Hochschulforum Digitalisierung

NO. 53 / JULY 2020

# Future Skills: a Framework for Data Literacy

## Competence Framework and Research Report

Katharina Schüller · STAT-UP

NO. 53 / JULY 2020

# Future Skills: a Framework for Data Literacy

## Competence Framework and Research Report

**Author**
Katharina Schüller · STAT-UP

**Illustrations**
Markus Armbruster · pictomind

# Table of Contents

# List of Figures



# List of Tables

# Hochschulforum Digitalisierung (HFD)

The Hochschulforum Digitalisierung (HFD) orchestrates the discourse on higher education in the digital age. As an innovation driver, it informs, advises, and connects stakeholders from higher education institutions, politics, business, and civil society.

Founded in 2014, HFD is a joint initiative by Stifterverband, CHE Centre for Higher Education and the German Rectors' Conference (HRK). It is sponsored by Germany's Federal Ministry of Education and Research (BMBF).

Further information is available at https://hochschulforumdigitalisierung.de/en.

A group of experts at Hochschulforum Digitalisierung conceptually accompanied the call for proposals and this working paper:

Arne Gerdes (University of Göttingen),
Priv.-Doz. Dr. med. Sebastian Kuhn, MME (University of Mainz),
Prof. Dr. Antje Michel (University of Applied Sciences Potsdam),
Prof. Dr. Tobias Seidl (Stuttgart Media University).

The expert group emerged from the Curriculum 4.0 working group. The aim of the working group was to identify overarching elements of curriculum development against the background of digital change and to demonstrate good practice approaches. The working group has developed a series of discussion and recommendation papers that universities can use in the revision process of their curricula to reflect on their course concepts as well as for concrete curriculum design. More information about the working group can be found at https://hochschulforumdigitalisierung.de/en/themen/curriculum-40.

Further contributions came from Paulina Busch, Carina Hindinger (research support, both STAT-UP), and Anne Koark (translation).

# Abstract

What knowledge, skills, and attitudes are needed in society, in the working world and in science, where data is considered a valuable, sometimes the most valuable resource and where decisions are increasingly taken based on data? Digitisation and datafication will undoubtedly change life and work in the 21st century. Artificial intelligence, networked production, communicating machines and self-driving cars are controlled by data and themselves produce data on the conveyor belt. Data is the starting point for knowledge and value creation as the basis for better decisions.

The process of knowledge creation comprises several steps: (A) Establish a data culture – (B) Provide data – (C) Evaluate data – (D) Interpret results – (E) Interpret data – (F) Derive actions. In order to systematically gain knowledge or value from data, the ability to deal with data in a planned manner and to be able to use and scrutinise it in the respective context will be of crucial importance in all sectors and disciplines. This is known as data literacy and includes the skills to collect, manage, evaluate, and apply data in a critical manner. Data literacy is far more than wide and extensive knowledge of constantly changing methods and technologies. Rather, the dimension of data ethics, motivation and value plays a central role in handling data successfully and confidently in the future.

Data literacy is a key competence of the 21st century. Therefore, it must be taught from the very beginning and across the disciplines at universities. This requires a competence framework, i.e. a model for the structured description of effective behaviour in each task context. It includes competencies, their definitions and behaviour indicators derived from them. Such a competence framework should map all stages of the knowledge or value creation process from data; it should cover all dimensions of competence: (a) knowledge, (b) aptitudes, (c) skills, (d) motivation and attitude. It should allow the competences to be translated into concrete and testable learning or competence goals; and it should reflect the interdisciplinarity of the task, i.e. reflect that, in addition to data experts, technical experts, data protection experts and data ethicists are also required.

Data literacy covers both the creation of data products by the methodically experienced specialist as well as the competent handling of data by the end user, i.e. the critical and adequate interpretation and application of the data. The Data Literacy Framework is designed primarily for universities and other educational institutions, but also for managers and human resources departments in private and public sector organisations and for political decision-makers. In addition, it can be used to formulate data literacy for the responsible citizen, whereby in this case rather low competence levels should be aimed for.

It is important to translate the results into proposals for curricula. For this, pilot universities and pilot courses must be selected, in which specific learning objectives for the respective disciplines are established in accordance with the competence framework. Another aspect to be discussed is the aspect of lifelong learning of key competencies: How can or should data literacy be taught at school and later in the working world and adult education? In any case, one needs didactic approaches regarding the interdisciplinarity of the task. Because complex data projects are executed in teams, they require multi-professional work that includes the ability to manage projects as well as in-depth knowledge of organisational, legal, and ethical framework conditions. Last, but not least, the question of how teachers can be qualified for this challenge needs to be answered.

# Chapter 1: Background & Objective

## 1.1 Starting Position and Objective

Almost a decade ago, Google's chief economist Hal Varian predicted in an article in the New York Times: "I keep saying that the sexy job in the next 10 years will be statisticians" (Lohr, 2009). In the same article, Erik Brynjolfsson, economist and director of the Massachusetts Institute of Technology Center for Digital Business, was cited: "We're rapidly entering a world where everything can be monitored and measured, but the big problem is going to be the ability of humans to use, analyse and make sense of the data."

But how do we get from data to meaningful action? Data is just raw material. Data represents abstractions of 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 (Foucault, 1980). How exactly this process works, whether it is a linear sequence or a cyclical one, where the starting point is and whether all process steps must always be carried out, remains open for now.

One thing is certain: Data is ubiquitous today; digitisation inevitably leads to datafication as the analogue world is mapped into a digital one. In almost all disciplines, data is collected in large quantities or arises as a by-product, used for ongoing monitoring, planning, control and evaluation. Examples include sensor data in production, digital insurance data, data from fitness trackers, traffic flow data, customer data in CRM systems and many others. That data is increasingly no longer used only for its original purposes, which were related to the optimisation of existing processes. Far beyond that new analysis, options arise that, for example, lead to the development of innovative business models.

In addition, data and its derivatives, such as statistics or charts, penetrate all traditional and new media as a result of the development of (more or less professionally conducted) data journalism. It is verbalised, contextualised, and must be extracted and rearranged by the recipient, if the interpretations of the producer are not to be adopted uncritically. Studies are no longer only carried out by universities, established research institutes and consulting firms. Rather, the democratisation of data access (e.g. through open data) and data processing (e.g. through intuitively usable tools such as the Google product "Data Play") also enables laypersons to draw (false) conclusions from data and use it on portals such as medium.com and social networks such as Facebook to present them to a broad audience and to contribute to the formation of knowledge or even opinion.

The study "Future Skills: A Framework for Data Literacy" (Schüller, et al., 2019) pursues the goal of developing a competence framework for digital competences using the example of data literacy and to make practical knowledge available for universities.

HFD Working Paper 37/18 (Heidrich, et al., 2018) formulates a competence matrix based on (Ridsdale, et al., 2015) with five areas of competence, individual competences and an evaluation according to the 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-competencies of digital literacy (European Commission, 2018). In this context, the terms data and information literacy are not differentiated from each other, and some skills that are defined as core competencies in (Heidrich, et al., 2018) are missing.

In contrast, the ProCivicStat project developed a far more extensive conceptual model (Nicholson, et al., 2018). 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. 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 special skills (data ethics, measurability of phenomena, classification of results in contextual knowledge, derivation of recommendations for action). Since the examinations that students must take are the strongest indicator of what they are learning, corresponding tests must also be developed for the "soft" competences.

## 1.2 Definition of the term "Data Literacy"

To carry out the process of transforming data into information, knowledge, and wisdom/"power", i.e. action knowledge and capacity to act in a controlled and responsible manner, data literacy is required. This competence enables people to address real problems through the use, analysis and interpretation of data that measures underlying phenomena. It is essential for understanding complex, socially relevant phenomena such as global economic and financial interdependence, migration, or climate change. In the age of digitisation, we not only have more data to measure these phenomena, but also possibly more inappropriate data that needs to be carefully evaluated. In addition, digitisation enables far more people than before to publicly discuss and influence these phenomena.

According to (Weinert, 2014), competencies are "the cognitive abilities and skills available in individuals or learnable through them to solve certain problems, as well as the associated motivational, volitional and social willingness and ability to solve problems successfully and responsibly in variable situations". This reference definition thus includes not only knowledge and the ability to apply this knowledge (skills) but also the willingness to do so (values).

With good reason, data literacy is regarded as a "central competence for digitisation and the global knowledge society in all sectors and disciplines" in (Heidrich, et al., 2018). It is important to systematically integrate this competence into education, especially into the curricula of universities.

Ridsdale's definition used in (Heidrich, et al., 2018) specifies data literacy as "the ability to collect, manage, evaluate and apply data in a critical way" (Ridsdale et al., 2015). The study also concludes that data literacy can serve various purposes. On the one hand, data literacy is needed to use data professionally within a specific subject area. This includes not only research at universities, which increasingly have access to new data sources ("big data") and tools developed for them from the discipline of "data science", but also research and development in companies. There, data serves as a raw material for innovations; it is considered the "oil of the 21st century".

On the other hand, data literacy can be understood as the ability that a responsible citizen needs in today's society to find his way through an overabundance of data and information and to make informed decisions – in everyday life and at various political levels. This ability to make decisions requires the ability to distinguish data and information from interpretations and opinions. "We are drowning in information and thirsting for knowledge," as futurologist John Naisbitt puts it (Naisbitt, 1982, p. 24), and science fiction writer Herbert G. Wells is (falsely) credited with the following prognosis: "If we want responsible citizens in a modern technological society, we have to teach them three things: Reading, writing and statistical thinking, that is, the rational management of risk and uncertainty." (Tankard, 1979)

Not only data literacy, but also "information literacy", "statistical literacy" and recently also "data science literacy" are terms that are repeatedly mentioned as essential competencies in relation to digitisation. Obviously, the delimitation of the terms is by no means conclusively clarified. Furthermore, there is the danger of definitions being deliberately extended because "data science" is becoming a fashionable term. There may also be an economic interest behind such extensions: Salaries paid on the job market in new professions such as "data engineer" or "data scientist" are considerably higher than those of statisticians or IT experts.

Even the question of how concepts of data, information and knowledge should be distinguished can no longer be answered so clearly, at the latest since "big data" also includes images or entire texts under the data term. Last but not least, the rapid development of technologies and tools for the automated collection, storage and analysis of such semi- and unstructured data contributes to the fact that these are fluid concepts: What is already information in the first context, i.e. data filled with meaning, may be a piece of data in the second context – it depends on the research question.

## 1.3 Requirements for a Data Literacy Competence Framework

This aspect becomes even clearer with Rainer Kuhlen's definition of information as "knowledge in action" (Kuhlen, 2013). For the design of a competence framework and the derivation of learning objectives for the teaching of data literacy – which is recommended as the next step according to the present study – first of all (at least) information literacy and data literacy should therefore be considered convergent. All stages of the process of creating value from data or decision-making with data must be mapped.

Secondly, the multidimensionality of the concept of competence must be covered by a competence framework. (Seidl, et al., 2018) call for the development of 21st Century Skills at universities to "create a framework in which (a) this complex knowledge, (b) skills, (c) abilities, and (d) motivational orientation and values can be acquired and developed". In view of the changing competence profiles during digital change, the dimension of motivation and attitude must be given a central role.

Thirdly, a competence framework should create the basis for the competences recorded there to be translated into concrete and testable learning or competence objectives. This applies "not only in the area of knowledge and skills/abilities, but especially in the area of motivational orientation and values" (Seidl, et al., 2018). For this purpose, the competences must be sufficiently operationalised and formulated, ideally with an initial classification into competence levels or performance levels. This is a prerequisite for selecting and developing measurement and testing instruments for data literacy.

Fourthly, the competence framework should consider the fact that the tasks to be solved with the help of data are increasingly complex and interdisciplinary. The transition areas between subject and methodological disciplines are becoming blurred and are also becoming the focus of legal and ethical considerations; it is not (any more) so easy to collect, process and use data, especially when third parties are to be involved in the evaluation. With the growth of data as a resource and its ever more intensive use, new professions such as data engineer, data scientist, data designer or data journalist are emerging. It is important to develop a picture of how these professions will interact in the future to understand which competencies should be taught across and which subject-specific, and who should be involved.

