Using Participatory Design to Gain Insight into How Students Make Sense of Data in Their Lives

Rotem Israel-Fishelson
University of Maryland
rotemisf@umd.edu

Peter F. Moon
University of Maryland
pmoon@umd.edu

Daniel Pauw
University of Maryland
dpauw@umd.edu

David Weintrop
University of Maryland
weintrop@umd.edu

ABSTRACT

Data is ever-present and increasingly shaping the lived experiences of today's students. This study examines high school students' perceptions of data using various analytical and methodological lenses. During participatory design sessions, students were given opportunities to express their understanding of the term 'data' and its relevance to their lives, their awareness of stakeholders collecting their data, and their concerns regarding data usage. This work sheds light on students' perceptions of data, discerning areas where their understanding is robust and identifying critical gaps. In doing so, it contributes an in-depth examination of an area that has received limited attention in previous studies but is of growing importance. Collectively, this research helps understand students' perceptions of data, concerns, and gaps and, in doing so, can guide educators and help inform data science curricula designers in developing learning opportunities to prepare students to be responsible and informed data-literate citizens.

CCS CONCEPTS

• **Social and professional topics** → Professional topics; Computing education; K-12 education.

KEYWORDS

Data Science, Data Literacy, Participatory Design

ACM Reference Format:
Rotem Israel-Fishelson, Peter F. Moon, Daniel Pauw, and David Weintrop. 2024. Using Participatory Design to Gain Insight into How Students Make Sense of Data in Their Lives. In *Symposium on Learning, Design and Technology (LDT '24), June 21, 2024, Delft, Netherlands.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3663433.3663461

1 INTRODUCTION

Today's students are increasingly immersed in a data-driven world where information and technology are central to their lives. In navigating this digital landscape, they constantly encounter data and data-driven technologies in both obvious and unexpected ways. The data today's students create and consume drive countless algorithms, shaping the content they see in their social media feeds and influencing the products they are shown in advertisements. Data informs countless yet very consequential facets of their lives relating to education, health, finance, and interactions with both peers and formal organizations. Therefore, helping today's students understand the role of data and its impact on their lives is an essential literacy [23]. Equipping them with the necessary knowledge and skills to recognize, comprehend, and interpret data is essential for their future success [34, 54]. In response to these needs, there is a growing demand to teach data literacy and foundational data science concepts and practices to all students as part of K-12 educational curricula [29, 53]. This has resulted in the rapid growth of research focused on K-12 data science education [38, 54].

Data literacy encompasses a range of skills, including the ability to analyze, evaluate, and draw meaningful insights from the data [16]. It involves understanding how data is collected, processed, and used to inform decisions. Importantly, students need to evaluate and understand the larger context in which data is situated, including questioning the data sources, data collection methods, and the potential biases within datasets [30]. A good understanding of data and its uses can significantly impact personal and civic engagement. It can help them comprehend and interpret the information they create and consume, leading to informed decision-making [3]. This can improve personal decision-making in short- and long-term financial planning and even lead to more thoughtful participation in civic processes such as voting [22, 31]. This is especially relevant for students historically excluded from computing fields, including BIPOC, female and non-binary, and neurodiverse individuals. These groups have been disproportionately impacted by biases and predatory applications of data-driven algorithms [7, 40, 41].

In their study, Biehler et al. [7] proposed that data science education must not only cover algorithms, data analysis, and data visualization but also equip students with the ability to think critically about collecting, producing, and utilizing data for various purposes. Data science curricula should educate students on their role as both producers and consumers of data and emphasize the risks associated with using data without considering its social effects [49]. Students need to comprehend how data influences their daily activities and understand their role as "data agents", i.e., their active engagement with data and its interpretation [51]. Such awareness could empower them to recognize the potential risks and benefits associated with data, enabling them to make informed decisions

LDT '24, June 21, 2024, Delft, Netherlands

ACM ISBN 979-8-4007-1722-2/24/06
https://doi.org/10.1145/3663433.3663461

regarding their digital footprint, privacy, and security [24, 33]. Simultaneously, it is crucial to assist students in recognizing the key stakeholders that collect their data and for what purposes [42]. In an interconnected world, data is gathered by various entities and stakeholders. By understanding their motivations, students can make informed decisions about the data they share, the platforms they engage with, and the privacy settings they utilize [9, 20].

