation, they do not understand it at all. That definition [...] is what people will shape their behaviour toward." (Bowker and Star 2000, p. 289)*

People cooperate across multiple communities of practice and bring different definitions and categories or category systems that they take for granted. In scientific work always members from more than one community of practice collaborate. Therefore, every scientific field is necessarily multi-disciplinary. We interpret information actively and operate in multiple contexts of use and practice (Bowker/Star, p. 293).

Within the DaLiCo project we accept and welcome these differences as a challenge and a chance. Instead of striving for a unified definition of data literacy we rather need to make the existing viewpoints visible in order to manage them with the aim to decide which definition(s) guide our work and in which contexts of use and practice we find our field(s) of action.

What are the prevalent definitions of the term "data literacy"? What do they have in common? In what respect do they differ? What are related and overlapping concepts? Do they coexist side by side or are they debated? Who are the main players/actors behind these definitions?

### ***What are the prevalent definitions of the term "data literacy"?***

The English version of the Wikipedia article from the 20<sup>th</sup> of January 2020 is marked with the following note: "It has been suggested that this article be merged into *Information literacy*". The explanation given for this suggestion on the corresponding talk page is that data is a synonym for information. Bendico42 strongly contradicts the suggestion arguing that recent educational approaches take a processual

<sup>22</sup>Due to the fast development in the field of data literacy education it is also necessary to reappraise our assumptions regarding the current state of data literacy education (awareness, teaching materials, integration into existing curricula) at the partner institutions. Based on an explorative investigation into student needs and the analysis of selected curricula the results of the explorative research will be summarized in a second working paper.

perspective on all data topics and that data is strongly related to key concepts of modern technology as computing, working with algorithms, encoding signals coming from sensors, treatment, aggregation, analysis and visualization of computable bits (Wikipedia 2019). The discussion is settled with the authors agreeing that although the exact meaning of the two concepts differs, they are both perceived as fairly new concepts. The articles have not been merged and the authors agreed on the following definition that since then has stayed stable

“Data literacy is the ability to read, work with, analyze, and argue with data. Much like literacy as a general concept, data literacy focuses on the competencies involved in working with data. It is, however, not similar to the ability to read text since it requires certain skills involving reading and understanding data.” (Wikipedia 2020a)

The shift of interest from information to data within the last ten years can be explained with the dynamic changes of methods of data collection and data sharing within the ongoing process of the digital transformation and globalisation, the increased accessibility of all kinds of data in the course of the open (access) movement and the new technical opportunities regarding the degree of automating the capture and analysis of data as well as the amount of data (big data) that can be processed (Heidrich, Bauer & Krupka, 2018, p. 14 ). Triggered by these developments people from outside the information technical field and with little or no technical understanding produce and work with data and as a natural consequence need a more thorough understanding of data. Simultaneously, new tools for e. g. data collection, coding or visualisation have become accessible and the threshold to work with data gradually decreases with an impact on decision making in business and democratic society (Gulsen 2019).

From the review of recent publications (Grillenberger 2018; Gulsen 2019; Schüller 2019) three ideal typical – partially overlapping but nonetheless different – understandings of data literacy can be distinguished.

- *The informed and critical application of data.* In 2015 (Ridsdale et al.) coined a generic rather open definition that follows the model of the other (especially information literacy) literacies and which strives to align the concept of information literacy to the demands of a data literate society. The definition is widely used and served as starting point for the DaLiCo project.

“Data Literacy is the ability to collect, manage, evaluate, and apply data; in a critical manner.” (Ridsdale 2015).

- *(Economic) innovation (or disruption)<sup>3</sup> starts with data literacy.* The economic and business management understanding that focusses on the economic application and use of data literacy skills either from the ‘leader’ perspective as a prerequisite for decision making or the employability perspective. As example for the first may serve a quotation from a blog entry of Information Age “Just as reading and writing skills needed to move beyond scholars 100 years ago, data literacy will become one of the most important business skills for any member of staff.” (Sommer 2017) and from a data literacy and digital marketing consultancy “Data literacy is the ability to understand data, how you can interact with it and what impact it can have.” (5 steps 2018). In Germany the Hochschulforum Digitalisierung, a “Think Tank”, as they like to call themselves, of higher education lobbyists (stakeholders) and private sponsors adopts this understanding.

„Data literacy is the ability to deal with data in a deliberate way and to use and question them consciously in their specific context. This includes the skills to collect, explore, manage, curate, analyse, visualise, interpret, contextualise, assess and apply data. Data literacy is a key competence for digitisation and the global knowledge society in all sectors and disciplines.“ (Hochschulforum Digitalisierung, 2017).

<sup>3</sup>“In business theory, a disruptive innovation is an innovation that creates a new market and value network and eventually disrupts an existing market and value network, displacing established market-leading firms, products, and alliances (Ab Rahman, Airini; et al. (2017).

A study commissioned by the Hochschulforum with the objective to identify overall necessities for the development of higher education curricula results in a suggested framework for data literacy (Schüller et al.). The authors of the framework build on the earlier definition from 2017 but shift the focus from abilities, skills and competencies to behaviour and attitudes and at the same time stress the procedural aspect and the relation to value creation and revenue.

“Data Literacy ist das Cluster aller effizienter Verhaltensweisen und Einstellungen für die effektive Durchführung sämtlicher Prozessschritte zur Wertschöpfung beziehungsweise Entscheidungsfindung aus Daten” (Schüller et al. 2019 p. 26).