# Chapter 2: Developing a Competence Framework

## 2.1 A Competence Definition based on the knowledge-creating process

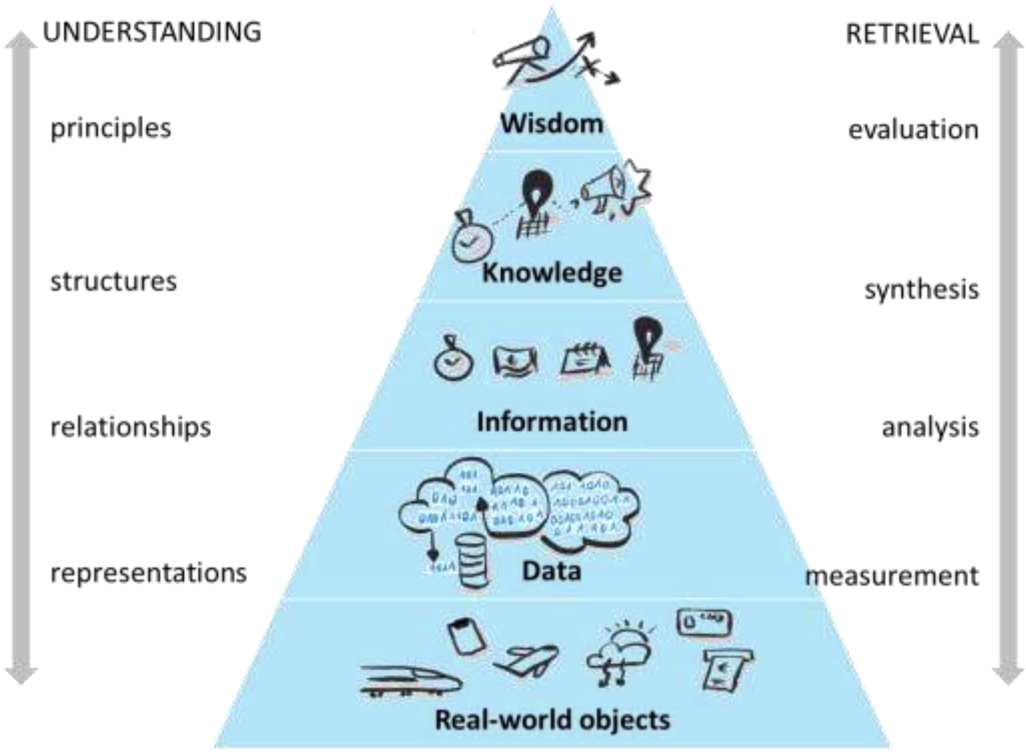
"To collect, manage, evaluate and apply data in a critical way", in Ridsdale's words, defines a process that requires data literacy to master (Ridsdale et al., 2015). It therefore makes sense to arrange the necessary competences based on the individual process steps of this "knowledge creation". In information literacy, a step-by-step model such as the DIKW model (Data, Information, Knowledge, Wisdom) is often chosen for this purpose. It shows schematically how raw data is processed into information, knowledge, and wisdom in the human brain, following an increasing degree of organisation. After cleansing and linking individual data elements to form meaningful information, we search for patterns, apply analysis principles and structure the information, for example by classification or categorisation. The measurement process is preceded by the mapping of real-world objects into data.

With the conceptual development of data (science) literacy and statistical literacy the perspective shifts towards a cyclical representation of the process (The Association of Independent Schools of New South Wales, n.d.). This representation emphasises the integration of data analysis into a concrete research question or decision situation, whereas the established statistics education at universities focuses on the acquisition of expert knowledge and the learning of methods. Nevertheless, the cycle depicted is in the tradition of established empirical research, which starts from one or more research questions ("science starts with a question") and models a hypothesis-driven or confirmatory approach. To answer that question, data is acquired, prepared, evaluated, and interpreted according to the respective purpose of analysis.

Digitisation, however, creates data "purposelessly" in large quantities and heterogeneity and even raises new questions ("data science starts with the data"). Data-driven or explorative procedures are becoming increasingly widespread in practice, so that new skills are needed in dealing with new forms of data – text, sound, image – beyond the familiar scale levels and forms of storage. Google describes its product "Data Studio" with the words: "Unlock the power of your data with interactive dashboards and beautiful reports that inspire smart business decisions. It's easy and free." (Data Studio, n.d.) This emphasises the explorative approach of "letting the data speak for itself". The risk of drawing wrong conclusions is high if there is a lack of a basic understanding of statistical errors such as the confusion between correlation and causality, or if there are unrecognised shortcomings in the quality of the data (e.g. a bias). Finally, ethical literacy is required when data is freely combined and analysed for purposes other than its original purpose of collection.

(Münster, 2019) proposes an equally cyclical process model in a visualisation, which schematically depicts visualisation literacy as productive coding competence and visual literacy as corresponding receptive decoding competence. It particularly considers the context from which data is derived or which is to be concluded based on the data. The model is obviously related to the model of statistical learning according to (Wild, 2006).

(a)


UNDERSTANDING
principles
structures
relationships
representations
Wisdom
Knowledge
Information
Data
Real-world objects
RETRIEVAL
evaluation
synthesis
analysis
measurement

(b)


Decision making and communication
Framing the issue and asking the right questions
Data informed decision making cycle
Data interpretation
Planning the analysis
Data analysis
Data collection

(c)

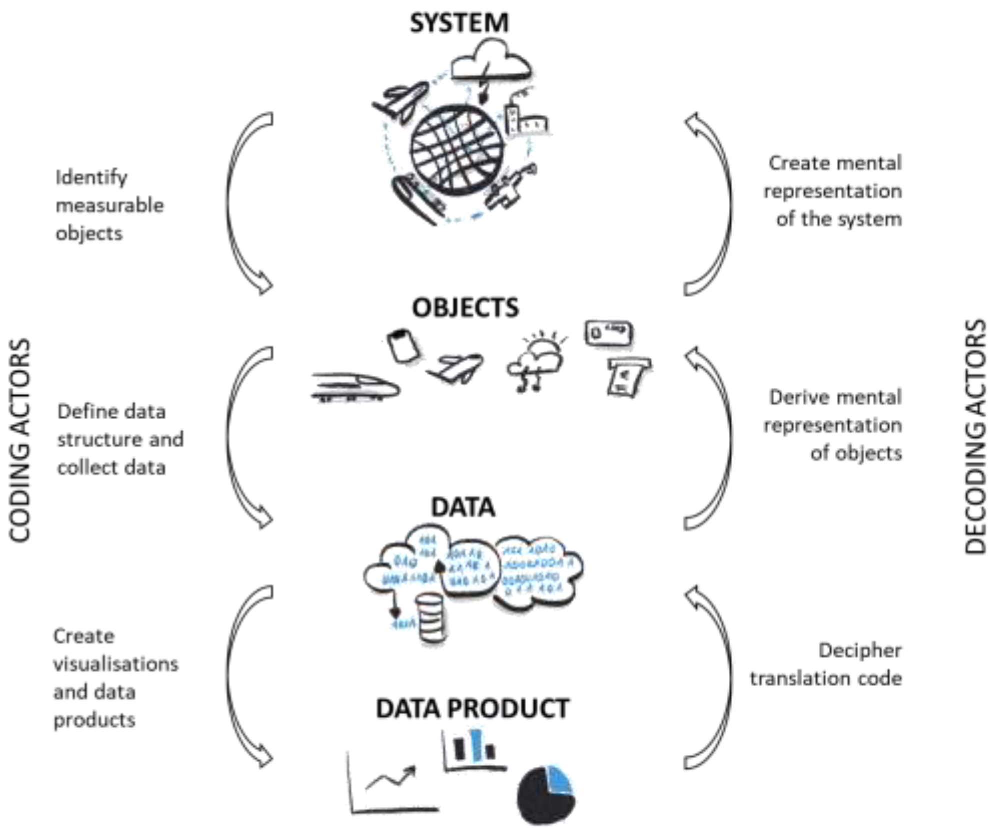



Figure 1: (a) Pyramid model, (b) cyclical model, (c) structural model of the process of data value creation (based on (a) (Awad & Ghaziri, 2004) & (Bellinger, et al., 2004), (b) (The Association of Independent Schools of New South Wales, n.d.), (c) (Münster, 2019)

The three process models are shown in Figure 1.

Any distinction between the term "data literacy" and concepts such as "information literacy", "digital literacy" or even "statistical literacy" is to a certain extent arbitrary and the transitions between the underlying concepts are fluid. Even if one attempts to assign sub-components, such as visualisation literacy or computational literacy to different competence profiles, overlapping remains.

Based on the literature, ethical literacy can be understood as a meta-competence. It is listed as an element of every other literacy. For example, statistics should not be deliberately manipulated in order not to tempt their users to make certain decisions ("nudging"). Even the accurate analysis of data can be ethically questionable, if it leads to discrimination, for example in the different pricing of health insurance tariffs depending on gender. The view that data literacy also requires ethical literacy is therefore based on the conviction that the collection, usage, processing, and analysis of data cannot be carried out independently of the context, i.e. separately from its interpretation and application.

***Ethical literacy in the context of data literacy is the ability to fully grasp the significance of data for decision making by reflecting and critically evaluating possible interpretations of these data in different contexts.***

Information literacy is usually understood as an overarching, receptive competence, since the competences typically assigned with refer to the general handling of information, not only to data or statistics. Data literacy and statistical literacy were defined as sub-areas, although there is some overlap. Data literacy primarily describes the handling of data, i.e. the acquisition and storage of data, data manipulation or data processing and data analysis for the purpose of transforming data into knowledge – i.e. the process of producing knowledge from data. Statistical literacy, on the other hand, primarily describes the handling of existing statistics and the ability to interpret them, i.e. the receptive process. Above a certain level of competence, both concepts can no longer be regarded independently of each other. The planned production of knowledge from data requires an understanding of the possible reception. Conversely, data-based decision making requires an understanding of the data-generating process, i.e. the origin of data including possible limitations, as well as the analysis tools used.

Further sub-competencies can be formulated within the two competences. In recent years, visual literacy or graph literacy has become increasingly important. It describes the ability to create and interpret visualisations correctly. Thus, it is part of both data literacy and statistical literacy. Model literacy shows similar overlaps. It is understood as the competence to know the different statistical models regarding their areas of application, strengths, and weaknesses and to apply and interpret them correctly. Finally, computational literacy includes the ability to use computers in general and programming skills. It is primarily to be assigned to data literacy. Occasionally statistical literacy can also require certain competences of computational literacy, for example, to assess the influence of a numerical simulation or optimisation method used on an analysis result.

The present study follows the approach of locating concepts of competence from a process perspective and thus represents the view of competences as a cluster of effective behaviours and attitudes towards the fulfilment of a delimitable task, i.e. a task context. The starting point is a combined process model that takes Münster's structural model as its basis and integrates elements of the pyramid model (differentiation between data, information and knowledge) and elements of the cycle model (process steps as competence fields). The concepts of competence can then be characterised in terms of the properties of the underlying task or process steps for the completion of this task:

Competences may enable one single process step (selective) or several process steps (comprehensive). The dimension is shown as "inside – outside" in the following Figure 2.

Competences may enable productive process steps (coding) or rather receptive process steps (decoding). The dimension is shown as "left – right" in the following Figure 2.

- Competences may enable process steps where knowledge is the focus of attention, or those where data is the focus of attention. The dimension is shown as "top – bottom" in the following Figure 2.