Advancing equity and social justice is a crucial driving force behind data science education in K-12. By incorporating these principles into the curriculum, we can ensure that all students, regardless of socioeconomic background or demographic factors, have equal access to the skills required to succeed in a data-driven world [16]. This approach helps students gain essential data analysis skills and encourages them to develop a critical lens for examining social disparities in representation. Understanding the impact of data on different communities allows students to relate to real-world issues and encourage fairness in decision-making processes [55].

This work used participatory design as a means to gain insight into high school students' experiences with and perceptions of data. Participatory design (PD) is a methodological approach that involves end-users in the design process to democratize design and remove the power differential between designers and users [17, 18]. Through this approach, designs reflect the voices and values of all stakeholders and recognize that the end users, in our case, high school students, have critical and essential insights and expertise that can productively inform the design process [8]. For this work, PD was used to help us understand student perceptions of the data they produce, where it goes, where the data they consume is coming from, who the stakeholders interested in their data are, and what concerns students have about data. We examine these perceptions and analyze which personal data literacies (as proposed by Pangrazio and Selwyn [42]) the students possess and what the gaps are. More concretely, this paper pursues the following set of research questions:

RQ1. How do students conceptualize the term data?

RQ2. How do students perceive their role as data agents?

RQ3. Who do students think the key stakeholders are who are collecting their data, and what are the stakeholders' motivations for doing so?

RQ4. What concerns do students have about how their data is collected and used?

In conducting this work, we seek to better understand how today's students make sense of the data-driven world in which they are living. In doing so, we contribute to the growing data science education literature on students' perceptions of data and data science, which is essential for understanding how best to introduce students to these critical concepts and skills. Further, this work can inform the design of future data science curricula and tools that leverage learners' lived experiences in a data-rich world as contexts for situating data science instruction.

2 PRIOR WORK

2.1 Data in Students' Lives

Data is playing an increasingly significant role in students' lives. Students growing up today are continuously generating and sharing personal data through social media, online learning platforms, or wearable devices [39]. Participatory culture theory [27] asserts that such digital environments enable students to engage actively in content creation, sharing, and collaboration. Students are empowered to consume content and generate their own data through text, images, videos, and other forms of self-expression. This active participation reflects their transformation from mere consumers to data producers, thus creating and shaping their digital identities [27]. These virtual and quantified identities consist of the digital traces created from online interaction with others and informational artifacts, consciously or unconsciously [9]. When we connect Participatory Culture Theory with data science, we observe a fascinating intersection. Participatory culture thrives on the creation of user-generated content, and data science techniques are applied to analyze this content, providing valuable insights into their behavior, preferences, and trends. Students can analyze and interpret the data they generate and foster a deeper understanding of the role the data plays in their lives.

As technological advancements continue to shape the digital landscape, the amount of data created and consumed by students is growing, with a vast and arrayed set of stakeholders participating in the data collection enterprise. These stakeholders include technology companies, financial organizations, educational institutions, and governmental agencies, both domestic and foreign [15]. Students experience a world where their personal data may be used to shape their opportunities and identity. The consequences of this are evident in various ways, from algorithmic biases to filter bubbles created through personalization [4, 45]. Therefore, students' perceptions of data can vary widely as they navigate this digital landscape. Some students may see data as a helpful tool for personalization and convenience, while others might view it with skepticism and concern over potential misuse or invasion of privacy [20]. It is thus important to ask what the term data means to students and how they perceive data's impact on their lives. These questions are especially important, given the tendency of students to mediate much of their lives through digital platforms that regularly capture their interactions [9].

Prior research shows that students have a general grasp of the fact that their digital information is being gathered through their online activities, which can have an impact on their privacy [35]. Students are often skeptical about how various entities collect, store, and use their personal data. This ambiguity sometimes arouses their concerns but does not necessarily change their behavior [20]. Given the central role of data in the lives of students and the complex relationship between personal data and privacy issues [33], it is important to foster a deep understanding of the impact of data on them. Furthermore, there is an increasing demand for education on data literacy that covers a broader range of viewpoints and emphasizes contextual understanding [24].

2.2 Personal Data Literacies

The growing importance of data and data literacy is reflected in both the popular press [32, 50] and the academic literature, which includes recurring calls to improve students' data literacy, data modeling, and data analysis skills [6, 19, 21, 25]. Given the diverse applications of data literacy across various fields, researchers have put forth several definitions to describe the essential competencies of data literacy [16]. While some definitions focus on computational and technical skills related to statistics, machine learning, data processing, and visualization, others also stress the importance of understanding the entire data life cycle and how our digital culture affects it [1]. Wolff et al. [56] compared different definitions of data literacy and found that there are different levels of granularity between and within them. While only a few definitions refer explicitly to the execution of the inquiry process, most definitions refer to data analysis skills, such as preparing data for analysis, understanding data types, and combining quantitative and qualitative data. The small overlap between these skills highlights the differences in the level of expertise required.