Data literacy is the cluster of all efficient behaviour patterns and attitudes necessary for the effective execution of all process steps to create value respectively support decision making from data. (translation Ulrike Spree)

- *Data literacy is a necessary prerequisite for societal and political engagement.* The Data-Pop Alliance, a collaborative initiative of the Harvard Humanitarian Initiative, MIT Connection Science, and Overseas Development Institute in their definition highlight the aspect of social engagement.

“Data literacy is the desire and ability to constructively engage in society through and about data” (Data-Pop- Alliance).

To underline the emancipatory dimension a group of scholars very much influenced by the critical pedagogy approach initiated by Paulo Freire coined the term critical data literacy.

“Critical Data Literacy is the set of abilities which allows one to use and produce data in a critical way” (Tygel; Kirsch 2016, p. 117).

### ***Different stakeholders privilege different competencies and approaches.***

Different interest groups and communities of practice within the larger network of data literacy education initiatives often identify with one of the above definitions.

The *library and information science community* has taken up the Ridsdale et al understanding of data literacy which is based on a synopsis of several data literacy education frameworks mapped to the Data information literacy framework (2013) originally published by Carlsen et al. (Carlsen et al. 2013, Schüller et al. 2019). In their activities they address a more generic set of competencies and a clear focus is on data management, especially research data management competencies. The main target group of these activities are professionals from the library and information science community themselves and students (undergraduate and graduate) and university teachers and professors. The community also reaches out into the field of education (higher education, college). A vital focus of this community of practice are activities supporting research data services. Proponents of this group, like the Hungarian researcher Tibor Koltay highlight the relationship between critical thinking and data literacy.

“This leads to the conclusion that one of the most important goals of data literacy education should be to foster critical thinking that keeps us away from the pitfalls of being overly optimistic or unduly pessimistic, behaving in an excessively critical or uncritical way. In any case, critical assessment of data includes giving emphasis to the version of the given data set, the person responsible for it (ACRL, 2013).” (Koltay 2015, p. 415)

*Open government and NGO’s initiatives* are basically active in two ways. a) educational in the sense that they strive to support everybody (the general citizen) in their everyday-use of data e. g. in search-engines, data-platforms, social media and as well wish to elucidate people how to secure their data as – as long as they do not truly understand the very concept of data, like behavioural data derived from social media and how deep the predictions can go based on this data - they might sacrifice their privacy over economic interests. b) with a stricter focus on maintaining democratic participation, striving for utmost transparency and supporting collaboration in all fields of society. In this approach data literate citizens are the ones who are able to understand the possible implications of data and from a more technical perspective, they also use open government data in order to improve society as well as reinforce democratic values and the accountability of governments. Exemplary players behind this

understanding are the Data-Pop-Alliance or the Open Knowledge foundation. The growing data-journalism community can also be partly attributed to this interest group and community of practice.

Representatives and lobbyists from the private sector, higher education (especially with a MINT or economic or business science background) as well as political decision makers adopt the second definition with the focus on innovation in order to secure the employability of the future workforce. This is based on the conviction that pupils of today will become workforce in a data-driven future, where not only technical skills but also problem-solving and computational thinking will be indispensable. Internationally active software and platform providers like Qlik (Qlik 2020) and industry (Future Skills Initiative 2020<sup>4</sup>, Bertelsmann Stiftung 2020<sup>5</sup>) in general cultivate this field of data literacy. Their efforts result either in the publication of policy guidelines and directed project grants as well as complete programmes for inhouse-trainings. Culture and data ethics in this context refer to an affirmative and approving corporate culture with respect to data<sup>6</sup> (Qlik 2019).

The above mapping of stakeholders and interest groups to different definitions and understandings of data literacy is of course one-dimensional and deliberately exaggerated to illustrate that data literacy needs to be understood within the respective context. De facto there are many initiatives and actors for example international organisations like the UN (UN's Data Revolution Group 2014) that strive to bridge between different stakeholder groups. The Data Revolution consultation subsumes data literacy under the general aspect of accessible data and places data literacy at the intersection of statistical literacy, information literacy and technical skills for working with data (UN Data Revolution Group 2014b). The United Nations World Data Forum narrows down the focus by locating data literacy at the intersection between numeracy and statistical literacy (United Nations World Data Forum).

What do the portrayed approaches have in common? All perceive data literacy as a set of practices within the data lifecycle with the steps of collection, management, evaluation, application.

In what respect do they differ? Main differences can be found regarding the focus on one of the steps (library and information science for example focuses on the collection and management aspect), the depth for example of data evaluation and analysis (is machine learning an integral part of data literacy?) as well as the purpose.

### Data Literacy is a dynamic concept

Since the coining of the term data literacy<sup>7</sup> in the late 1970s it has been object of a couple of still ongoing debates regarding the 'correct path' to a more data literate society. Gulsen identifies the following main debates:

<sup>4</sup> The Future Skills Initiative list of supporters ranges from insurance companies like Alliance, over pharma industry (BAYER, BASF) to automobile industry and financial institutions to huge consultancies. <https://www.future-skills.net/initiative>

<sup>5</sup> Bertelsmann Stiftung (2020)

<sup>6</sup> Qlik defines data culture as "organizations or individuals work to acquire, cultivate and utilize skills within data literacy". Relevant ethical questions according to this understanding are questions like "How representative is (your) data set", does a "data model accounts for biases", if "the predictions based on big data are accurate" and if the "reliance on big data raise ethical or fairness concerns".