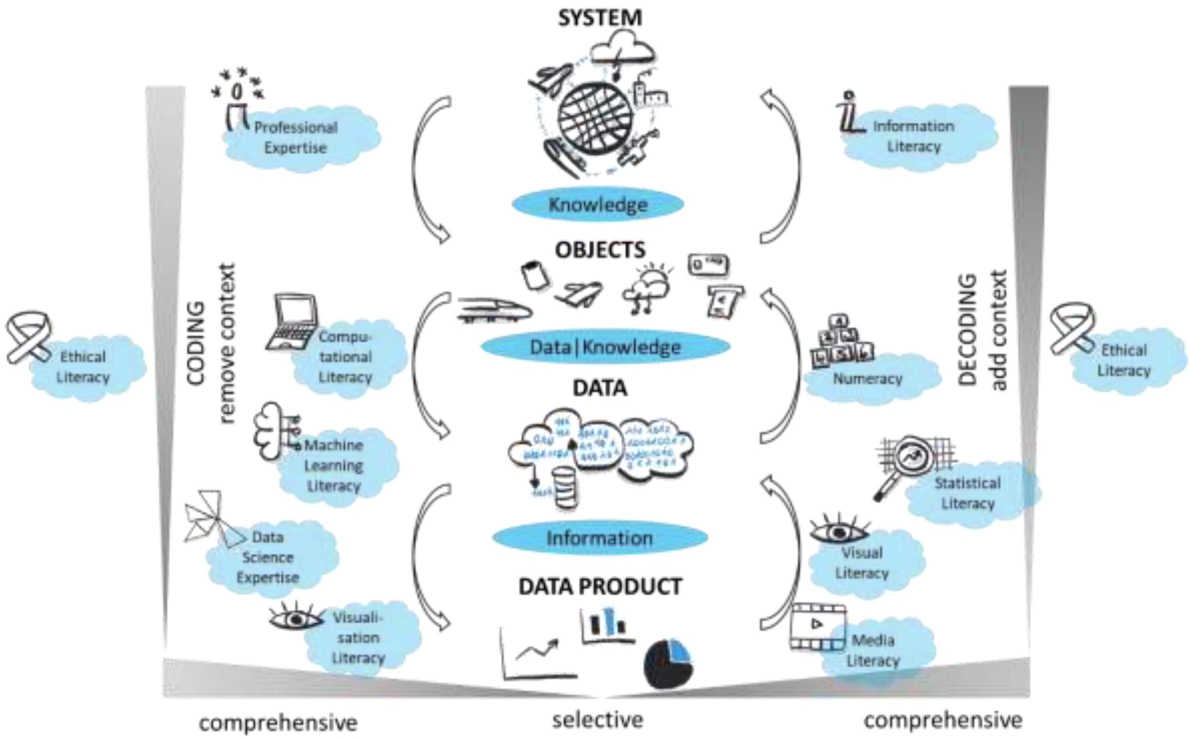



Figure 2: Classification of competence terms on the three-dimensional, integrated process model for data value creation

Visualisation literacy (in the literature sometimes also called graph literacy) describes a competence, which is selectively assigned to the process step of the production of data visualisations, while visual literacy is correspondingly assigned to the process step of the reception, i.e. the interpretation of data visualisations.

For decades, information literacy was defined primarily as receptive literacy and classified accordingly. Due to the changing production conditions of scientific publications (Web 2.0, Open Science, Open Access) and the associated erosion of the "gatekeeper function" of publishers, a producing dimension in information literacy is increasingly being considered and respected.

Data science literacy is here referred to as data science expertise, to specify it as the competence profile of a specific profession (professional expertise) and not as an interdisciplinary competence. This follows Kitchin's opinion (Kitchin, 2014). Ethical literacy takes an outstanding role as a general competence.

The present study further defines the term data literacy very comprehensively:

> ***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 implementation contributes to the goal of 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 goal with the least possible use of resources, answering the question of how the goal can be achieved in the best possible way ("doing things right"). (Seidl, et al., 2018) emphasise the importance of attitudes, values and ethics for key competences.

## 2.2 Attitudes, Values and Ethics as a Competence Dimension

A competence framework is a model for the description of effective behaviour in each task context and comprises competences, their definitions and derived behaviour indicators (i.e. operationalisation).

Competence frameworks often follow a thematic order; for example, the OECD Competence Framework groups various key competences into competence categories (OECD, 2005). The European eCompetence Framework (e-CF) (European Commission, 2016) arranges competences in fields of competence, defining levels in each case in accordance with the European Qualifications Framework for Lifelong Learning (EQR) (European Commission, 2008) and formulates examples of knowledge and abilities or skills. For the competence field "planning", for example, knowledge of "concepts for business strategies" is given as an example of knowledge and the ability to "contribute to the development of the business strategy" is given as an example of 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). (Deutscher Qualifikationsrahmen, n.d.) (DQR), 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 attributed to someone who "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 (European Commission, n.d.).

So far it remains ambiguous whether "knowledge", "skills" and "values" represent different dimensions of a competence or whether they are different categories of competence. The latter view is reflected in the EQR and DQR, the former in the KSAVE model and is also represented in this study (Heidrich, et al., 2018).

### 2.3 The Framework in the Context of Existing Literature

An overview of the different competence frameworks defined in scientific articles is presented in (Schüller & Busch, 2019). In addition, several data handling guidelines have been consulted to define the individual competences and levels more precisely.

A closer look at the competence profiles reveals that orientation based on the individual analysis steps is a common procedure. First, competencies are considered to arise before the analysis, i.e. during data collection. During the analysis other competences are needed, which often requires a deeper knowledge of statistical models. Less attention is paid to the competences for communicating the results. But they are essential: a wrong or misleading communication of data and statistics can render the entire analysis obsolete.

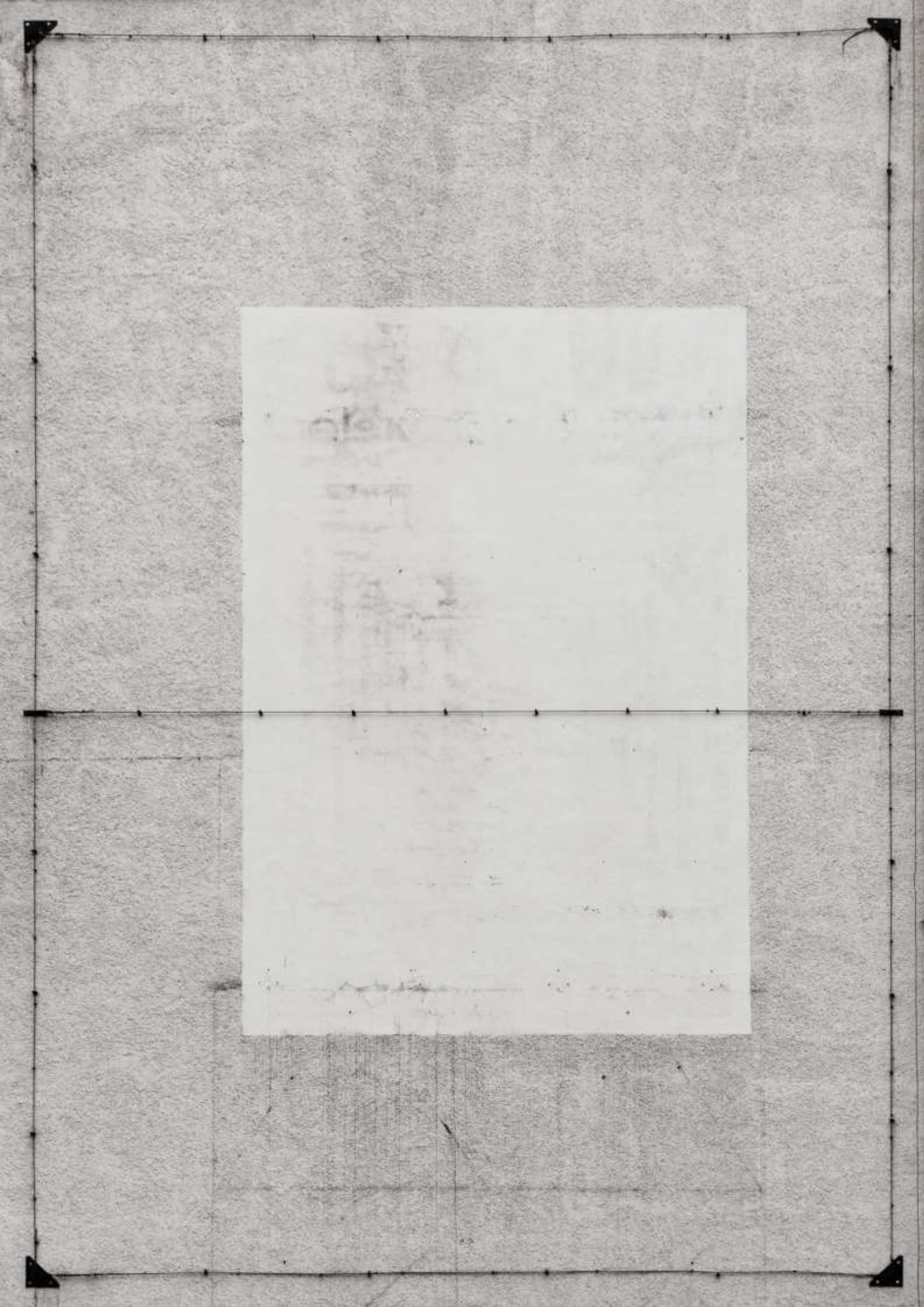
In addition to the competences of the analysis steps, some more general competences are described, which deal with data storage or ethical aspects, for example. These competencies are relevant for all analysis steps. Less attention seems to be paid to the various tools that are required for the analysis steps and competencies on the side of the productive process steps. However, mastery of these is an essential skill and must therefore be considered at various levels when describing the individual competences. For example, a higher level of competence is required if an analysis is to be carried out using an expert tool such as R instead of Excel. Furthermore, a detailed knowledge of the different programs and the implementation of the algorithms increasingly replace a detailed knowledge of the exact formulas of the methods. While in the days of slower computers it was inevitable to specify a linear regression manually or even to calculate it by hand, today appropriate user interfaces simplify this. Although a rudimentary knowledge of the functionality is important, it is more important today to have a good overview of the many different techniques. The number of possibilities has grown considerably due to faster machines, so it is important to be able to estimate exactly which method will yield the best and most reliable results for which data set and for which problem.

In general, little attention is paid on the part of the receptive process steps to the interpretive capacity that individuals possess in relation to statistics and other data products. This could be a result of the conceptual separation of data literacy and statistical literacy. As already mentioned, this separation no longer seems appropriate. To create a common competence profile, competence frameworks and competence profiles for statistical literacy were therefore also researched. A frequently cited framework was defined by (Watson & Callingham, 2002) and further developed (Watson & Callingham, 2004). It describes 6 levels of competence for interpreting and handling data. At the lowest level, idiosyncratic, one has only the most rudimentary statistical and mathematical skills. At the next level, one already has knowledge of individual statistical terms and can interpret basic tables, graphs, and probability calculations. The third level requires the recognition of conclusions, although still without justification or qualitative use of statistical terms. In the following level a wide range of terms can be used in a differentiated way and statistical skills related to mean values, probabilities and graph properties are learned. In the second highest level one can question simpler contexts that do not have too complex terminology and methods. At the highest level, one can question any text and interpret even the smallest aspects of statistical language.

Based on the literature and monographs of (Kitchin, 2014) and (Linoff & Berry, 2004) the framework is also divided into main categories (areas of competence) and subcategories (competencies), which are particularly inspired by (Matthews, 2016). There categories are divided into data acquisition, which takes place before the analysis, data analysis and data communication, which is important after the analysis. In addition, general competences in data management, as well as ethical, legal, and social competences, which are set on a meta-level. For each field of competence individual competences are defined, depending on which different levels of competence can be achieved. These are subdivided into competencies that are required on the producer side, i.e. in the coding process, and competencies that are required on the recipient side, i.e. in the decoding process. The ethical competencies, which are often considered separately in the literature, are understood as a competence dimension on both sides, i.e. they represent an essential facet of each individual competence in dealing with the data and comprise motivation and values. Data protection and data security aspects, on the other hand, are included in a competence in the competence field "Providing data", since they are primarily required there.

The levels outline the degree of complexity of the task corresponding to the respective process step that can be mastered by an individual. While the first level requires only basic knowledge and skills in simpler methods, the highest level can be compared to that of a data scientist. It should be noted that it is not the aim that every individual reaches the expert level. Instead, depending on the discipline and the individual's points of contact with data, a level of competence to be defined there should be mastered. It is crucial to understand the limits of one's own competence and to be prepared to consult the opinion of experts for more complex tasks.

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 (eCompetence Framework) or ESCO (European Skills and Competencies of Occupations). Consequently, the derivation of subject- or course-specific learning objectives, measurement or testing instruments requires considerable further development effort. There is a clear need for additional research, which should be based on the concepts developed in this study.



# Chapter 3: Description of the Competence Framework

## 3.1 Intention and possible Applications

The competence framework, hereinafter referred to as the "Data Literacy Framework", consists of 18 competencies that are used in the process of creating value from data or making decisions from data. It is based on a uniform, cyclical process model that divides the individual process steps and the corresponding competencies into productive and receptive steps. The Data Literacy Framework deliberately avoids understanding data literacy as data science expertise (often misleadingly referred to as data science literacy). The latter is primarily assigned to the fields of statistics/mathematics/computer science, emphasises the technical side of the competencies and, from the author's point of view, corresponds to a profession-specific competence profile.

## 3.2 Outline Levels of the Competence Framework

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**: 6 fields of competence, derived from the process steps: (A) Establish a 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 productive process steps from system to data to data products, the competence fields (D) to (F) correspond to the receptive process steps from data products to data to system.