Data literacy goes beyond grasping numbers, graphs, and data manipulation. It encompasses critical thinking about personal data intricacies and their wider implications on our society [11]. Pangrazio and Selwyn [42] proposed the concept of "personal data literacies," which extends beyond traditional data literacy by focusing on the nuances of personal digital data. It refers to the ability to critically engage with, analyze, and make informed decisions about one's personal digital data. The personal data literacy framework encompasses five key domains: (1) Data Identification, (2) Data Understandings, (3) Data Reflexivity, (4) Data Uses, and (5) Data Tactics. *Data Identification*, the first domain, focuses on recognizing the types of personal data generated and collected online and offline. This refers to data that students generate (e.g., uploading a post to a social network), data extracted from systems used by the student (e.g., health apps tracking the steps made by the student), and data processed about the student (e.g., ads on search engines based on student's previous searches). By being able to identify what constitutes personal data, students can make informed decisions about their online activities and digital footprint. *Data Understandings*, the second domain, goes beyond mere recognition and delves into comprehension. It focuses on identifying the sources and purposes of collecting and using personal data. Additionally, it involves grasping the implications of sharing personal data, understanding the concepts of data privacy and security, and comprehending the potential risks associated with data breaches and misuse. *Data Reflexivity*, the third domain, introduces a critical element to personal data literacy. It encourages students to reflect upon their data-sharing practices and consider the societal norms that shape these practices. This domain prompts students to question the motivations behind data collection and contemplate the potential biases embedded in algorithms. The fourth domain, *Data Uses*, is a call to action. Students are encouraged to explore the manifold ways in which their personal data is used. This involves understanding how data is harnessed for targeted advertising, decision-making processes, and even influencing political opinions. Such knowledge empowers them to make conscious choices about the data they share. The last domain, *Data Tactics*, equips students with practical strategies to manage their personal data, gain a sense of agency over their data, and protect their privacy. By cultivating these five domains, students can develop a holistic understanding of how data affects their lives, become more discerning consumers of digital content, and make informed decisions that align with their values and interests [42]. In this study, we examine the students' perceptions of data in their lives and analyze which of the domains of personal data literacies they hold.

2.3 Students Perceptions of Data and Privacy

As the use of digital platforms and online services becomes ubiquitous, understanding students' perceptions of data and privacy has become imperative. In recent years, various studies have examined the way young people perceive the data in their lives, focusing on their awareness, attitudes, and behaviors regarding the privacy of their digital footprint. One aspect in this context examines how students concretely perceive the term data and how they understand the data lifecycle. [9] found that students aged 11-18 had different interpretations of the nature of data and a broad understanding of the data life cycle. However, most of them had difficulty connecting to data on a concrete and personal level. Understanding the knowledge base of young individuals in this regard is crucial for developing effective educational initiatives and privacy protection measures. [33] evidence review distinguished between students' awareness of interpersonal, institutional, and commercial privacy and pointed to the influence of personal expression and belonging on these definitions. They also pointed out the gaps in the understanding of institutional privacy, which refers to data collected by educational, governmental, and medical organizations. Another crucial aspect of the study examines students' attitudes towards privacy, including analyzing behaviors regarding privacy settings and taking steps to protect their personal information. This includes exploring the extent to which young people actively manage their privacy settings on different platforms, their understanding of consent mechanisms, and the adoption of security measures to protect their online presence. A study that examined these aspects among 3000 teenagers found that while they are aware of privacy issues, they are not willing to take additional measures to protect their personal data [20].

Alongside these studies, others have examined methods and tools designed to foster data literacy and raise awareness towards aspects of privacy among students through PD sessions. In a study conducted by [28], game- and story-based resources were utilized to teach students about privacy, while another study explored three projects and examined the students' level of understanding and awareness of critical digital practices [43]. We continue this line of research with the aim of examining how students perceive data in their lives and what personal data literacy they possess.