<sup>7</sup> The first article with the term "data literacy" accessible via Google scholar dates from 1978. In a discussion on higher education and critical thinking literacies in general are discussed as a prerequisite of establishing a connections between the world of the teacher and the pupil. Theodore L. Gross (1978) The Organic Teacher, Change: The Magazine of Higher Learning, 10:6, 37-77, DOI: [10.1080/00091383.1978.10569454](https://doi.org/10.1080/00091383.1978.10569454). The ERIC thesaurus does not list the term data literacy. It only refers to literacy and data. The first article with the term "data literacy" listed in Scopus dates from 2005. Charles Humphrey (2005) Collaborative Training in Statistical and Data Library Services, Resource Sharing & Information Networks, 18:1-2, 167-181, DOI: [10.1300/J121v18n01\\_13](https://doi.org/10.1300/J121v18n01_13)



- What is the relationship between data and information?
- Do data literacy competencies differ depending on the ‘type of data’ considered and is a common unified definition of data literacy desirable and achievable?
- What is the core competency, common ground of data literacy and how is the relationship between data literacy and data science?

1) The discussion about the relationship between data versus information has reached a kind of consensus and the DIKW (data, information, knowledge, wisdom) pyramid is widely used. In the words of Jennifer Rowley "Typically information is defined in terms of data, knowledge in terms of information, and wisdom in terms of knowledge" (Rowley 2007). Often data, information and knowledge are mapped to the 4 W questions. Data/information to what, knowledge to how and wisdom, understanding to why.

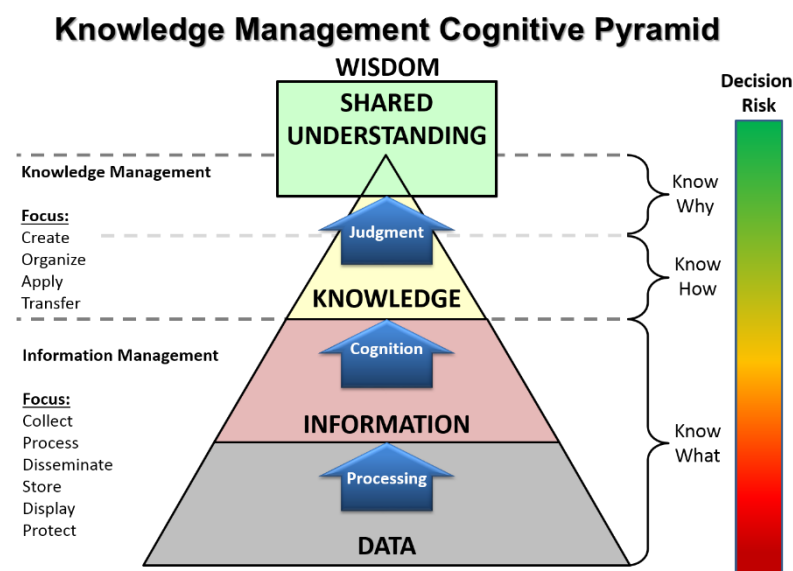


Fig. 1: DIKW Adaptation in use within the US Army KM Community of Practice. Matthiew.viel. CC-BY-SA. 4.0.

[https://en.wikipedia.org/wiki/DIKW\\_pyramid#/media/File:KM\\_Pyramid\\_Adaptation.png](https://en.wikipedia.org/wiki/DIKW_pyramid#/media/File:KM_Pyramid_Adaptation.png)

Whereas the DIKW pyramid is widely accepted in the knowledge management community of practice it needs to be mentioned that within in the philosophy of science community there exists an ongoing debate on the epistemic and ontological relationship between data, information and knowledge. The information scientist Rafael Capurro argues that data, information and knowledge cannot emerge out of each other as they refer to different “taxonomies of knowledge” (Capurro 2018).

2) Are data literacy skills and competencies essentially different depending on the domain and the kind of data and is a common and unified definition desirable and achievable?

There is a general consensus that data literacy education needs to address all steps of the data lifecycle (collect, manage, evaluate, and apply data) and that simultaneously the competencies need to be contextualized. Apart from this the approaches differ. Whereas the diverse frameworks (Carlson, Ridsdale, Schüller) which originate in the context of university education strive to provide a common, interdisciplinary framework that allows to map the different activities dependent on the domain and context data literacy is applied to alternative approaches insist on the necessity to spell out data literacy for different subject-areas. To arrive at a specific data literacy competence framework for information science education at secondary school level Grillenberger for example combines aspects from data management and data science (Grillenberger 2018, p. 147). As a further element the model distinguishes between specific content fields and processes of data literacy. For the final model he interlaces content areas and processes.

<b>C</b>	Data and information		
<b>O</b>		Capture, clean, model	<b>P</b>
<b>N</b>	Data storage		<b>R</b>
<b>T</b>		Implement and optimize	<b>O</b>
<b>E</b>	Data analysis		<b>C</b>
<b>N</b>		Analyse, visualise, interpret	<b>E</b>
<b>T</b>	Data ethics and protection		<b>S</b>
		Replace, archive, delete	<b>S</b>

Fig. 2: Interlacing of content fields and processes (based on Grillenberger 2018)

3) What is the core competency, common ground of data literacy and how is the relationship between data literacy and data science?