**Outline level 2:** Essential competences for each field, each with a generic description. A further subdivision is made in competence field (B).

**Outline level 3**: Examples of knowledge, skills/aptitudes, and motivation and attitudes. The dimension "knowledge" refers to the knowledge needed to master the process step. This dimension deals with the (complex technical) knowledge. The dimension "skills" describes the aptitudes and skills needed to master the process step. These aptitudes or skills describe the application of knowledge, i.e. the abstraction of what has been learned. The dimension "attitudes, values, ethics" (AVE) describes the motivation and values that an individual should possess. It describes the ethical requirements, for example to ensure a certain degree of objectivity and to exclude the misuse of data and analyses. In addition, it is about motivation, openness, and the willingness to learn from mistakes.

The subdivision into these three dimensions is intended to provide a comprehensive picture of the competence. 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**: Level specifications that represent a simplification of the EQF levels. EQF levels (EQF = European Qualifications Framework) 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.

### 3.3 From a Competence Definition to the Competence Framework

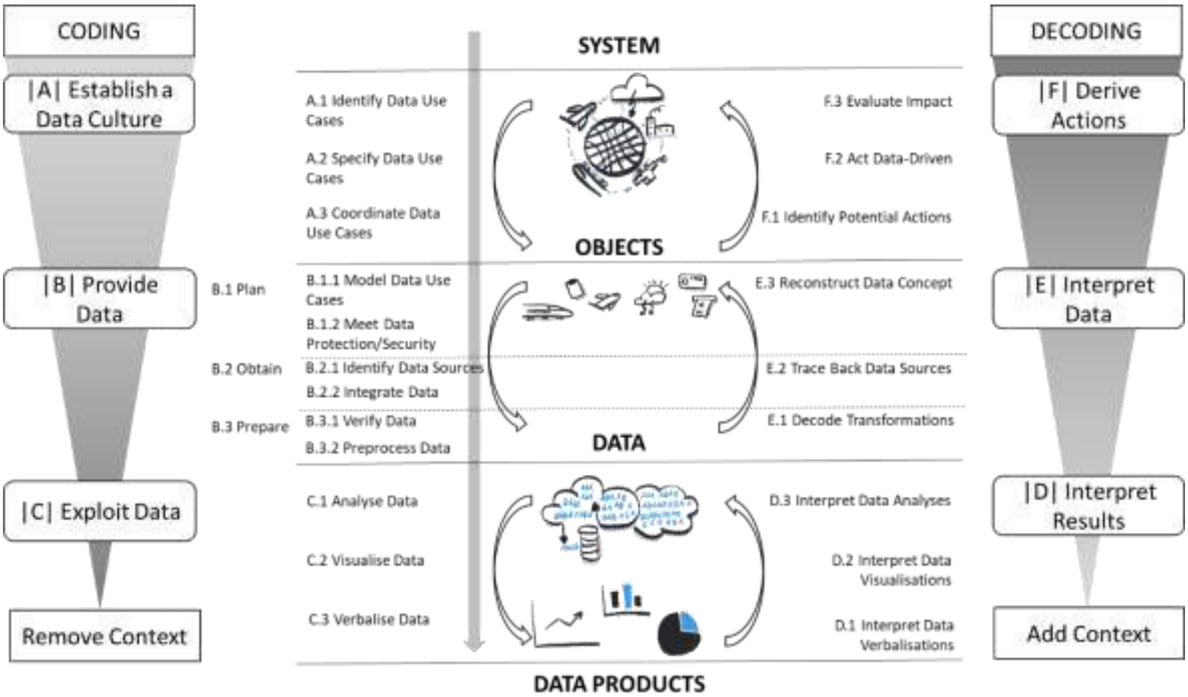



Figure 3: Competence fields and essential competences of the Data Literacy Framework in the process of value creation from data

The Data Literacy Framework results from a detailed consideration of the individual process steps. Following the definition from chapter 2, data literacy is 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. This is based on existing literature, which was researched, analysed, and evaluated in the systematic review accompanying the study (Schüller & Busch, 2019). In addition, practical experience from many years of consulting experience in the professional field of "data science consulting" was incorporated.

The competence fields "acquiring data", "evaluating data", "interpreting results" and "interpreting data" with the respective competences mentioned there are typically included in the Data Literacy Framework developed so far. The last two are often combined and not presented in such detail.

(Mandinach & Gummer, 2013) define data literacy as the ability to understand and effectively use data to support decision-making. The skills they define cover the entire value chain by addressing the handling of data on the one hand, but also by focusing on defining hypotheses and deriving options for action, implementation, and impact analysis on the other.

The core competencies for "data information literacy", as defined by Carlson et al. (Carlson, et al., 2013), deal in detail with the skills and aptitudes that arise on the coding side and consider ethical behaviour as a separate, important core competency. As this framework was specifically defined for working with research data, the decoding side is not included here. The twelve core competencies that are listed are shown (Carlson, et al., 2013).

Table 1: Core competencies from (Carlson, et al., 2013)

| | |
|---|---|
| Data processing and analysis | Data transfer and reuse |
| Data management and organisation | Data conversion and compatibility |
| Data protection | Data visualisation and representation |
| Databases and data forms | Finding and acquiring data |
| Ethics and classification | Metadata and data descriptions |
| Data quality and documentation | Process options |

Sternkopf (Sternkopf & Mueller, 2018) deals with the competence field "Establishing a data culture" in its Data Literacy Maturity Model. There are also references to this in (Kitchin, 2014), (Goldsmith & Crawford, 2014) and (Schüller & Wrobel, 2018). (Kitchin, 2014) as well as (Williford & Henry, 2012) especially emphasise the interdisciplinarity and the resulting necessity of data project management. This includes the knowledge and mastery of appropriate tools (e.g. Git). Using a practical example, (Schüller & Wrobel, 2018) also address the competence field of "deriving action" and the attitude it requires to establish data-driven decisions during digital transformation in an organisation.

The decision to integrate values and motivation in this framework as a competence dimension was taken because data analysis and interpretation – in contrast to mathematics or programming – do not represent a one-to-one transformation. On the way from the system to the data product, context knowledge is continuously removed (extraction of "essential" information), and on the way back context knowledge is continuously added (interpretation and contextualisation). An essential learning progress in the dimension of values is getting aware of the evaluation process throughout that context removal and addition and the risk of evaluation failures, accidentally or on purpose. This also includes the awareness that resources are being used and potentially wasted: To create data products, manpower, physical energy, and resources are invested that may not belong to you.

On the production side, a basic ethical attitude means evaluating what the data really contains, what cannot be assessed and when nothing at all can be assessed (i.e. where grey areas lie). To be aware that the data analysis is carried out for a certain purpose (and that this inevitably involves certain interests to which one is obliged under certain circumstances) is an expression of attitude and the ability to reflect.

On the side of reception, a critical basic attitude means being aware that one's own previous knowledge and values are inserted into the interpretation and that both are subjective. In addition, it is important to take in that there are subtle possibilities of manipulation by the producer and how these should influence one's own interpretation. Part of the attitude is also to make oneself aware that one has to deal with one's own inner resistance if the data does not deliver the desired result – and that one should therefore not immediately give in to the impulse to reject the results. Conversely, the same applies if results confirm one's own (pre-)judgements – still they should be questioned critically. The responsibility for decisions under uncertainty cannot and must not be delegated to data analysis.

## 3.4 The Competencies in Detail

**Field of competence A: Establishing a data culture – from systems to measurable objects**

This includes the intention to establish a data culture. In the first stage is the identification of data applications. Here, gaps in knowledge and background information are to be identified, which provide the basis for a concrete task. In the second stage, for example, requirements are defined and communicated to experts and the application is distinguished from other tasks. This stage specifies the data applications. The third level comprises the planning and coordination of the data application.

Table 2: Field of competence A: Establishing a data culture

| Competence | | Dimensions | | | Levels |
|---|---|---|---|---|---|
| **Labelling** | **Description** | **Examples of knowledge** | **Examples of skills** | **Examples of attitude** | **Examples of ascending levels** |
| A1: Identifying data application | 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 | Deep theoretical and practical knowledge of the field of application or the discipline and, if applicable, related disciplines<br><br>Knowledge of the relevant literature or professional requirements (norms, rules, | Skill to identify and assess relevant gaps in knowledge regarding the potential value of data analysis<br><br>Skill to separate relevant from irrelevant information about the system | Openness to answer specific questions with the help of data and its analysis<br><br>Willingness to learn from data<br><br>Willingness to question existing rules and processes | (1) Identifies obvious use cases, e.g. by copying questions from the literature based on typical problems and data sources<br><br>(2) Identifies unusual applications, for example by creative recombination of known applications |

| | | | | | |
|---|---|---|---|---|---|
| | contribution of the data | quality standards, processes, restrictions)<br><br>Knowledge that and how data and its analysis can be used for decision-making in their own discipline | | Willingness to admit and accept non-knowledge | (3) Identifies disruptive use cases, for example by anticipating new technologies, data sources, user groups |
| A2: Specifying data application | 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 | Knowledge about concrete practical processes, participants, and effect models<br><br>Knowledge about problems that can result from a wrong understanding of the system<br><br>Knowledge of the information that a data expert needs to measure and model the system | Skill to break down complex tasks into individual elements and to describe these elements and their relationships<br><br>Skill to formalise uncoded or only partially codified expert knowledge about systems, their elements and interaction in (hypotheses about) rules<br><br>Skill to communicate the knowledge required to understand a task in an interdisciplinary team or to a (data) expert | Willingness to share knowledge with others<br><br>Willingness to trust the results of data analyses, even if they are not understood in every detail<br><br>Willingness to accept the limits of the significance of data analysis | (1) Can specify simple problems with few, clearly identifiable elements and few direct relations<br><br>(2) Can specify complex problems with partly nested elements and partly indirect or non-linear, static relationships<br><br>(3) Can specify highly complex problems involving a variety of complex objects, dynamically changing relationships and feedback |
| A3: Coordinating data application | Plans and coordinates a data project, if necessary, with participation of further persons (from interdisciplinary areas) | Knowledge about the capabilities (strengths/weaknesses) of employees<br><br>Knowledge about the expertise (specialist area) of the respective employees | Skill to perform meaningful assignment of tasks to different employees based on their competencies<br><br>Skill to bring in and use the experience from previous projects to make the | Willingness to deal with problems of individual experts<br><br>Openness to the wishes of other employees<br><br>Readiness to help out in emergency situations | (1) Can plan and coordinate a data project that is handled by a single data expert<br><br>(2) Can plan and coordinate a data project with few participants, especially data experts, whose interests converge |

| | | | | | |
|---|---|---|---|---|---|
| | | Time Management<br><br>Selection of suitable tools | project/collaboration more efficient<br><br>Skill to work with collaboration/versioning tools like git | | (3) Can plan and coordinate a data project with many interdisciplinary participants and diverging interests |

**Field of competence B: Providing data – from measurable objects to data**

This includes the provision of data. The first stage deals with the design and is divided into two parts: modelling of the data application and compliance with data protection and security. The second stage involves the procurement of data, more precisely, the identification of data sources and the integration of the data. The third stage is the standardisation of the data. This includes the verification and preparation of the data.

