Next Step: Data Literacy Measurement

Marketa Smolnikova
Department of Information Technology, University of Economics, W. Churchill Sq. 4, Prague, Czech Republic

Keywords: Data, Data Literacy, Data Literacy Indicator, Data Literacy Measurement, Knowledge Management.

Abstract: As data became a new business commodity, affecting our everyday lives from shopping to voting, it smoothed the way for data literacy as a tool for full participation in a modern society. This paper argues for data literacy development and accelerated research of its measurement which has been lagging behind countless studies on teaching data skills. Data literacy is in this paper approached as an ability to understand and use data and differentiates itself from information or statistical literacy. As a prerequisite of information literacy, data literacy is inevitable part of knowledge development. While the term of data literacy has been well established and used for developing best practices and methodologies to teach data skills, measurement of data literacy seems to be still in its infancy. As a result, this paper includes research plan for developing a data literacy indicator based on quantitative methods.

1 INTRODUCTION

Data is the new oil. It has become a common phrase and easily accepted fact in the recent years. Why? This “call for action” of Big Data and education specialists (ODI 2015) briefly summarizes why we label the present day as “a golden era of data” and introduces the business case for data literacy:

1. Our world economy and our jobs are increasingly defined by data and by the knowledge and skills required to use them effectively.
2. We are all perpetually producing streams of data, which we need to shape and manage to ensure our privacy and personal security.
3. Effective use of data empowers us to make objective, evidence-based inferences and fundamental decisions affecting our lives, both as individuals and among societies.”

No wonder, Gartner also recognizes data as the new core capability of business along with people, processes, and technology. (Gartner 2018a) Grillenberger and Romeike (2018) argue that “knowing about the possibilities offered by data and data analysis plays an increasing role for developing an understanding of the world.” We manipulate data in everyday processes regardless of sectors or domains. That supports Ridsdale’s et al. (2015) view that data literacy “is an essential ability required in the knowledge-based economy”.

2 DATA LITERACY

However, to go “data-mindful” at full scale and to enhance organization’s lead in the fierce competition or to expand individual’s understanding and its future options in employability, businesses, institutions, schools, and its members require a certain level of data literacy.

2.1 What Is Data Literacy

Ridsdale et al. (2015) define this type of literacy as “the ability to collect, manage, evaluate and apply data, in a critical manner”. Gartner (2018b) further elaborates on the definition of data literacy by articulation of four key barriers of data literate society – the individual as well as organizational incapacity to derive insights from their data, to understand the analytical methods, to use analytical services to get the insights or the incompetence to comprehend and to integrate company’s data sources.

Yet we shouldn’t get intimidated with the above definition and we should rather expound data literacy as “the ability of non-specialists to make use of data”.

https://orcid.org/0000-0002-2631-920X
As Wolff et al. (2016) emphasizes, a data literate person follows the same phases of data processing as data scientists and knows how to make use of them for its objectives. Nevertheless, the devil is (obviously) in the detail as proposed in the Figure 1 in which the leftmost vertical arrow pointing downwards suggests the level of expertise. While a data literate person gets along with basic understanding of the process and methods, data scientists are professionals with profound knowledge and skills in data management and advanced statistical methods. In other words, data literacy translates into being able to "read and speak data", to understand data and being able to make use of them, in order to take a full part in society affected by the availability and accessibility of vast volume of data.

2.2 It Does Not Equate with Information or Statistical Literacy

In "the increasingly pervasive nature of data" (Gartner, 2018b) first we need to learn to handle the volume and characteristics of data (discrete, objective facts) before we can draw information from it (make data useful, enrich them with meaning). That is where we demarcate a line between data literacy and information literacy which ACRL (1989) specifies as: “To be information literate, a person must be able to recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information”. The difference arises from the relation between data and information in which information plays a role of a result of data processing during which meaning is assigned to data and which is explained in detail in the data-information-knowledge-wisdom (DIKW) hierarchy (Rowley 2007). Thus in the world brimming over with data, data literacy is a prerequisite of information literacy.

Even though data literacy inevitably draws on statistical methods, it differentiates itself from statistical literacy. According to Gould (2017), goal of statistical literacy is “developing critical consumers of statistics”. Gal (2002) calls it people’s “ability to interpret and critically evaluate” statistical products, as well as their ability to “discuss or communicate their reactions” to statistical products. Either way, both definitions anticipate statistical literate persons to be only consumers of statistical products which is in opposition to the view of a data literate individual who is both consumer and producer of data. It resonates with the view of Frank et al. (2016) who argues that data literacy adds to statistical literacy which developed first – in the era of limited access to data when people had to rely on intermediaries like press to access and interpret data for them.

2.3 Its “Fit For” Knowledge Management

The direct relationship between data literacy and knowledge management does not seem to be in the “research spotlight” yet. However, thanks to its position in the data-information-knowledge-wisdom pyramid, data literacy is evidently well-connected to successful knowledge acquisition. As data literacy serves as a precondition to become information literate, the relation of data literacy to knowledge management can be derived through its middleman – the information literacy.