The last – often only implicit debate – revolves around the question whether ‘soft skills’ respectively ‘personal skills’ like critical thinking, ethics is also mentioned in this context, or ‘technical skills’, or primarily statistics and programming/coding should form the basis of data literacy education (Gould 2017). Closely related to this question is the debate concerning the adequate relationship between data literacy and data science. Some stakeholders for example from the library and information science community who primarily deal with tasks around (research) data management and the provision of access to usable data view data literacy and data science as two distinct independent approaches in their own right; others perceive data science as a more mature and higher level of data literacy (Gulsen 2019, p. 33, Schüller 2019, p. 16, Kelleher 2018). Especially, this last debate is not free from ideology. For example, Schüller worries that out of economic considerations the term data literacy could be unduly broadened because “data science” has become a buzzword and data scientists are better paid than statisticians” (Schüller 2019, p. 17).

#### **Excursus: Data literacy terminology**

DaLiCo partners agreed to focus for their activities on resources and material accessible in English. To better pigeonhole specific unique national paths, we provide a short excursus on the special terminology chosen in German, Dutch, Hungarian and Spanish.

The German Wikipedia presents a literal translation of Ridsdale. In the definition Datenkompetenz (literally data competence) or the English term data literacy are suggested.

**“Datenkompetenz bzw. Data Literacy** umfasst die Fähigkeiten, Daten auf kritische Art und Weise zu sammeln, zu managen, zu bewerten und anzuwenden.“(Wikipedia Datenkompetenz 2020)

Data competence respectively data literacy comprises the ability to collect, manage, evaluate and apply data in a critical manner. (translation Ulrike Spree)

Scraping google for “data literacy” and “Datenkompetenz” generates largely different results, also both terms are used interchangeable in some titles (Data literacy: Datenkompetenz im Unternehmen / <https://www.heyde.ch/data-literacy-datenkompetenz-im-unternehmen>). Two searches on Google with „data literacy“ and “data science” have no similar results among the first 49 hits. The results list of the search “Datenkompetenz” is dominated by sources from industry and school education whereas “data literacy” is favoured in resources with that relate to universities. Among the top 100 for “Datenkompetenz” 5 come from idw-online and report about projects in the context of industry 4.0. Among the top 100 hits 7 are from the hochschulformdigitalisierung and 5 from the dataliteracyproject.

The Spanish, Dutch and Hungarian Wikipedia have no self-contained entries on literal translations of the term data literacy.

In the Spanish Wikipedia there is no entry on “alfabetización de datos”, however an entry “alfabetización informática” which focuses on inclusion and social justice through the use of information and communication technologies. The article states a fierce contrast to the concept of “alfabetización digital” which is reduced to the objective to support the use of new technologies with the sole aim to increase administrative and economic efficiency. The article mainly refers to the Latin American discussion.

Se llama **alfabetización informática** al conjunto de tareas e iniciativas que tienen como objetivo incluir, en la Sociedad del Conocimiento, los sectores marginados de la misma por razones económicas, sociales y/o culturales. Inclusión que se realiza cuando las personas que integran sectores socialmente marginados desarrollan las destrezas necesarias para utilizar tecnologías de información y comunicación, que les permiten utilizarlas como herramientas para su recalificación social y el mejoramiento de su calidad de vida.

Por el contrario, la alfabetización digital tiene como objetivo el uso de las nuevas tecnologías en función de la eficiencia de los entes administrativos. Los sujetos de la alfabetización digital son considerados consumidores o usuarios, lo único que deben aprender es ejecutar operaciones básicas, tanto para conseguir los bienes que desean como para usar mecanismos masivos. Esta alfabetización está centrada en el desarrollo de equipamientos, quedando el conocimiento de los procesos en manos solamente de los especialistas (Wikipedia Alfabetización informática 2020)

Within the academic discourse the term alfabetización en datos or de datos is common. It is closely related to information de datos and digital literacy. Similar to the English language discourse the three terms are used more or less interchangeable. The results of a Google Scholar query on the search term “alfabetizacion de datos” from July 2020 reveals the same plurality as the English results with reference to all three ideal-typical understandings of data literacy and different stake holders. The articles fall into four overlapping clusters dealing with pedagogical issues (school and university level), especially referring to the teaching of ICT, societal issues, for example egovernment, economic/business demands and scientific needs. More than half of the first 40 hits also deal with the concept big data (Appendix 1. List Google Scholar Alfabetizacion de datos).

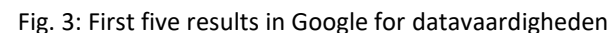
In the Magyar version of Wikipeida there is also no entry on the literal translation of data literacy which is “adattudás” or data science “adattudós.

Scraping the first 100 google results reveals a clear distinction between adattudás (data literacy) with more sources from library and information science (for example Koltay 2015; Koltay 2017) and adattudós (data science) from the IT domain. The English “data scientist” is seen as a superordinate term to (Adattudós, Gépi tanulási mérnök, Adatmérnök, Adatelemző) which are described as different roles that in the professional practice may overlap and be carried out by the same individual. The list also implies a kind of hierarchy the data scientist at the top of the hierarchy.

Adattudós	Datenwissenschaftler /data scientist
Gépi tanulási mérnök	Ingenieur für maschinelles Lernen /engineer for machine learning
Adatmérnök	Dateningenieur /data engineer
Adatelemző	Daten Analyst/ data analyst

Catherine D'Ignazio and Lauren Klein observe the same mechanism where low status “data analysts” (low-status number crunchers) have become labelled as high status researchers “data scientists” (D'Ignazio; Klein, 2020, Chapter 1, p.)