Table 3: Field of competence B: Providing data

| Competence | | Dimensions | | | Levels |
|---|---|---|---|---|---|
| **Labelling** | **Description** | **Examples of knowledge** | **Examples of skills** | **Examples of attitude** | **Examples of ascending levels** |
| B1.1: Modelling a data application | Maps the measurable objects into variables with definable properties and their relationships in a model structure | Knowledge about the relationships between real objects and digital representations of these objects<br><br>Knowledge that information is lost when measuring objects or their properties and that a selection and evaluation process is necessary to determine which information can be dispensed with<br><br>Knowledge about possible distortions that occur, for example, in | The skill to understand the requirements of the technical expert<br><br>The skill to map a process model in a data model<br><br>The skill to map the relevant process characteristics in data<br><br>The skill to determine the view on the data required for the research question (e.g. historical data views for forecasts)<br><br>The skill to determine the necessary | Willingness to make decisions regarding the information to be measured<br><br>Willingness to weigh up and accept information losses in the measurement process<br><br>Willingness to acquire technical terms from various disciplines and to adapt one's own communication to the knowledge of the technical expert, including the willingness to abandon precision | (1) Can determine variables corresponding to models of simple problems and simple functional relationships in the form of hypotheses<br><br>(2) Can derive suitable variables and goal-oriented hypotheses for more complex questions, which can be tested statistically<br><br>(3) Can produce various possible variables and alternative hypothesis systems on highly complex issues, anticipating possible models |

| | | | | | |
|---|---|---|---|---|---|
| | | surveys or crowd-generated data<br><br>Knowing which types of questions require which data structures and perspectives (e.g. customer or transaction view, depending on what the object is) | granularity of the data | in favour of comprehensibility<br><br>Willingness to accept restrictions, e.g. due to the delimitation of the research question, as well as willingness to communicate additional specifications, extensions, or restrictions and to enforce them against resistance if necessary | and testing procedures |
| B1.2: Compliance with data protection and security | Observes guidelines for secure and ethically sound data processing and implements it accordingly where no clear guidelines are defined | Knowing which criteria characterise data that is particularly worthy of protection<br><br>Knowledge of which data protection and data security requirements are necessary for the respective data situation<br><br>Knowledge, which statistical and technical possibilities exist to protect and secure data | The skill to apply data protection and data security requirements in a specific situation<br><br>The skill to implement data protection and data security by analogy if no clear regulation exists for a situation | Willingness to be guided by data protection and data security as well as ethical principles, even if there are no clear rules for the specific situation, or compliance with these rules is not checked<br><br>Understanding data protection, data security and informational self-determination as values<br><br>Willingness to force information loss for ethical or legal reasons<br><br>Willingness to follow the development of new technologies for data management and | (1) Detects gross violations of data protection<br><br>(2) Knows basic rules that must be observed in data protection and is basically aware of what data he may use and how, or when caution is required<br><br>(3) Has a sound knowledge of data protection and has sufficient competence to be given responsibility for data protection for entire projects |

| | | | | | |
|---|---|---|---|---|---|
| | | | | to train independently | |
| B.2.1: Identifying data sources | Identifies various common and novel data sources (internal, external) and evaluates their accessibility, relevance, and usability | Knowledge of possible data sources<br><br>Knowledge about search engines<br><br>Knowledge of criteria for assessing the quality of data sources (e.g. neutrality, quality standards)<br><br>Knowledge of the characteristics of data based on the survey method (observation, measurement, interview)<br><br>Knowledge of the new types of data sources created by digitisation ("Big Data")<br><br>Know which standards and best practices are common in the respective subject area when searching for data, e.g. whether working with primary or secondary data | The skill to search for and select data suitable for the research question<br><br>The skill to formulate queries for search engines and databases<br><br>The skill to assess the quality of data sources<br><br>The skill to design independent surveys to carry them out or to contract them out | Willingness to make decisions regarding the information to be procured<br><br>Willingness to weigh up and accept information losses in the data selection process<br><br>Sceptical attitude towards the quality of data sources<br><br>Openness to new types of data sources (e.g. sensor data, app data, etc.) | (1) Accesses data portals, makes enquiries via search engines, commissions simple surveys, but the quality of the source is not additionally checked<br><br>(2) Uses search engines, internal and external data sources for more demanding queries, commissions sophisticated surveys and checks the quality of the sources<br><br>(3) Accesses sources that are accessible through complex database queries or interfaces to machines, apps, etc., plans sophisticated surveys independently and follows a standardised quality assurance process |
| B.2.2: Integrating data | Automatically reads data in various formats, integrates it and documents | Knowledge of data models and their description (meta data)<br><br>Knowledge of how data can be made accessible, stored, and | The skill to find, acquire and integrate appropriate data<br><br>The skill to collect data in a new way using different methods | "Data fairness" as a basic attitude<br><br>Willingness to actively provide data to facilitate the acquisition of information and | (1) Can read common data sets<br><br>(2) Can troubleshoot for correct reading of data (encoding etc.) |

| | | | | | |
|---|---|---|---|---|---|
| | the integration | integrated into existing infrastructures<br><br>Knowledge of technologies for (automated) procurement of secondary data and storage<br><br>Knowledge of methods for collecting primary data<br><br>Knowledge of which standards and best practices are common in the respective field when data is collected, procured, and stored | The skill to select and use appropriate tools for this purpose | knowledge for others<br><br>Understanding that data collection always involves subjective decisions, and willingness to make these decisions transparent and justify them with arguments<br><br>Willingness to question the appropriateness of the data to be collected for the fulfilment of tasks, i.e. to ensure the effectiveness of the work steps | (3) Can read data automatedly, can save data automatedly in databases |
| B.3.1: Verifying data | Checks the data quality regarding various criteria (correctness, relevance, representativ eness, completeness )<br><br>Documents the audit systematicall y | Knowledge about possible quality problems and ways of dealing with them<br><br>Knowledge about possible causes of data quality problems, e.g. in the form of collection, transformation, or data protection<br><br>Knowledge of criteria for the evaluation of data quality | The skill to systematically identify quality problems, errors and gaps in the data and to remedy them appropriately | Objectivity as a basic attitude, i.e. not manipulating data to force a certain statement (e.g. when determining outliers)<br><br>Willingness to actively think through and communicate the possibilities and limits of interpreting the collected data and its possible transformations | (1) Identifies wrong type of variable (text vs. numeric)<br><br>(2) Recognises strongly correlated data, can do a first correction of content problems<br><br>(3) Eliminates problems in all common areas and can obtain missing, new data for a question and link it meaningfully to the previous data |
| B.3.2: Preparing data | Cleans data, corrects errors, imputes missing | Knowledge about possible technologies and tools for | The skill to transform data e.g. by standardisation, recoding, | Willingness to invest time and effort in elaborate and unexciting processing steps | (1) Performs simple creation of new variables from existing ones |

| | | | | | |
|---|---|---|---|---|---|
| | values, standardises, and transforms data, filters relevant data for a given question, links data | transforming data<br><br>Knowledge about data formats and other characteristics of data (e.g. scale levels, degree of structuring)<br><br>Knowledge about possibilities to transform data by standardising or linking it<br><br>Knowledge about how transformation processes change the information content of data | aggregation, combination, or other methods<br><br>The skill to link different data sources<br><br>The skill to identify redundant information in the data and to eliminate redundancies<br><br>The skill to provide data in various forms or aggregated data for specific purposes<br><br>The skill to anticipate future or alternative uses of the data<br><br>The skill to anticipate future or alternative interpretations of the data and its transformations | to increase the relative value of the data<br><br>Understanding of the value of data; understanding that storage space and analysis capacities are scarce resources and willingness to use them sparingly, e.g. in terms of working time and energy consumption<br><br>A basic attitude of expediency and user orientation, i.e. adherence to the effectiveness of the work steps (e.g. when imputing missing values) | (2) Merges of data records with prior modification of values<br><br>(3) Performs complex changes of data records (long and wide format), performs complex processing of character strings |

**Field of competence C: Evaluating data – from data to data products**
This includes the evaluation of the data. The first step is the analysis of the data. Here, the analysis methods from various fields (statistics, analytics, machine learning) are applied in a factual and purpose-oriented manner with the help of suitable tools. In the second stage the data is analysed. Static and dynamic visualisations are used here together with suitable tools. The third stage comprises the verbalisation of data. Data analyses are presented in various text forms and made available for communication.

Table 4: Field of competence C: Evaluating data

| Competence | | Dimensions | | | Levels |
|---|---|---|---|---|---|
| **Labelling** | **Description** | **Examples of knowledge** | **Examples of skills** | **Examples of attitude** | **Examples of ascending levels** |
| C.1: Analysing data | Uses analysis methods from various fields (statistics, analytics, machine learning), with the help of suitable tools in a factual and purpose-oriented manner | Knowledge of procedures for different tasks (description, exploration, prognosis) as well as their requirements, strengths, and weaknesses<br><br>Knowledge of procedures for directed and undirected questions<br><br>Knowledge about estimation methods and algorithms<br><br>Knowledge about possibilities of model diagnostics and model optimisation<br><br>Knowledge about measures to ensure the robustness and general validity of the models (e.g. cross-validation)<br><br>Knowledge about possible causes of artifacts<br><br>Knowledge of best practices and standards of model development and | The skill to map measurable relationships in models<br><br>The ability to identify and select appropriate analytical methods based on the issue at hand and available data<br><br>The skill to specify the model appropriately (e.g. by defining optimisation criteria)<br><br>The skill to examine the model for weaknesses and artifacts (e.g. overfitting, multicollinearity problems) and to counteract these<br><br>The skill to assess the uncertainty of the model results and to determine the required accuracy<br><br>The skill to anticipate future uses of the analysis results<br><br>The skill to separate relevant from irrelevant information in the | Willingness to implement and adapt models in an iterative and often time-consuming process<br><br>Sceptical basic attitude in data analysis<br><br>Willingness to weigh up and accept information losses in the analysis process<br><br>Willingness to comply with "good analytics standards", even if they are not explicitly defined<br><br>Willingness to work in a resource-saving manner, e.g. not to "over-engineer" the model<br><br>Willingness to enforce the required precision even when resources are scarce and against resistance to prevent fallacies<br><br>"Analytical fairness" as a basic attitude, i.e. willingness not to | (1) Can handle basic statistical methods such as mean value and standard deviation<br><br>(2) Can handle and use more complex models, can assess which methods provide meaningful results for which questions and data, and recognises the limitations of analytical results<br><br>(3) Masters and uses highly complex models and recognises when the analysis cannot provide meaningful results or when the information from the analysis is not relevant to the issue and the analysis process should be terminated |

| | | | | | |
|---|---|---|---|---|---|
| | | data analysis, e.g. validation<br>Knowing that information is lost when analysing data and that a selection and evaluation process is necessary to determine which information can be dispensed with | analysis process (e.g. to select variables) | carry out analyses if the risk of misuse is high<br>Objectivity as a basic attitude, especially in situations where the data situation and the question posed leave room for analysis | |
| C.2: Visualising data | Utilises static and dynamic visualisations with the help of suitable tools in an appropriate and purpose-oriented manner | Knowledge of the different diagrams, types of slides and their effect on the presentation<br>Knowledge of the different display options within the diagrams (colours etc.)<br>Knowledge that different conclusions can be drawn depending on the presentation form | The skill to select the correct diagrams and display options for the data set<br>The skill to accurately identify and minimise potential uncertainties<br>The skill to classify information in the data set in such a way that its importance is also reflected in its presentation<br>The skill to design a visualisation in such a way that exactly the relevant knowledge is conveyed | Objectivity in the representation, no conscious manipulation of the viewer<br>Willingness to invest time in a more elaborate presentation if it leads to correct conclusions<br>Willingness to correctly visualise data that does not correspond to one's own basic attitude | (1) Can create the basic visualisations<br>(2) Has mastered a variety of visualisations and can estimate when which is best suited<br>(3) Masters various visualisations in detail, knows their strengths and weaknesses and quickly recognises the advantages and disadvantages of new visualisations |
| C.3: Verbalising data | Verbalises the results of data analyses in various text forms in a factual and purpose- | Knowledge of the technical vocabulary in statistics and the fine differences<br>Knowledge of the methods and | The skill to formulate the statistical results in such a way that they can be understood and interpreted by | Willingness to present statistical results correctly and objectively, so that no false or distorted | (1) Can describe and explain simple statistical methods<br>(2) Can describe a variety of statistical methods and can explain |

| | | | | | |
|---|---|---|---|---|---|
| | oriented manner | their descriptions and explanations | non-statisticians without leading to false conclusions | conclusions are suggested<br><br>Willingness to objectively present data and results that do not correspond to the own basic attitude | them to laypersons<br><br>(3) Can describe various statistical methods briefly and concisely to non-specialist users, both orally and in writing, so that the analytical process can be understood. |

**Field of competence D: Interpreting data products – from data products to data**

This includes the interpretation of the data products. In the first stage, data products (statistics, model results) are first interpreted in a previously verbalised form and the explicitly or implicitly provided interpretation is critically reviewed. In the second stage, graphics are then analysed, and conclusions are drawn about essential elements and relationships. Here, too, the delivered interpretation is critically examined. In the third stage, statistical parameters and models are interpreted in such a way that conclusions can be drawn about underlying data points and relationships and forecasts can be made.