Proceeding from his analysis, O’Farrill (2008) excellently names “the state of the relationship” between knowledge management and information
literacy as “preparation for arranged marriage”. The profound link between those fields is obvious, but it is still waiting for its interconnection by academics from two camps. O’Farrill emphasizes the learning processes as the main meeting point of information and knowledge and along with Marcum (2002) criticizes that the “reception of information is equated with knowledge acquisition in a rather unproblematic way”. According to O’Farrill (2008) knowledge management also “lack a robust understanding of effective information use in the organization” which is supported by Oman (2001), Cheuk (2008) who examined information literacy in company environment and realised that the failure of many knowledge management projects was caused by inadequate information literacy skills.

In 2016 Virkus conducted a content analysis of the literature about knowledge management and information literacy published in the period of 1990-2016. His study confirmed a strong link between these fields which was supported by Whitworth (2014) who believed that information literacy was “an essential and integral competency for both knowledge worker and effective knowledge management” or van Rooi and Snyman (2006) who acknowledged corporate information literacy as one of main knowledge management areas where library and information science professionals can contribute. Saito (2007) in his doctoral thesis even claims that “knowledge management seemed to be a natural extension to the field of library and information sciences”. Nevertheless, Virkus concluded that research of information literacy in the context of knowledge management was insufficient and short of empirical studies and he naturally followed the call for further research in this topic of several authors before him (e.g. Thompson 2003, De Saulles 2007).

The presented research of O’Farrill (2008) and Virkus (2016) implies that not only data literacy, but information literacy as well deserve more attention in the context of knowledge management and require further research.

2.4 Current State of the Data Literacy Research

The term of data literacy is well established which resulted in many different approaches to its definition. Van der Wal et al. (2017) strengthen the importance of data literacy as one of the technological skills necessary for graduates of technical universities. Koltay (2015) circumscribes data literacy in relation to other types of literacies like information or statistical literacy; on the other hand, Gould (2017) emphasizes data literacy as a part of statistical literacy. Wang, Wu, Huang (2019), Burns, Matthews (2018) or Halliday (2019) underline data literacy in context of a specific field like safety management or journalism while Prado and Marzal (2013) call for complex approach to data literacy definition. Gray, Gerlitz, Boungre (2018) and D’Ignazio, Bhardava (2015) also ask for expansion of the term of data literacy (e.g. Big Data literacy or data infrastructure literacy) to emphasize obvious aspects. Pedersen and Caviglia (2019) perceive data literacy as a compound competence and what is more, the authors explore data literacy as a group competence. Research of Grillenberger and Romeike (2018) enriches the topic with a model of data literacy competences which is clearly inspired by Ridsdale et al. (2015).

Nevertheless, the measurement of data literacy seems to be still in its infancy. Pratama and his team (2020) has published a preliminary study of their assessment instrument tested on 94 junior high students which is according to their conclusions ready to test initial level of data literacy. Another initiative to measure data literacy also originates in south-east Asia where a team of Lusiyana (2020) aims to prove effectiveness of MIRECAL learning model.

Furthermore, there are also business initiatives like QlikTech’s (2018) Data Literacy Project which focuses on corporate data literacy whose measurement has three components: employees’ individual data literacy skills, the accessibility of the right data for decision-making in a given job position and the widespread use of data across the company. Based on the scores of corporate data literacy QlikTech also came up with Data Literacy Index which correlates data literacy levels to measures of corporate performance and thus points out what business value can company attain with a given level of corporate data literacy.

In the field of teaching data literacy, the research has been richer and more fragmented. In 2015 Maybee and Zilinski came up with a framework for teaching data literacy based on a method of informed learning, while D’Ignazio, Bhardava (2016) set pedagogical principles to keep in mind when developing tools or interactive applications for teaching data literacy. The conceptual approach to teaching data literacy was extended by Wolff, Wermeling and Petre (2019) who pilot a method for teaching data literacy at middle schools on complex data. Moreover, the social benefits of data literacy and current educational models were examined in Pangrazio, Sefton-Green (2019).
More specific methods or approaches are brought by the research of Wallner, Kriglstein (2011) or Gäbler et al. (2019) who focused on design of interactive application and games. On the other hand Nolan and Perret (2016) come up with “ideas and assignments” how to effectively involve statistical visualizations into undergraduate courses or Halliday (2019) who developed set of exercises for economic students.

3 PLANNED RESEARCH

To improve anyone’s data literacy, first of all, it is necessary to determine what the start line is and what is to be achieved. That is why my research aims to define Data Literacy Indicator to measure data literacy as a level of maturity. Thus creating a Capability Maturity Model (CMM) for Data Literacy is an essential part of it.

The second part of the research is focused on the design of methodology for teaching data literacy for different levels of Data Literacy Indicator. As the Data Literacy Indicator will generally define sets of abilities needed to achieve a certain level of data literacy, the indicator then will naturally serve as a prescript of what skills need to be taught to move up to a higher level.