If one searches google.nl for the Dutch term “datageletterdheid” already in some titles of the search results the English term data literacy appears. Similar to the other languages in the 100 list there are sources from business (quilk) with a focus on secondary school and university education. The hits on “datavaardigheden” (data skills or competencies) are completely different, although again the terms are used synonymously in the titles of the retrieved resources (Datageletterdheid | Data literacy | Hoe datageletterd ben jij?/ <https://www.passionned.be/bi/datagedreven-werken/datageletterdheid/>) and show a strong focus on programming (for example python or R).



### Conclusion and implication for DaLiCo

Although, the three ideal-typical directions outlined above display a holistic understanding of data literacy and the Ridsdale definition serves as a kind of lowest common denominator we observe (regarding curricula, user demands, offering of data literacy training and courses) a clear distinction regarding the determination of core topics and focus between three different subfields of data literacy:

- 

In addition, the literature review reveals a kind of “data literacy” paradox or dilemma. On the one hand the authors stress the need to deal with data in an interdisciplinary, respectively multidisciplinary way and at the same time emphasise that the particularly needed competencies differ largely between domains.

The authors of the framework for data literacy (Schüller 2019) acknowledge these apparent paradoxes and try to bridge these differences by allowing for intersections between the three fields and at the same time taking into consideration the special domain specific or better disciplinary requirements of data literacy education. The framework consists of a three-dimensional matrix, describing data literacy activities along three main facets.

1) Firstly, 6 main fields of competencies are determined along the data lifecycle. Each subfield is specified by the assignment of 3 to 6 competencies.

- A) Establishing a data culture – from the system to measurable objects
- B) Provide access to data – from measurable objects to data
- C) Analyse data – from data to data products
- D) Interpret data products – from data products to data
- E) Interpret data – from data to measurable objects
- F) Deduce recommendations for actions – from measurable objects to the system (Schüller 2019, Anhang C)

2) Contextuality is taken into consideration by taking up an idea of the design consultancy designation (Münster 2019) describing data literacy as a binary process that involves coding (releasing data from context) and decoding (contextualizing). In the categories of the framework this is expressed by the varying focus from systems to measurable objects to data to data products.

3) With recourse to the KSAVE (Knowledge, skills, attitudes, values and ethics) model (Binkley 2012) three competence dimensions (knowledge, skills, attitude) are added to the framework.

In line with this framework data literacy is defined as

“the cluster of all efficient behaviours and attitudes for the effective execution of all process steps for the creation of value and/or decision making from data.” (Hochschulforum Digitalisierung 2020, p 25)

It is suggested DaLiCo broadens the initially planned grounding on the Ridsdale matrix and adopts the categories from the Future Skills Framework. This will be implemented in the data model for the DaLiCo Map.

However, the carved out paradox not only indicates the need for an inclusive understanding of data literacy that allows for the intersection of different perspectives. It also alludes to a fundamental underlying problem of existing data literacy (education) approaches, the general lack of an explicit discussion of the particular theoretical backgrounds the diverse understandings and deviated interventions to increase data literacy competencies are based on.

### 3 Theoretical contextualisation – theories of knowledge and political engagement

Looking at the ideal-typical definitions quoted above it becomes obvious that the different understandings of data literacy vary according to the different theoretical background they are approached from. These are however only rarely explicitly stated in the screened publications like frameworks, definitory approaches or case studies and educational initiatives. In most cases the approach is either completely theory agnostic ‘hands on’ or relies on established educational frameworks (EQR; HQR; TUNING PROJEKT). Generally, the approaches either demonstrate a narrow

disciplinary focus on individual subject fields, areas of data expertise or they concentrate on the teaching of assumed generic skills and competencies.

A didactical reflection and the backing on theories how people learn is often somewhat tacitly implied. For example, Schüller only explicitly addresses the aspect of learning outcomes when it comes to measuring and testing learning success (Schüller 2019, p. 17). It is only in this context that for example taxonomies for learning objectives like the Bloom taxonomy are mentioned. Overall, a very functionalistic understanding of learning focussed on the economic benefit is applied based on an understanding of knowledge as commodity/raw material and skills as tools for task solving (Schüller 2019, p. 27). Even the consideration of ethical aspects is presented as ultimately subordinated to economic goals<sup>8</sup> (Schüller, 2019, p. 29).

Alternative approaches come from various academic and social areas. To complete and counterbalance the predominant focus on economic benefit selected alternative approaches are presented below. They have their roots in educational theory building, in sociology and political science (e. G. critical theory or information infrastructure (Bowker, Star 2008)) and journalistic and media and communication studies (Sander 2020). Data literacy is perceived as embedded in social practices like in data feminism and participatory design (D'Ignazio, Klein 2020).

#### *Recognizing the learner's perspective – Constructivist perspective*

Constructivist views of learning emphasise the importance of understanding the learners' perspective (Limberg; Sundin 2020, p. 11, Koltay 2019). Before educational interventions are planned the studying of variations in people's ways of experiencing different phenomena, for instance the understanding and use of data and information is inevitable. The Library Carpentry Movement offers a very good example how much learning interventions benefit from a clear pedagogical model like the Dreyfus model of skill acquisition (Koch, Wilson 2016).