Table 5: Field of competence D: Interpreting data products

| Competence | | Dimensions | | | Levels |
|---|---|---|---|---|---|
| **Labelling** | **Descriptio n** | **Examples of knowledge** | **Examples of skills** | **Examples of attitude** | **Examples of ascending levels** |
| D.1: Interpreting data analyses | Interprets data products (statistics, model results) in verbalised form or critically checks the explicitly or implicitly delivered interpretatio n | Knowledge of statistical key figures such as mean values, percentages, percentage points, their significance, and limitations<br><br>Knowledge of the relationships between statistical key figures and the underlying data<br><br>Knowledge about the relationship between key figures (e.g. | Can draw conclusions about which characteristics of the data a key figure/ makes statements about<br><br>Understands which key figures (also verbalised) are used equivalently<br><br>Can question whether special data situations influence the result | Willingness to question explicitly communicated, given interpretations in data verbalisations<br><br>Willingness to search for and question implicitly communicated interpretations<br><br>Willingness to question one's own contextual knowledge regarding its | (1) Can understand simple statistical terminology and interpret its relation to data, knows basic forms of manipulation by statistics and reports and the criteria to be observed<br><br>(2) Has an advanced understanding of terminology and can differentiate clearly between |

| | | | | | |
|---|---|---|---|---|---|
| | | relative/absolute frequencies)<br><br>Knowing that the choice of a certain key figure can be the result of a conscious decision process<br><br>Knowledge of statistical terminology<br><br>Knowledge of statistical fallacies (e.g. correlation vs. causality)<br><br>Knowledge that statistical statements are generally not individual case statements<br><br>Knowing that interpretation requires the addition of contextual information | Can question whether the presentation form influences the message of a data product<br><br>Can question the extent to which the interpretation of the results depends on own contextual knowledge<br><br>Knows to question the extent to which the presentation of the results manipulates the interpretation through the choice of form (e.g. percentages) or verbalisation | influence on interpretation<br><br>Openness to new insights, even if they contradict previous convictions<br><br>Willingness to question the significance of the results for the facts in which they are presented | different terms, knows the building blocks of explicit communication, and can check whether explicit interpretations can be derived from the results<br><br>(3) Has a deep understanding of the terminology, can recognise implicit statements and interpretations, and can weigh them up against each other and in relation to the facts of the case |
| D.2: Interpreting Data Visualisation s | Interprets graphics and draws conclusions on essential elements and connections or critically examines the explicitly or implicitly delivered interpretatio n | Knowing which conclusions can and cannot be drawn from slides (correlation and causality)<br><br>Knowledge of the advantages and disadvantages of the individual slides | The skill to identify and read out different and relevant points in a diagram<br><br>Ability to interpret special markings<br><br>The skill to draw conclusions from model equations on data points<br><br>The skill to question and to critically assess the statements and the choice of visualisation | Willingness to question the visualisation and the conclusions<br><br>Willingness to search for implicitly communicated information and to look at it critically<br><br>Openness to new conclusions and new knowledge, even if this contradicts the current convictions | (1) Basic diagrams can be interpreted<br><br>(2) Has a basic understanding of the simple diagrams and can critically question them; in addition, more complex visualisations can be interpreted<br><br>(3) Both simple and complex diagrams can be interpreted and questioned in detail; both implicitly and explicitly |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | transmitted knowledge is recognised |
| D.3: Interpreting Data Verbalisations | Interprets statistical parameters and models to draw conclusions about underlying data points and relationships or to make forecasts | Knowledge of the basic statistical methods | The skill to understand and critically question the analysis process based on the description<br><br>The skill to read and understand statistical texts | Willingness to question the descriptions and conclusions<br><br>Willingness to search for implicitly communicated information in the subtext<br><br>Willingness to question the method of communication<br><br>Openness to new conclusions and new findings, even if they contradict the current convictions | (1) Can understand simpler explanations of statistical methods<br><br>(2) Can understand and comprehend more complex analytical processes based on written or oral explanations<br><br>(3) Can understand and question more complex analytical processes based on written or oral explanations |

**Field of competence E: Interpreting data – from data to measurable objects**

This includes the interpretation of the data. In the first stage, the standardisations must first be decoded by identifying the statistical methods used and thus the underlying transformation of the data. In the second step, the data retrieval is traced. Based on the analysis and the information provided, the exact original procurement and source can be identified, and its integrity assessed. Finally, in the third stage, the data concept is reconstructed by drawing conclusions about the data basis and potential misconceptions.

Table 6: Field of competence E: Interpreting data

| **Competence** | | **Dimensions** | | | **Levels** |
|---|---|---|---|---|---|
| **Labelling** | **Description** | **Examples of knowledge** | **Examples of skills** | **Examples of attitude** | **Examples of ascending levels** |
| E.1: Decrypting standardisation | Recognises, assesses, and interprets the statistical methods used; recognises | Knowledge of the different statistical methods, their similarities, differences, and potential interpretations | Ability to interpret statistical key figures and question their use<br><br>The skill to interpret statistical figures and their potential differences | Objectivity in the interpretation of the key figures<br><br>Openness to question obvious conclusions | (1) Has a basic understanding of statistical ratios, data transformations and their definitions<br><br>(2) Has a detailed understanding of |

| | | | | | |
|---|---|---|---|---|---|
| | the transformation of the data | | objectively and not to be led astray | | the different transformation options and their differences<br>(3) Has a detailed understanding of the transformations, their calculations, and their different effects on the conclusions |
| E.2: Trace data acquisition | Can trace back, based on the analysis and the information provided, how the data was obtained, from which source it originated and what confidence can be placed in the data | Knowledge of potential data sources<br>Content knowledge about the quality of certain data sources<br>Knowledge about possible sources of error in data acquisition and data collection | The skill to carry out a (systematic) literature search<br>Ability to question data sources and data collection | Objectivity in the interpretation and assessment of data sources | (1) Searches for a simple data record (if necessary, with link etc.)<br>(2) Obtains information from different sources<br>(3) Can combine such information appropriately |
| E.3: Reconstructing the data concept | Can draw conclusions about the data basis and potential fallacies | Knowledge of the theoretical basis of the data set and possible misconceptions that could lead to incorrect interpretations | The skill to identify possible weaknesses in the evaluation and implicit transmission of information | Attitude not to immediately accept the author's implicit conclusion, but to question it<br>Understanding that data does not allow all conclusions to be drawn, but that even data cannot make all statements | (1) Has a basic data knowledge that helps interpret the data<br>(2) Can critically examine data in general regarding the conclusions<br>(3) Has a detailed knowledge of the data set, general data basis, and can challenge conclusions based on this knowledge |

**Field of competence F: Deriving action – from measurable objects to systems**
This includes the derivation of action. In the first stage, concrete options for action are first identified, the assessment and evaluation of which can be evaluated with data. In addition, an idea of the possible value contribution of the data in deriving possible courses of action is developed. The second stage describes the integration of results into the decision-making process and the actions based on the results. Finally, in the third stage, data-based action is tested and evaluated for its effectiveness.

Table 7: Field of competence F: Deriving action

| Competence | | Dimensions | | | Levels |
|---|---|---|---|---|---|
| **Labelling** | **Descriptio n** | **Examples of knowledge** | **Examples of skills** | **Examples of attitude** | **Examples of ascending levels** |
| F.1: Identifying possibilities for action | Identifies concrete possibilities for action, the assessment and evaluation of which can be evaluated with data; has an idea of the possible value contribution of the data when deriving possibilities for action | In-depth theoretical and practical knowledge of the field of application or discipline and, if applicable, related disciplines<br><br>Knowledge of the relevant literature or professional requirements (norms, rules, quality standards, processes, restrictions)<br><br>Knowledge that and how data and its analysis can be used for decision making in their own discipline | The ability to assess the possibilities that have arisen due to data-driven developments based on their meaningfulness<br><br>The skill to separate relevant from irrelevant information about the actions | Openness to accept unexpected possibilities for action<br><br>Willingness to learn from data<br><br>Willingness to question existing rules and processes | (1) Can identify basic fields where data-driven action would be beneficial<br><br>(2) Can assess exactly in which fields data-driven action would be possible and where it would be too complex<br><br>(3) Can make a detailed assessment of the advantages and disadvantages of data-driven actions, evaluate the possibilities |
| F.2: Data-driven action | Describes how to integrate results into the decision-making process and how to base actions on these results | In-depth theoretical and practical knowledge of the various options for action and their implications<br><br>Basic statistical knowledge to assess which | The skill to select the appropriate analyses and data for decisions on actions<br><br>Ability to formulate and quantify the various options for action in such a | Willingness to base actions also on data-driven decisions<br><br>Openness to accept the results<br><br>Willingness to question current actions | (1) Basic decisions on action can be based on the results of an analysis or statistics<br><br>(2) More complex decisions for action can also be |

| | | | | | |
|---|---|---|---|---|---|
| | | analyses are important for the decision | way that a data-driven decision is possible<br><br>The skill to transform the results of the analysis in such a way that an action can be derived from them<br><br>The skill to identify the optimal solution and recognise potential unwanted side effects | Objectivity in the analysis of the possible courses of action | quantified and justified with data<br><br>(3) The action is completely data-driven, not only decisions are made regarding the action, the system is continuous, and the action strategy is adjusted according to the data |
| F.3: Evaluating the effect | Describes the evaluation of data-based action based on its effectivenes s | Knowledge about the possibilities for action and their quantification to the knowledge of the effectiveness | The ability to put into words and quantify the effect, impact, or difference<br><br>The skill to focus on the relevant and essential aspects when assessing the impact and not to be confused by unimportant aspects | Objectivity in assessing the impact<br><br>Openness to the results<br><br>Willingness to implement the results of the effectiveness assessment | (1) The basic effect of major actions and their differences can be estimated<br><br>(2) The effectiveness of the data-based actions can be tested in detail<br><br>(3) The effectiveness can be tested, and the action can be fine-tuned based on the results |

# Chapter 4: Reflection and Outlook

### 4.1 Changes in the world of work through Data

The need for data literates and data experts will continue to increase with the advancement of digitisation and datafication. Data is playing an increasingly important role in all areas. For example, the importance of image processing is increasing rapidly in the medical field. This makes it possible to automatically generate diagnoses and recommendations for action from pictures that a person could not derive with the same accuracy or speed.

Figure 4 gives an overview of some new professions in which the handling of data is central. The figure also contains long-existing roles (decision-makers, technical experts, citizens) that need data competencies at certain points. Vertical arrows indicate which areas of competence are covered by the respective professions and roles. The areas of competence partially overlap.

In the present study, data literacy is seen as a comprehensive competence, but learning goals must be derived in a subject-specific manner. It is therefore important in further work to develop such learning objectives for the individual disciplines in interdisciplinary cooperation. In addition, data literacy competence levels need to be defined, which can also serve to delimit expert levels (job description data scientist, data engineer and other related professions, see Figure 6) and to delimit the levels to be aimed for in the individual disciplines. From the author's point of view, a basic level should also exist in the individual specialist disciplines regarding each individual competence in the defined competence fields.

In addition to the established profession of statisticians, numerous new professional fields are emerging, such as that of data journalists. The data designer designs concept studies, collects data and visualises evaluations. He specialises either in the design of data-related processes and architectures (technical data designer) or in the design of visualisations (data artist). The terms data scientist, data engineer and data/business analyst are often used in an unsharp manner and are not clearly delimited. The term data scientist is often found in connection with activities related to the development and application of automated learning processes, especially with very large and unstructured data sets ("big data"), whereas a data engineer deals more with technical aspects, for example with the management of databases and IT infrastructures.

Data ethics will play an increasingly important role in the professional world. The data ethicist must "think ahead" which sort of data handling could be problematic, even if current practices do not violate applicable law. With a view to visionary questions, he stands above both the producing and the receiving side in the process of adding value from data. The data ethicist deals with social and normative change through digitisation and the consequences both for the economy and for us as individual citizens. The data ethicist is concerned with where a social system is developing to, what is good and bad for mankind and what the goals are. Currently, data ethicists are more of a role than a learned profession – a role that is becoming increasingly important in business and politics, as the involvement of data ethics in decision-making bodies during the Corona crisis has shown.