3 METHODS

To answer our stated research questions, we conducted three PD sessions, each with different participants. The first two sessions took place after school, each lasting 90 minutes, and the third was more extensive and carried out over three five-hour sessions during the school day as a part of the end-of-year workshop. In these sessions, the participants went through several design activities (as

will be detailed below) to provide us insights into their understanding of the term "data", how they perceive themselves as consumers of data, their awareness of the stakeholders who collect their data, and the concerns they have about data usage. Additionally, we used the domains proposed by Pangrazio and Selwyn [42] to serve as an analytic lens into their personal data literacies.

3.1 Participants

A total of 28 high-school students participated in the PD sessions (21 in the after-school sessions and 7 in the during-school sessions). Participants included 17 males and 11 females, 22 Black/African American students, 2 American Indian or Alaska Native students, one Hispanic, and one White student. The average age was 16. All participants had taken a computer science course before our sessions, and only one had taken a data science course. Parent/guardian consent and child assent were received from all participants.

3.2 Participatory Design Structure and Data Collection Activities

The PD sessions took place in school classrooms and had students work through a series of discussions and hands-on design activities drawn from the participatory design literature (e.g., [13]). This methodological approach was used to provide a means for the voices and values of the high school students to be shared while removing the power differences between them and the researchers [17]. Moreover, this approach was found to be essential in fostering fundamental understandings of learning processes while prioritizing the development of forms of knowledge and effective interventions among individuals and groups that have historically been marginalized, all in the interest of achieving more equitable and cooperative outcomes [2]. The goals of these design activities were to provide different ways for students to share their perceptions and experiences with data. The PD facilitators introduced each activity and circulated between groups while the design activities were conducted. Here, we briefly describe a subset of PD activities to provide a sense of the PD sessions. The order they are described below matches the order of the sessions.

3.2.1 Data in Your Lives (20 minutes). During this activity, the students were asked: *"What do you think when you hear the word data?"*. They scanned a QR code and logged their responses using an online data collection tool that provides real-time results that can be shared with participants. Then, the facilitator presented the word cloud from their responses to engage the participants in an open discussion about the data in their lives. This activity provided an initial venue for students to share their understanding of the term "data" and provided a useful starting point for both our investigation and the PD that followed.

3.2.2 Empathy Map (15 minutes). This activity was based on the User-Centered Design concept of a persona [36]. It was intended to help students rapidly reflect on their identities and present how data affects their lives. Participants were divided randomly into groups of 2-3 students. Each group received a large sticky paper that had a circle in the center and was separated into four quadrants labeled as follows: "Does", "Uses", "Concerns", and "Interests". In the center of the paper, participants were asked to sketch a typical student and create a name and brief description of who the imaginary student is. The goal of this activity is to allow students to use the imaginary student persona as a proxy to express their own habits, interests, and concerns without having to personally claim them, allowing for honest expression behind a protective "other." Next, each participant was asked to add sticky notes to each quadrant to describe the student they were creating (Figure 1). The groups were provided with guided questions to help them reify their student: *What do they do that creates data? What do they do that uses/consumes data? What concerns do they have about data? What are they interested in?* This activity resulted in nine Empathy Map artifacts describing different personas aged 16-19.

3.2.3 What's on Your Plate? (25 minutes). In this design activity, which took place in the third session, students were given a paper plate and markers and were asked to reflect and illustrate what data they create and consume during their daily routines. This activity was inspired by the work of Panke [44]. The students were given the liberty to express themselves on the plate creatively, with both explicit descriptions and more abstract drawings. Students first worked on their plates alone before sharing them with their peers. The goal of this activity was to illustrate to students that their daily activities, such as using social media, generate data and help them draw connections between their personal lives and data science.

3.2.4 Stakeholders Board (20 minutes). In this activity, students were given a big piece of paper divided into three parts. In the center of the page, they were asked to write all the stakeholders interested in the data they create daily. On the top half of the page, they were asked to add sticky notes describing why companies collect data. On the bottom half of the page, they were asked to add sticky notes to name the different uses of the data. The exercise aimed to gain insights into students' perceptions and comprehension of the various stakeholders involved in data collection, and groups that they associate with data science.