#### *Sociocultural perspective*

The sociocultural perspective is rooted in the concept of community of practice (Snow 1959). In order to participate in a practice and its activities individuals need to learn a specific language, often referred to as "vocabularies" (Markham 2019, p. 755) where certain concepts, theories and beliefs are central. Hence, information literacy implies learning to communicate appropriately within a specific practice and at the same time experiencing the crossing of borders between disciplines. The sociocultural perspective further implies that this is also about being able to use artefacts and real experiences for communication in a way that corresponds with the purpose of the practice. Markham describes her approach of provoking ethical reflections by telling a story that submits students to an 'ethical dilemma' (Markham 2019, p. 756). Sociocultural approaches resort to critical theory and take their point of origin in the assumption that 'there might be a problem'. As Markham puts it "Applied deliberately and consistently to everyday lived experiences, this stance is a strong response to datafication, as it questions everything from the perspective of "who benefits and who loses?" (Markham 2019, p. 757). Instead of taking existing data 'for granted' data is constantly reflected as made and somewhat problematic. This critical or sceptical attitude towards data is derived from investigating which forms data take and how they are produced. As a consequence the focus of data literacy interventions shifts from teaching statistical and research methodology to also incorporate the organization of science and the production of knowledge (Špiranec 2019). Literacy education is geared to carving out the process of meaning making, of sense making as well as dissenting interpretations (Markham 2019, p. 758).

---

<sup>8</sup> "Daten werden dabei als eine werthaltige Ressource verstanden, die von Bürger\*innen ohne unmittelbare Gegenleistung zur Verfügung gestellt würden, für die er jedoch einen moralischen Anspruch auf Partizipation an den Ergebnissen der Datenleistung erwerben. Gewisse Interessen stünden möglicherweise in Konflikt, etwa der Wunsch der Bürger\*innen auf Transparenz und das Bedürfnis einer Behörde nach Diskretion im Falle von Strafverfolgung." Schüller 2019, p. 29



“Shifting the conversation away from data as thing or object—including its basic characteristics but also its collection, storage and protection—and toward data as ideology.” (Markham, 2017)

### *Data infrastructures perspective*

A special direction of the sociocultural perspective that is especially relevant in the context of dealing with data and information is the information respectively data infrastructure perspective that goes back to works of Bowker and Star in the 1990s. In “Sorting Things out” they expand the concept of infrastructure and not only refer to technical infrastructures in a narrow sense but also to categorical infrastructures like vocabularies or classifications systems and carve out a special characteristic of infrastructure; it only becomes visible when it ‘breaks down’ (Bowker, Star 2000). The proponents of this approach in data literacy education shift the perspective from competencies in reading and working with datasets to the ability to account for, influence and participate in the socio-technical infrastructures through which data is created, stored and analysed (Gray et al. 2018). Especially from this school of thought critical data (infrastructure) literacy is perceived as subject for public involvement and experimentation around data infrastructures (e. G. platforms or protocols) and reflecting the social as well as individual consequences of datafication (Gray et al. 2018).

### *The feminist view*

The data feminism movement – basic approaches are summarized in the recent publication “Data feminisms” - offers a clearly stated alternative concept as it grounds data literacy on a fundamental challenge of the presumption of scientific objectivity. Data feminists share the conviction that all knowledge (including sciences) is generated from a certain perspective. Instead of trying to avoid or deny bias (e. G. in data) one should strive for transparency and making different viewpoints visible (D’Ignazio, Klein 2020, Chapter 2, <https://mitpressonpubpub.mitpress.mit.edu/pub/8tjbs2x5/release/2>). Through the publications the term data science is favoured over data literacy but data science is demystified and the authors strive to counterbalance the assumed hierarchy between data cleaners at the bottom of the hierarchy and the analysts at the top.

“Embracing pluralism in data science means valuing many perspectives and voices and doing so at all stages of the process—from collection to cleaning to analysis to communication. It also means attending to the ways in which data science methods can inadvertently work to suppress those voices in the service of clarity, cleanliness, and control.” (D’Ignazio, Klein 2020, Chapter 5, <https://data-feminism.mitpress.mit.edu/pub/2wu7aft8/release/2>)

Data as well as data science is perceived as a form of power which can likewise be used to uncover discrimination or suppression as well as an instrument to discriminate, surveil, exploit or suppress. The approach is very much informed by the ideas of intersectional feminism (with reference to the work of Patricia Hill Collins) based on the concept of a matrix of domination to explain how power is exercised and experienced. The seven chapters of the book are organized around seven principles that are summarized to a kind of manifesto of a feminist approach to dealing with data. Data is considered as never neutral and contextualised. It is used to examine and challenge power by collecting, analysing, dissemination data to overcome power structures and arrive at greater equity. In contrast to other understandings of data literacy the elevation of emotion, which is in line with starting from personal and/or local experiences, and the role of embodiment is emphasized. Data feminists query simple binaries and encourage pluralism and multiple perspectives (D’Ignazio, Klein 2020)

In contrast to similar politically inspired approaches like Data for Good (<https://dataforgood.ca/>) or more than code (<https://morethancode.cc/report/ecosystem/>) data feminism is firmly based on a participatory community based approach termed “Co-Liberation” (D’Ignazio, Klein p. 35).

### *Conclusion and implication for DaLiCo*

The cursory literature report confirmed that the Ridsdale matrix is a useful smallest common denominator for a shared understanding of data literacy. The additional dimensions of coding and

decoding data as well as the orientation along the data life cycle suggested in the framework for data literacy should be considered for future activities within DaLiCo.

Apart from this common ground the review also reveals the existence of different communities / interest group within the data literacy field and that the exchange between these groups is rather limited. One reason for this lack is that position papers, guideline, frameworks rarely make the theoretical background they come from explicit. These different orientations (research data management, metadata; statistics and data science; business and economic benefit of data literacy activities) are also present within the DaLiCo team.