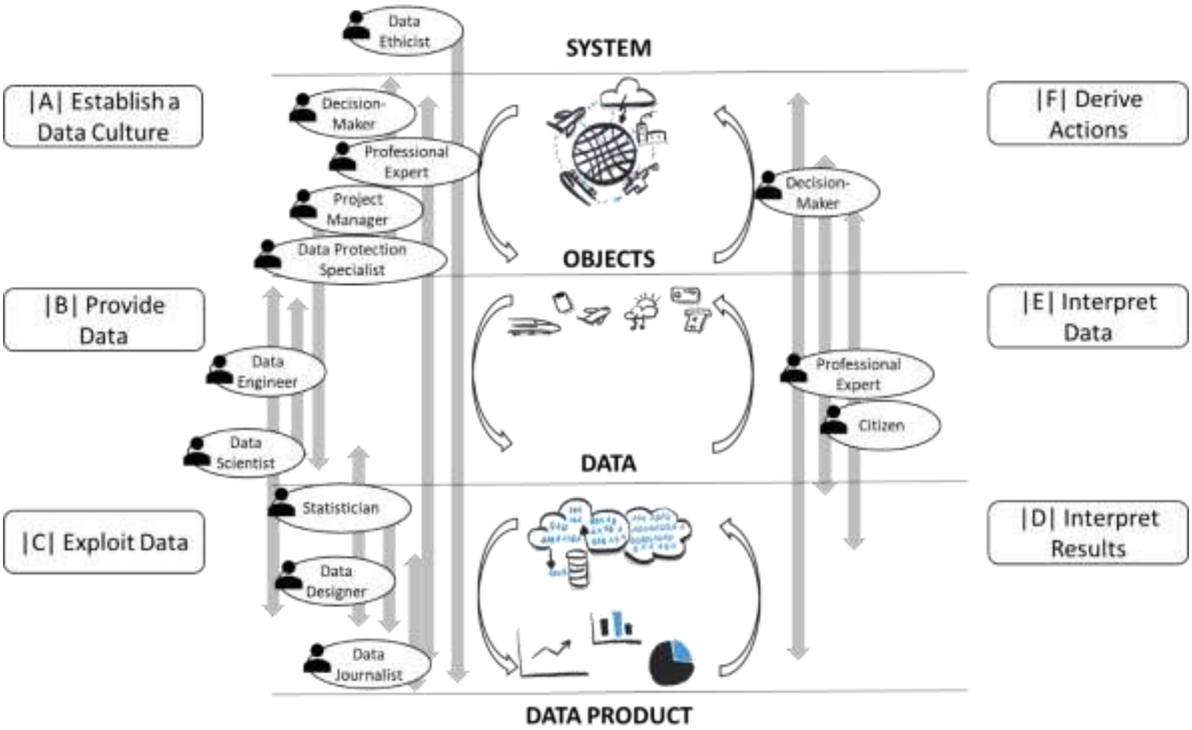



*Figure 4: The Data Literacy Framework and related disciplines and roles*

## 4.2 Change in the meaning of Data Literacy

In the context of the present study, experts from various universities were asked about their understanding of data literacy. The key messages of the interviews conducted with them are reflected below in the light of the Corona crisis. Nobody would have dreamed that the Corona crisis would make the value of data literacy so clear and inspire so many people to try their hand at analysing data. Even if the occasion is an unpleasant one, this is a huge opportunity.

At the beginning of the crisis, data experts faced a major challenge. They were expected to analyse numbers and make forecasts, but the Corona case numbers were not suitable because they contained too little information, such as the number of unreported cases. The crux of the matter was and still is that the number of cases depends significantly on how testing is carried out. As before, this is not done based on a representative selection of test subjects, but according to different regional and international strategies, which have also changed over time. This had to do with the availability of tests but was also a question of cost. The registered case numbers therefore do not allow any reliable statements about the overall situation that are comparable in terms of time or space.

A possible way of commenting as an expert would be the honest and transparent communication of the lack of secure knowledge. But there is a danger that numerous recipients will receive the message that the model calculations are wrong (because the probability that it will work exactly as in the model is practically zero) and deduce from this that the recommended measures, such as curfews, must also be wrong (although they were probably necessary and correct).

This is a big ethical dilemma. Is it better to stick to the truth and hope that experts will still be believed when interpreting uncertainty? Or should one suggest more certainty in order to achieve the right thing, even if one cannot “prove” in the strict sense that it is right?

Data literacy is a key competence, particularly in a crisis phase like the current one, in which scenario calculations are created with incomplete data. However, the understanding of what data literacy means and how this competence can be imparted is not yet uniform at German universities, although data competencies are currently gaining importance there. One university is currently planning a Master’s degree course in "Data Science", in which the entire curriculum will be covered by "thinking in data" as a cross-sectional topic, said an interviewed statistics professor, who is currently an economist dean of the business department at a university of applied sciences. According to a unanimous statement, the topic will increase significantly in importance.

A professor of business research (economist) commented: "The term data literacy has not yet appeared, at most in guest lectures." Data skills, or rather the lack of them, were mainly shown when evaluating model results. In the past three years, there has also been a great deal of interest in algorithmic data analysis, i.e. the processing of large amounts of data.

A professor of spatial development at a technical university addressed the problem that official data was not sufficient for spatio-temporal analyses to derive action from it: "Spatial data has so far been provided primarily by governing bodies, and its spatial resolution is no longer sufficient for our spatial understanding of active players. We therefore generate our own geo-referenced company and household data through a primary survey. Literacy here means that it must be understood exactly how the data is created, what the underlying facts are, what implicit temporalities resonate in order to be able to interpret the resulting information in a targeted and action-oriented manner.”

There were parallels in the Corona crisis when the debate about regionally different easing strategies was conducted in May 2020. Parameters before the reproduction number R and the 7-day incidence should serve as a benchmark. However, such parameters, which can only be calculated after a considerable delay and are based on data that allows anything but precise statements, cannot support decisions in advance. They do not provide control knowledge, but rather resemble driving a car with a look in the rear-view mirror. If the road is straight, this may still work, but not if the road – figuratively speaking – is changed by the measures taken.

Action knowledge and control knowledge require the interpretation of data and its integration into a context. In its key point paper on data strategy, (German federal government, n.d.) formulated the idea that knowledge is only created from data as follows: "In the digital age, data is a key resource for social prosperity and participation, for a prospering economy and the protection of the environment and climate, for the scientific community progress and for state action. The ability to use, link and evaluate data responsibly and independently is equally the basis for technological innovation, for the generation of knowledge and for social cohesion." The closeness of this formulation to the definition of data literacy used here is unmistakable.

### 4.3 Attitudes, Values, Ethics as a Competence Dimension.

The surveyed university professors perceived the meaning of the topics, attitudes, values, ethics in connection with data competence rather differently. Those with a background in statistics / econometrics or social science saw the strong role of attitudes. It is a modular component in various courses.

The professor of spatial development assessed (value) attitudes and motivation more from a regulatory or competitive point of view – it seems as if one should adhere to ethical standards as a precautionary measure because one has to: "EU regulations (GDPR) determine the framework for action. When analysing spatial preferences and motivational situations that we generate ourselves through primary surveys, moments of bias, personal protection, and confidentiality can lead to questions about relevant issues not being asked in the first place. Research methodical self-limitation ultimately comes before regulations that of course increasingly restrict data generation in the EU."

His comments show clear parallels to the discussion about whether cell phone data should be used to monitor people with Corona, but also to the admissibility of a Corona app. While in China rigorous surveillance of the population with the help of cameras, fever-measuring sensors and a mandatory app probably contributed significantly to the rapid containment of the first wave of pandemics, such an approach in Europe seemed to be an unenforceable encroachment on the constitutionally guaranteed freedom of the individual. The fact that the weighing up of individual liberties and the need for protection by society in no way leads to clear results that cannot be changed over time is borne out by the very sharp debates about the obligation to register when visiting restaurants or a possible immunity certificate.

In an interview for the present study, a medical professor was convinced that scientific data collection and evaluation required an underlying set of values. Above all, medical researchers must be careful not to carry out unnecessary studies on humans, to work scientifically with integrity and correctly, and to interpret the results regardless of personal preferences or wishes.

The importance of this assessment can be illustrated by two controversial studies by the virologists (Streeck, et al., 2020) and (Jones, et al., 2020), which have received numerous comments in specialist circles as well as in the public and social media after publishing an interim analysis respectively a preprint. Problematic was the sometimes quite explicit expectation of the scientists to generate results that, in retrospect, would justify political decisions already made. To counter this requires a high degree of data ethical competence.

### 4.4 Data Literacy through Diversity and Interdisciplinarity

In its key point paper on data strategy, (German federal government, n.d.) defined increasing data literacy as one of four key areas of action and of establishing a data culture. The Corona crisis offers a great opportunity for this, because, born out of the awareness of its necessity, people and institutions that previously had hardly any points of contact are increasingly working together. This interdisciplinary approach is the greatest lever for building data literacy because the process in which knowledge is created from data first requires "translating" a technical or social issue into a data model. Firstly, this includes determining what such a question encompasses.

In his final report on the "Heinsberg study" on page 3, (Streeck, et al., 2020) wrote: "However, epidemiological modelling is urgently needed to design the most appropriate prevention and control strategies to counter the pandemic and to minimize collateral damage to societies."

Obviously, the crisis is not a purely medical or epidemiological problem, but also a social one. It is an economic problem, a psychosocial one, it is an educational problem, and maybe even an ecological opportunity. But "you cannot manage what you cannot measure", it says. To measure all dimensions of the crisis, these must be recognised as being relevant at all. Interdisciplinary programs at universities for the establishment of data literacy education can make a valuable contribution to this. In Germany, for example, the nationwide data literacy education network, in which the Stifterverband has played a key role, has been supporting participating universities in the development and implementation of good practices and data literacy curricula since autumn 2019 (Stifterverband, n.d.).

Corona offers the opportunity to better measure and manage future crises if more diversity is created in the bodies that deal with such big problems. In the future, there should be more people involved in the expert councils who reflect the diversity of perspectives. It is certainly not only about the perspective of women, but it is striking how few women are represented in key positions and are involved in strategies to deal with the crisis. This, in turn, is reflected in the data that is available as the basis for decisions.

The German Federal Statistical Office (Destatis) publishes an extensive dossier with statistics on the COVID-19 pandemic every two weeks (Statistisches Bundesamt (Destatis), 2020). Eleven pages of the June 8, 2020 issue report statistics on the number of cases, deaths, and health care. 28 pages deal with the economy and the labour market, another 15 with branches and companies. An entire two pages deal with education, and this is limited to the presentation of tables on the number of pupils, of teachers and day-care children in the individual federal states. A single page deals with the topic of the environment and a quarter of it is taken up by a picture of the air quality app of the Federal Environment Agency.

Many questions, such as the impact of the closing of schools on families, the increase in mental disorders due to isolation or the consequences of short-time work, especially for employees of lower income groups, remain unanswered, because the data that would be needed for the answer is hidden in other databases or will not be collected at all. At least the German Federal Employment Agency provides clear indications in the monthly updated time series of the jobs reported that job opportunities for the low-skilled, as well as in typical "women's jobs", decreased significantly in the spring of 2020 (Bundesagentur für Arbeit, 2020). Statistics professor Ulrich Rendtel from the Free University of Berlin said in an interview: "People in the lower third of income experience a greater loss of income through short-term work – from a socio-political point of view, this is of course highly relevant " (Rendtel & Holly, 2020). Rendtel is one of those who have coordinated and advanced the cooperation between RKI and the Socio-Economic Panel (SOEP) to enable a representative panel study on as many aspects of the Corona crisis as possible.

The second major opportunity for increasing data literacy is given by the enhancement of cooperation between state institutions and private partners. Despite the test capacities that have meanwhile been extended, a representative panel sample for the entire population will presumably only be available in Germany from September 2020. Even for an institution like the RKI, this task alone seems too great and can only be solved in cooperation with the SOEP at the German Institute for Economic Research Berlin (DIW) and the University of Bielefeld.

It appears that joint efforts by the public sector, universities and private research institutions are necessary for the combining of high-quality data and research with agility and for the ability to obtain data quickly. Exactly this mixture will be needed in the future since the big challenges of our time – not only pandemics, but also migration movements, financial crises, climate change, economic interdependencies – can only be mastered by large, interdisciplinary teams.

## 4.5 Outlook: Data Literacy without Borders

All this results in the clear realisation that such crises can neither be solved by individual disciplines nor by national solo efforts. That is all the truer since data is now an indispensable basis for possible solution strategies. A coordinated, international approach across institutions and disciplines is required, and this applies to the question of which data can be used to measure and manage a crisis. This is the only way to create reliable, comparable information.

If you look at the European comparison table for the Corona case numbers in (Statistisches Bundesamt (Destatis), 2020), Belgium appears to be by far the most affected country. The statistics show over 84 deaths per 100,000 inhabitants, significantly more than in Italy and Spain (56 and 58) and Germany (10). This is mainly because the Belgian authorities also count suspected cases as Corona deaths. This shows just how little case numbers can be compared internationally.