3.2.5 "Selfiecity" Data Exploration (40 minutes). In this activity, participants were introduced to the "Selfiecity" project (selfiecity.net), a research project that examines the patterns and characteristics of selfies (self-portraits typically taken with smartphones) shared on social media platforms [52]. The project explores the cultural phenomenon of selfies by analyzing thousands of selfies from five cities around the world. The website for the project presents rich media infographics and findings about the demographics of people taking selfies, their poses, and expressions [26]. In addition to these summary graphics and statistics, users of the site can also click on individual selfies within groups to see the raw data being used to generate these infographics. During the design session, participants developed three interesting insights and discussed the following questions with the group: *How could they use data like this to examine their lives?; What data do they create like this in their own life?;* and *What (or who) do they think is missing from this project?* This activity encouraged students to be critical and reflect on the potential biases inherent in the data collection and presentation methods, such as which cities were chosen and how the data was categorized. Afterward, the participants were asked to develop a research question, plan a study related to their lived

Table 1: Coding Manual

Code	Definition	Examples from Participants' Data
What	Student-provided definitions of data	"A set of gathered information"; "Various amounts of information being stored into a specific location (a computer or other virtual forms)"
Who	Student-provided entities and organizations interested in data	"Government"; "Foreign Agencies"; "Schools"; "Banks"; "Hospitals"; "Social Media Platforms"; "Hackers and Phishers"
Why	Student-provided reasons for entities and organizations to collect data	"To recommend something better"; "To improve their work"; "To make decisions"; "To control you and monitor your moves"
Where	Student-articulated sources of data (i.e., defining where data comes from)	"Social Media (e.g., Instagram, Twitter)"; "Netflix"; "Hulu"; "Canvas"; "Google"; "Online shopping"; "iMessage"
Concerns	Student-articulated issues regarding the collection, storage, sharing, and use of data	"Being tracked by the apps used"; "Being recorded without his consent"; "Someone wants to use your data against you"

experiences inspired by Selfiecity, and address the issues and concerns they raised regarding data collection. This activity revealed what students understood about data collection and their concerns about data sharing.

3.2.6 Summative Questionnaire (5 minutes). At the end of each session, participants completed a short online questionnaire that included demographic questions and summative questions about their perceptions and interests related to data and data science. The students were asked about their usage of social networks and mobile apps and their preferences for topics to study within data science courses.

3.3 Data Collection Procedure and Analysis

The design sessions were led by two researchers acting as facilitators and were audio and video recorded. All design artifacts created during the sessions were collected for analysis. After being transcribed, the data was analyzed using an open coding approach to identify emergent themes [48]. First, the researchers reviewed each segment of text, identifying and labeling prominent concepts or themes without preconceived ideas or categories. The researchers then went through the descriptive labels and assigned abstract categories to group them. The analysis process resulted in five group-level codes attending to different aspects of student perceptions of data. Table 1 presents each code, including the definition and an example from the data. Two researchers systematically and independently coded the data using the coding scheme in Table 1 [46]. The researchers went through each segment of text and coded it according to one of the five codes. In addition, the researchers added notes or memos about their thoughts, questions, and insights. The resulting codes were compared, and an inter-rater reliability assessment using Cohen's Kappa [14] yielded a coefficient of 0.67. The differences in coding results were primarily related to the categories of "What" and "Concerns". In light of this, the researchers refined their definitions and performed another round of independent coding. The inter-rater reliability assessment of the second coding yielded a coefficient of 0.8. The researchers discussed and resolved the remaining discrepancies between their analyses until full agreement was achieved.

4 FINDINGS

The various design activities helped reveal students' attitudes toward and perceptions of data and its impacts on their lives. In the following sections, we synthesize the findings from the design activities through five thematic lenses, according to our research questions: data conceptualization, students as data agents, data-interested stakeholders, reasons for data collection, and concerns regarding the data. In addition, we illustrate how these themes correspond with the various domains of the personal data literacies framework proposed by Pangrazio and Selwyn [42].

4.1 RQ1: Data Conceptualization

During the PD sessions, students were given several opportunities to express their conceptualizations of data. In analyzing the various expressions of what data is, two distinct ways of thinking about data emerged: pre-analysis and post-analysis. Pre-analysis conceptions of data refer to students talking about raw, unprocessed data. For example, one of the participants claimed that data is *"Anything that can be stored"*, while others claimed data refers to *"numbers"*, *"binary numbers"*, and *"your grades"*. This is largely in line with the concept of raw data or individual observations that might be stored in a data table or database. In contrast to this pre-analysis view, other participants provided definitions referring to structured and organized data and data that has undergone some kind of processing or analysis. We classify this second type of definition as post-analysis. Among the definitions of data classified in this category were: *"An organized set of information"* and *"Something that gives out information based on a study"*. This perspective seems to be more in line with what in data science would term "information" – data that has been analyzed with insights distilled and shared. There were also four cases that revealed a lack of clarity about the distinction between data and information. For example, one of the participants stated that data is *"information that is being collected"*. Although this answer does include the term "information" in it, it blurs the distinction between raw collected data and information or insight derived from that data after being analyzed. Responses such as this highlight the need to introduce students to the meaning of specific data science terminology, as understanding how data becomes information is a foundational concept in data science. This analysis reveals an interesting nuance in the practice of data science, that data can be seen as both an input and output of the process,

Figure 1: Examples of "April" (left) and "Tiffany" (right) personas from the Empathy Map activity.

helping students understand this difference should be a part of data science instruction. More concretely, instruction in this area could help students understand the difference between data and information and help them understand the process of turning one into the other [10].