Notwithstanding the diverse directions different proponents take they agree on a number of aspects.

- Data literacy needs contextualisation and precise problem definitions and examples.
- Ethical aspects as well as critical thinking should be communicated/discussed/taught in relation to the appending of technical concepts and skills (tool training).
- The sustained success of data literacy interventions is not always achieved and needs to be evaluated and monitored.

Alongside the sort of 'mainstream' discourse on data literacy there exists a tradition of 'critical' data literacy that in addition to promoting and developing data literacy skills also addresses political questions like who benefits and who loses.

Altogether, for the further course of the project, the results of the literature report encourage us to work on allowing several perspectives on data literacy and to make them explicit by

- striving to explore different access points and approaches to data literacy and make these different access point visible in our future activities and
- cater for relations and linkings (shaping context)

With the format and concept of the Summer School 2020 on Open Government Data already a couple of identified problems and challenges are addressed implicitly:

- Contextualisation of Data Literacy interventions (World Happiness Report)
- Facilitators from different communities (academia (Data Science), civic society, journalism, library and information)
- Combining tool training and critical thinking and ethical principles



## Resources

5 steps in becoming a data literate leader. Retrieved from: <https://missiondrive.io/5-steps-to-becoming-a-data-literate-leader>

Ab Rahman, A.; et al. (2017). Emerging Technologies with Disruptive Effects: A Review. PERINTIS eJournal. 7 (2). Retrieved 2017-12-21

Bellinga, Pim (2017). Creating a curriculum to help civil society become more data literate. Retrieved 2020-02-18, from: <https://medium.com/ihatestatistics/creating-a-curriculum-to-help-civil-society-become-more-data-literate-6b0649c0f31a>

Bertelsmann Stiftung (2020). Wer wir sind. Retrieved: <https://www.bertelsmann-stiftung.de/de/ueber-uns/wer-wir-sind/ansprechpartner/mitarbeiter/cid/ralph-mueller-eiselt/>

Binkley, M.; Erstad, O.; Herman, J.; Raizen, S.; Ripley, M.; Miller-Ricci, M.; Rumble, M.. Defining Twenty-First Century Skills (2012). In: Griffin, P.; Care, E. (Hrsg.): Assessment and teaching of 21st century skills. Methods and approach. Dordrecht, S. 17–66

Bowker, G. C., & Star, S. L. (2000). Sorting things out: Classification and its consequences. MIT press.

Carlson, J., Johnston, L., Westra, B. & Nichols, M. (2013). Developing an Approach for Data Management

Data-Pop-Alliance. Retrieved: <https://datapopalliance.org/beyond-data-literacy-workshop/>

D'Ignazio, C., & Bhargava, R. (2016). DataBasic: Design Principles, Tools and Activities for Data Literacy Learners. The Journal of Community Informatics, 12(3). Retrieved from: <http://ci-journal.net/index.php/ciej/article/view/1294>

Education: A Report from the Data Information Literacy Project | Inter-national Journal of Digital Curation, 8 (1).

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

Gray, Jonathan; Gerlitz, Carolin; Bounegru, Liliana (July 10, 2018). Data infrastructure literacy. Big Data & Society. 5 (2): 1–13. Retrieved from DOI: 10.1177/2053951718786316

Grillenberger, A. (2018). Von Datenmanagement zu Data Literacy: Informatikdidaktische Aufarbeitung des Gegenstandsbereichs Daten für den allgemeinbildenden Schulunterricht. Berlin: FU Berlin. Dissertation. Retrieved 2020-01-19, from: <http://dx.doi.org/10.17169/refubium-1932>

Heidrich, J., Bauer, P. & Krupka, D. (2018). Future Skills: Ansätze zur Vermittlung von Data Literacy in der Hochschulbildung. Arbeitspapier Nr. 37. Berlin: Hochschulforum Digitalisierung. Retrieved 2020-01-19, from DOI: 10.5281/zenodo.1413119

Hochschulforum Digitalisierung (2020). Future Skills: A Framework for Data Literacy. Working Paper No. 53. Berlin: Hochschulforum Digitalisierung.

D'Ignazio, C., & Klein, L. F. (2020). Data feminism. MIT Press.

IMB (2020). HAW Hamburg. Information, Media and Library Management. Retrieved 2020-07-31,

from: <https://www.haw-hamburg.de/en/study/degree-courses-a-z/study-courses-in-detail/course/courses/show/information-medien-bibliothek/Studieninteressierte/>.

Information Age (Journal). Retrieved from: <https://www.information-age.com/about-information-age/>

Khan, H. R., Kim, J., & Chang, H. C. (2018). Toward an Understanding of Data Literacy. iConference 2018 Proceedings.

Kelleher, J. D., & Tierney, Brendan (2018). Data science. Cambridge, Massachusetts: The MIT Press.

Christina Koch and Greg Wilson (eds) (2016). Software Carpentry: Instructor Training. Version 2016. Retrieved 2016-05-06, from: <https://github.com/carpentries/instructor-training> DOI: 10.5281/zenodo.57571.

Koltay, T. (2015). Data literacy: in search of a name and identity. Journal of Documentation, Vol. 71 No. 2, pp. 401-415. Retrieved from DOI: 10.1108/JD-02-2014-0026

Koltay, T. (2017). Data literacy for researchers and data librarians. Journal of Librarianship and Information Science, 49(1), 3-14.

Koltay, T. (2019). Data literacy in academia: Basics and pedagogical views. Opus et Educatio, 6(4).