That is why the Federation of European National Statistical Societies (FENStatS) has set itself the goal of harmonising the corresponding statistics with a specially created COVID-19 working group, which the author of this study is chairing (Federation of European National Statistical Societies, 2020). Nearly 30 experts from 17 countries and different areas of work have been working on proposals since June 2020 on how data on crisis management can be better collected and analysed in the future. Only with information-rich, reliable, timely and relevant data is it possible to develop reliable indicators for the vulnerability risks and adaptation costs of our society.

Cross-border cooperation is needed – crossing borders between nations, borders between public authorities and private institutions, borders between different groups of people affected, and borders between specialist disciplines. Only in this way will it be possible to understand major risks for our society faster and in all dimensions in the future. Universities play a key role in this. For them to be able to fill this position competently, the Data Literacy Education Network in Germany promotes professional exchange, peer-to-peer formats, and friendly advice. Thus, the universities involved can mutually benefit from their experiences and thus support each other in the implementation of their data literacy programmes.

This can lead to data being increasingly shared as "open data" so that as many people as possible can work and research with it. Or regarding the use of new data sources, such as in the experimental data from Destatis: for example, the daily evaluation of truck toll data shows that a highly current and at the same time reliable economic indicator can be calculated from (Statistisches Bundesamt (Destatis), 2020).

After all, the crisis painfully teaches that not every decision can be "predicted" by data. But data literacy is not the ability to create an illusion of certainty by using as much data and the most complex analysis methods as possible, but rather the ability to deal with uncertainty and not to delegate responsibility for decisions to data and algorithms. That would be statistical or data literacy, as the author H.G. Wells might have had in mind over 100 years ago as one of the three core competencies of the responsible citizen in a modern technological world: the ability to deal sensibly with risks and uncertainty.

# List of References

Awad, E. M. & Ghaziri, H. M., 2004. Knowledge Management. Upper Saddle City, NJ, Pearson Educational International.

Bellinger, G., Castro, D. & Mills, A., 2004. Data, Information,. [Online] Available at: https://www.systems-thinking.org/dikw/dikw.htm [Accessed 23 June 2020].

Bundesagentur für Arbeit, 2020. Gemeldete Arbeitsstellen - Deutschland, West/Ost, Länder, Kreise, Regionaldirektionen und Agenturen für Arbeit (Monatszahlen). [Online] Available at: https://statistik.arbeitsagentur.de/nn_1751878/SiteGlobals/Forms/Rubrikensuche/Rubrikensuche_Form.html?view=processForm&resourceId=210368&input_=&pageLocale=de&topicId=1601172&year_month=aktuell&year_month.GROUP=1&search=Suchen [Accessed 23 June 2020].

Carlson, J., Johnson, L., Westra, B. & Nichols, M., 2013. Developing an Approach for Data Management Education: A Report from the Data Information Literacy Project. s.l.:INTERNATIONAL JOURNAL OF DIGITAL CURATION.

Data Studio, n.d. Product Overview. [Online] Available at: https://datastudio.google.com/overview [Accessed 22 June 2020].

Deutscher Qualifikationsrahmen, n.d. DQR-Niveaus. [Online] Available at: https://www.dqr.de/content/2315.php [Accessed 22 June 2020].

European Commission, 2008. Der Europäische Qualifikationsrahmen für lebenslanges Lernen. [Online] Available at: https://ec.europa.eu/ploteus/sites/eac-eqf/files/leaflet_de.pdf [Accessed 22 June 2020].

European Commission, 2016. European e-Competence Framework 3.0. [Online] Available at: http://www.ecompetences.eu/wp-content/uploads/2014/02/European-e-Competence-Framework-3.0_DE.pdf [Accessed 22 June 2020].

European Commission, 2018. The Digital Competence Framework 2.0. [Online] Available at: https://ec.europa.eu/jrc/en/digcomp/digital-competence-framework [Accessed 22 June 2020].

European Commission, n.d. European Skills/competences. [Online] Available at: https://ec.europa.eu/esco/portal/skill?uri=http%3A%2F%2Fdata.europa.eu%2Fesco%2Fskill%2F25a26ff6-af18-40b0-b2e2-cb58471015eb&conceptLanguage=en&full=true [Accessed 22 June 2020].

Federation of European National Statistical Societies, 2020. FENStatS News. [Online] Available at: https://www.fenstats.eu/news/COVID-19_WG [Accessed 23 June 2020].

Foucault, M., 1980. Power/knowledge: Selected Interviews and Other Writings, 1972-1977. s.l.:Pantheon Books.

German federal government, n.d. Eckpunkte einer Datenstrategie der Bundesregierung. [Online] Available at: https://www.bundesregierung.de/resource/blob/997532/1693626/e617eb58f3464ed13b8ded65c7d3d5a1/2019-11-18-pdf-datenstrategie-data.pdf [Accessed 23 June 2020].

Goldsmith, S. & Crawford, S., 2014. The Responsive City: Engaging Communities Through Data-Smart Governance. s.l.: Jossey-Bass by John Wiley & Sons, Inc.

Heidrich, J., Bauer, P. & Krupka, D., 2018. Future Skills: Ansätze zur Vermittlung von Data Literacy in der Hochschulbildung, Kaiserslautern: Hochschulforum Digitalisierung.

Jones, T. C. et al., 2020. An analysis of SARS-CoV-2 viral load by patient age. [Online] Available at: https://zoonosen.charite.de/fileadmin/user_upload/microsites/m_cc05/virologie-ccm/dateien_upload/Weitere_Dateien/analysis-of-SARS-CoV-2-viral-load-by-patient-age.pdf [Accessed 23 June 2020].

Kichin, R., 2014. The Data Revolution. s.l.: Sage Publications Ltd.

Kuhlen, R., 2013. A 1 Information – Informationswissenschaft. In: R. Kuhlen, W. Semar & D. Strauch, eds. Grundlagen der praktischen Information und Dokumentation. s.l.: De Gruyter Saur, pp. 1-24.

Linoff, G. S. & Berry, M. J. A., 2004. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. s.l.: John Wiley & Sons, Inc..

Lohr, S., 2009. For Today's Graduate, Just One Word: Statistics. [Online] Available at: https://www.nytimes.com/2009/08/06/technology/06stats.html [Accessed 22 June 2020].

Mandinach, E. B. & Gummer, E. S., 2013. A Systemic View of Implementing Data Literacy in Educator Preparation. Educational Researcher, 42(1), p. 30–37.

Matthews, P., 2016. Data literacy conceptions, community capabilities. The Journal of Community Informatics, 12(3).

Münster, E., 2019. The Cycle of Encoding and Decoding. [Online] Available at: https://www.designation.eu/en/2019/01/31/why-does-no-one-understand-my-data-visualization/ [Accessed 22 June 2020].

Naisbitt, J., 1982. Megatrends: Ten New Directions Transforming Our Lives. New York: Grand Central Publishing.

Nicholson, J., Gal, I. & Ridgway, J., 2018. Understanding Civic Statistics: A Conceptual Framework, s.l.: A product of the ProCivicStat Project.

OECD, 2005. Definition und Auswahl von Schlüsselkompetenzen. [Online]
Available at: http://www.oecd.org/pisa/35693281.pdf

Rendtel, U. & Holly, L., 2020. Die Coronavirus-Pandemie im Fokus der Statistik. [Online]
Available at: https://www.fu-berlin.de/campusleben/forschen/2020/200711-statistik-interview/index.html
[Accessed 23 June 2020].

Ridsdale, C. et al., 2015. Strategies and Best Practices – Knowledge Synthesis Report. [Online]
Available at: http://www.idees-ideas.ca/sites/default/files/sites/default/uploads/general/2015/sshrc-ksg-smit.pdf
[Accessed 22 June 2020].

Schüller, K. & Busch, P., 2019. Data Literacy: Ein Systematic Review zu Begriffsdefinition, Kompetenzrahmen und Testinstrumenten, Munich: Hochschulforum Digitalisierung.

Schüller, K., Busch, P. & Hindinger, C., 2019. Future Skills: Ein Framework für Data Literacy, Munich: Hochschulforum Digitalisierung.

Schüller, K. & Wrobel, C., 2018. Unlocking the Doors of Frankfurt Airport's Digital Marketplace: How Fraport's Smart Data Lab Manages to Create Value from Data and to Change the Airport's Way of Thinking. In: C. Linnhoff-Popien, R. Schneider & M. Zaddach, eds. Digital Marketplaces Unleashed. Berlin, Heidelberg: Springer, pp. 637-649.

Seidl, T. et al., 2018. (WERT-)HALTUNG ALS WICHTIGER BESTANDTEIL DER ENTWICKLUNG VON 21ST CENTURY SKILLS AN HOCHSCHULEN, s.l.: Hochschulforum Digitalisierung.

Statistisches Bundesamt (Destatis), 2020. Dossier: Statistiken zur COVID-19-Pandemie, s.l.: s.n.

Statistisches Bundesamt (Destatis), 2020. Lkw-Maut-Fahrleistungsindex. [Online]
Available at: https://www.destatis.de/DE/Service/EXDAT/Datensaetze/lkw-maut-artikel.html
[Accessed 23 June 2020].

Sternkopf, H. & Mueller, R. M., 2018. Doing Good with Data: Development of a Maturity Model for Data Literacy in Non-governmental Organizations. s.l., Proceedings of the 51st Hawaii International Conference on System Sciences.

Stifterverband, n.d. Data Literacy Education. [Online]
Available at: https://www.stifterverband.org/data-literacy-education
[Accessed 23 June 2020].

Streeck, H., Hartmann, G., Exner, M. & Schmid, M., 2020. Vorläufiges Ergebnis und Schlussfolgerungen der COVID-19 Case-ClusterStudy (Gemeinde Gangelt). [Online]
Available at:
https://www.land.nrw/sites/default/files/asset/document/zwischenergebnis_covid19_case_study_gangelt_0.pdf
[Accessed 23 June 2020].

Streeck, H. et al., 2020. Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. [Online]
Available at: https://www.medrxiv.org/content/10.1101/2020.05.04.20090076v2.full.pdf
[Accessed 23 June 2020].

Tankard, J. W., 1979. The H.G. Wells Quote on Statistics: A Question of Accuracy. Historia Mathematica, Volume 6, pp. 30-33.

The Association of Independent Schools of New South Wales, n.d. Data Informed Decision Making Cycle. [Online]
Available at: https://www.aisnsw.edu.au/Resources/WAL%204%20[Open%20Access]/Data%20Informed%20Decision-Making%20Cycle.pdf#search=Data%20informed%20decision%20making%20cycle
[Accessed 22 June 2020].

Watson, J. & Callingham, R., 2002. Statistical literacy: A complex hierarchical construct. Statistics Education Research Journal, 2(2).

Watson, J. & Callingham, R., 2004. Statistical Literacy: From Idiosyncratic to Critical Thinking, s.l.: Curricular Development in Statistics Education.

Weinert, F. E., 2014. Leistungsmessungen in Schulen. 3rd ed. s.l.:Beltz Verlagsgruppe.

Wild, C., 2006. THE CONCEPT OF DISTRIBUTION. Statistics Education Research Journal, 5(2), pp. 10-26.

Williford, C. & Henry, C., 2012. One Culture. Computationally Intensive Research in the Humanities and Social Sciences, Washington, D.C.: CLIR – Council on Library and Information Resources.

# Publishing Information



**Cite as**

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

**Published by**

Geschäftsstelle Hochschulforum Digitalisierung beim Stifterverband für die Deutsche Wissenschaft e.V.
Hauptstadtbüro • Pariser Platz 6 • 10117 Berlin • T 030 322982-520
info@hochschulforumdigitalisierung.de

**Editor**

Johanna Ebeling

**Publishing House**

Edition Stifterverband – Verwaltungsgesellschaft für Wissenschaftspflege mbH
Barkhovenallee 1 • 45239 Essen • T 0201 8401-0 • mail@stifterverband.de

**Graphicy & Layout**

TAU GmbH • Köpenicker Straße 154 A • 10997 Berlin
Philipp Kürten

**Images**

p. 9: Photo by Joshua Hibbert on Unsplash / p. 14: Photo by pine watt on Unsplash / p. 23: Photo by Paweł Czerwiński on Unsplash / p. 42: Photo by Johannes Plenio on Unsplash

The Hochschulforum Digitalisierung is a joint project of the Stifterverband, the Centre for Higher Education Development and the German Rectors' Conference.
It is funded by the Federal Ministry of Education and Research.

www.hochschulforumdigitalisierung.de

STIFTERVERBAND
CHE Centrum für Hochschulentwicklung
HRK Hochschulrektorenkonferenz
GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

hochschulforumdigitalisierung.de/publikationen