3.1 Expected Contributions

The contribution of my research resides mostly in the research artefacts. The Data Literacy Indicator as a product of a maturity model brings two beneficial aspects – while it naturally measures the level of data literacy at a given moment, it also allows to generally define sets of abilities needed to achieve a certain level of data literacy. From this assumption, I derive the usage of the Data Literacy Indicator.

I expect that the constructed indicator will be suitable to measure individuals’ as well as organizations’ data literacy. In companies and state institutions, it can serve as a tool for specification of data literacy competencies linked to different job positions and of how to gain the required knowledge and skills and eventually for the development of analytical culture. In schools, it should be used to define appropriate capabilities of data literacy for different grades, to continuously and critically measure its students’ progress in data literacy along their studies and most importantly to prepare adequate educational programs to acquire these capabilities.

While the Data Literacy Indicator serves as a critical assessment where we are and where we need to go, the methodology for teaching data literacy appropriately translated into tailor-made educational programs brings solution to the knowledge and skills gap in data literacy. Based on the measured level of indicator it will offer a path of concrete steps to follow in order to reach the targeted level of data literacy. As the measurement of data literacy should take into account differences of subjects’ domains or students’ highest level of education acquired, the methodology aims to be tested and tailored to these specifics as well. However, the main objective and a stepping stone is to create a methodology for teaching data literacy at schools (from middle schools to universities).

3.2 Selected Research Methods

The proposed research is clearly design-oriented and intends to contribute to the academic world as well as to the public with two artefacts – the Data Literacy Indicator (CMM) and methodology for teaching data literacy in relation to the measured level of the indicator.

The development of the Data Literacy Indicator is based on the People Capability Maturity Model (PCMM) as maturity assessment models are used as “an instrument for systematically documenting and guiding the development and transformation of organizations” (Paul et al. 1993). Its offspring, the People Capability Maturity Model, then focuses on the development of people competences and the measurement of the competences maturity. We generally understand maturity as a level of sophistication, here it clearly serves as a measure for capability evaluation (De Bruin et al. 2005). The capabilities are characterized by specific areas, so called dimensions, which encompass “different aspects of the maturity assessment’s object” and are “further specified by a number of characteristics (practices, measures or activities) at each level” (Raber, Winter, Wortmann 2012).

I plan to approach the maturity assessment with quantitative methods as used by Lahrmann et al. (2011) or Raber, Winter, Wortmann (2012). Their quantitative analysis is based on Item Response Theory (IRT) which is “a collection of mathematical models and statistical methods used for two primary purposes: item analysis and test scoring” and is “used with data arising from educational tests of ability, proficiency or achievement” (Millsap, Maydeu-Olivares 2009). The theory builds on the hypothesis that the probability of correct answer to an item (the question) is a mathematical function of the respondent and the characteristics of the item.
The IRT will be applied to a test/questionnaire to measure data literacy. The test questions aim to assess all dimensions of data literacy which represent areas of knowledge concepts and skills required to be able to “read and speak data”. By expanding Ridsdale’s et al.’s (2015) and Grillenberger and Romeike’s (2018) models of data literacy competencies, I derived five dimensions of data literacy: (1) Data concepts, ethics and protection; (2) Analytical principles and models; (3) Data collection and preparation; (4) Data analysis and evaluation; (5) Data communication and decision-making. Every dimension comprises of a specific set of competencies which will be measured by the test (e.g. an ability to assess relevant data sources or an ability to access data are competencies covered in Data collection and preparation dimension). By clustering the respondents’ test results which indicate respondents’ level of sophistication in different areas of data literacy will be used to establish the maturity levels.

Regarding the second artefact, the methodology for teaching data literacy, I would like to base its design on the use of various case studies. More precisely I would like to follow the approach in Wolff, Wermelinger and Petre (2019) which for the design of data literacy activities applied method called research through design in which “design practice is applied to the creation of artefacts as a way of exploring solutions to problems”. The method comes from interaction design research in human-computer interaction and as stated in Wolff, Wermelinger and Petre (2019), “new knowledge is constructed by undertaking activities associated with design, such as iteratively creating and testing prototypes to understand and solve a problem and to act as a focal point for discussion by making interactions observable”.

At the moment I am at the beginning of the first case study preparation. Its main objective is to verify the pilot version of a questionnaire measuring the data literacy maturity of the freshmen students at university. The test would be included in the first lesson of a new subject introducing application of data analysis by means of an interactive online game which promises high number of respondents with several classes each semester and allows to measure the entry level of the high school graduates (without the interference of higher education).

4 CONCLUSIONS

With the accessibility of vast volumes of data everywhere, data literacy is a must in order to fully participate in a modern society. Nevertheless, if we want to be effective in enhancement of our knowledge and skills in the domain, we have to be able to mark our start line and to specify what level of data literacy we want to achieve. Thus appropriate measurement of data literacy is required and should swiftly complement the recent boom in data literacy teaching initiatives and research.

REFERENCES

Next Step: Data Literacy Measurement