Mackey, T. P., & Jacobson, T. E. (2014). Metaliteracy: Reinventing information literacy to empower learners. American Library Association.

Markham, A. N. (2020). Taking data literacy to the streets: critical pedagogy in the public sphere. Qualitative Inquiry, 26(2), 227-237.

Markham, Annette. (2018). Critical Pedagogy as a Response to Datafication. Qualitative Inquiry. Retrieved from DOI: 10.1177/1077800418809470.

Münster, E. (2019, Januar 30). The Cycle of Encoding and Decoding. Retrieved 2020-07-24 from: <https://www.designation.eu/en/2019/01/31/why-does-no-one-understand-my-data-visualization/>

Presmarmethuen.org (2020). Retrieved from: <https://presmarmethuen.org/hu/dictionary/is-there-a-difference-in-salary-between-a-data-scientist-and-a-machine-learning-engineer/> Retrieved 2020-07-28.

Qlik (2020). About Qlikc: Our Vision for the future. Retrieved from: <https://www.qlik.com/us/company>

Juliana E. Raffaghelli & Bonnie Stewart (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. Retrieved from DOI: [10.1080/13562517.2019.1696301](https://doi.org/10.1080/13562517.2019.1696301)

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S & Wuetherick, B. (2015). Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report. Retrieved 2020-01-19, from DOI: [10.13140/RG.2.1.1922.5044](https://doi.org/10.13140/RG.2.1.1922.5044)

Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information and Communication Science. 33 (2): 163–180. Retrieved from DOI: [10.1177/0165551506070706](https://doi.org/10.1177/0165551506070706).

Sander, I. (2020). What is critical big data literacy and how can it be implemented?. Internet Policy Review,[online] 9(2). Retrieved 2020-07-30, from: <https://policyreview.info/articles/analysis/what-critical-big-data-literacy-and-how-can-it-be-implemented>

Snow, C. P. [1959]. *The Two Cultures*. London: Cambridge University Press

Schüller, K., Busch, P., Hindinger, C. (2019). *Future Skills: Ein Framework für Data Literacy – Kompetenzrahmen und Forschungsbericht*. Arbeitspapier Nr. 47. Berlin: Hochschulforum Digitalisierung. Retrieved from DOI: [10.5281/zenodo.3349865](https://doi.org/10.5281/zenodo.3349865)

Sommer, Den (2017). 2017. The year of information Literacy. <https://www.information-age.com/2017-year-data-literacy-123463401/>

Špiranec, S., Kos, D. & George, M. (2019). Searching for critical dimensions in data literacy. In *Proceedings of CoLIS, the Tenth International Conference on Conceptions of Library and Information Science*, Ljubljana, Slovenia, June 16-19, 2019. *Information Research*, 24(4), paper colis1922. Retrieved from: <http://InformationR.net/ir/24-4/colis/colis1922.html> (Archived by the Internet Archive at <https://web.archive.org/web/20191217174118/http://informationr.net/ir/24-4/colis/colis1922.html>)

Tuning. Educational Structures in Europe. Retrieved from: <http://www.unideusto.org/tuningeu/>

Tygel, A., & Kirsch, R. (2016). Contributions of Paulo Freire for a Critical Data Literacy: a Popular Education Approach. *The Journal of Community Informatics*, 12(3). Retrieved from: <http://ci-journal.net/index.php/ciej/article/view/1296>

UN Data Revolution Group (2014a). Data Revolution Group. Retrieved from: <https://www.undatarevolution.org/>

UN Data Revolution Group (2014b). Accessible Data: open data, accountability and data literacy. Retrieved from: <https://www.undatarevolution.org/data-use-availability/>

Wikipedia contributore. (2020). Alfabetización informática. In: Wikipedia, La enciclopedia libre. Version: 2020-04-20. Retrieved: 2020-06-18 06:39 UTC, from: [https://es.wikipedia.org/w/index.php?title=Alfabetizaci%C3%B3n\\_inform%C3%A1tica&oldid=125341516](https://es.wikipedia.org/w/index.php?title=Alfabetizaci%C3%B3n_inform%C3%A1tica&oldid=125341516)

Wikipedia (2019). Talk: Information Literacy. Retrieved: 2020-06-04, from: [https://en.wikipedia.org/wiki/Talk:Information\\_Literacy#Proposed\\_merge\\_with\\_Data\\_literacy](https://en.wikipedia.org/wiki/Talk:Information_Literacy#Proposed_merge_with_Data_literacy)

Wikipedia contributors. (2020). Data literacy. In Wikipedia, The Free Encyclopedia. Version: 2020-04-15. Retrieved 2020-06-04 15:40 UTC, from: [https://en.wikipedia.org/w/index.php?title=Data\\_literacy&oldid=951191278](https://en.wikipedia.org/w/index.php?title=Data_literacy&oldid=951191278)

Wikipedia contributors. (2020). Data literacy. In Wikipedia, The Free Encyclopedia. Version: 2020-01-23. Retrieved 2020-02-11 18:43 UTC, from: [https://en.wikipedia.org/w/index.php?title=Data\\_literacy&oldid=937273215](https://en.wikipedia.org/w/index.php?title=Data_literacy&oldid=937273215)

Wikipedia contributors (2020). Datenkompetenz. In: Wikipedia, die freie Enzyklopädie. Version: 2020-05-10, 2020, 15:46 UTC. Retrieved 2020-06-17 16:42 UTC: <https://de.wikipedia.org/w/index.php?title=Datenkompetenz&oldid=199810749>