4.2 RQ2: Students as Data Agents

Throughout the design sessions, we examined how students perceive the data in their lives and, the data they produce and consume, and whether they perceive themselves as "data agents" [51]. We asked to see if and how Data Identification, the first domain in Pangrazio and Selwyn's [42] framework, is reflected in participants' answers and whether they identify the types of personal data they generate. Interestingly, and not surprisingly, students repeatedly stated that they create and consume data mainly through social networks and mobile applications. The questionnaire analysis showed that 92% of the participants (26 out of 28) use social media applications in their daily lives. Indeed, when they were asked to provide the five most used applications, the social media category was the most prominent (mentioned 39 times), followed by entertainment (14 times), communication (12 times), and gaming (9 times).

The Empathy Map activity and the personas artifacts also reveal how students think about the ways they create and consume data. The personas often shared characteristics with members of the study group, indicating that they did use these personas as proxies for their own interests and feelings. Eight of the nine personas included a reference to social media as part of the description of how typical students create/consume data. Some groups expanded and described the specific uses. For example, one group created the persona of April (Figure 1, left), an African American bilingual basketball player and TikTok lover who posts pictures on Locket with friends, posts on her Instagram stories, and listens to music. Through this description of April, we can see how central social media and smartphone applications are to the identity of today's high school students. Another example is the persona of Tiffany (Figure 1, right), a 17 y/o athletic student who posts videos on social media, submits assignments, and attends classes, again showing how online tools that collect student data are an important part of what it means to be a high school student today.

The analysis of the artifacts from the "What's on your plate?" activity revealed that all students depicted their consumption and production of data in reference to social media. Communication apps, such as Facetime and iMessage, were also present in all plates. Additionally, utility apps, like Calendar and Chrome, were featured on four plates, while navigation apps, e.g., Find My Location or Google Maps, and entertainment apps, e.g., Hulu, appeared three times. Video game apps were mentioned twice. Notably, in two plates created by participants, activities such as reading books, working, and walking were also highlighted as data creation and consumption sources. Only one student referred to the school's online learning management system, i.e., Canvas, as a means of creating and consuming data.

As part of the discussions held during the PD sessions, they were introduced to the infographic "Data Never Sleeps", which shows the data being created each minute on the internet. The students were surprised by the amount of data collected from the various applications they use on a daily basis. A thematic analysis of the discussions revealed that social networks are the main source from which students see themselves as data producers. For example, when asked how they create data, one of the participants said: *"Anything we do on social media"*. Another participant referred to watching and uploading videos to YouTube: *"What you watch on YouTube. . . if you watch this type of stuff, they will collect data from that and say: Hey, you like this type of stuff, so we'll show you more"*. Besides social media, other ways that they create data mentioned were media platforms, mobile applications, online searches, emails, eCommerce websites, medical monitoring, educational/learning platforms, and financial transactions. Some participants noted how pervasive data creation is, saying things like "*Everything you do in daily life that's being recorded or someone's learning from it"*.

Upon examining the answers provided by the participants, it was found that they have a good understanding of when and how

they create data, especially in relation to social networks. However, when it comes to processed data concerning their habits, their only reference was to exposure to advertisements or personalized content. They gave an example of data collected about their viewing habits on a streaming service, which is then used to recommend movies they might like. These findings indicate that the participants partially possess the domain of Data Identification in Pangrazio and Selwyn's [42] framework. This is because most participants did not identify the third type of personal data, that is, data extracted from systems used by them. Even though one participant referred to the learning management system his school uses, there was no mention of the data extracted from such a system, for example, by the teachers and administration at the school. Also, it is likely that many participants produce data that is extracted from systems they use, such as sports and health apps, but these systems and platforms were not mentioned by the participants, suggesting a potential lack of awareness of the role they play in data collection.

4.3 RQ3: Who is Interested in Student's Data and Why?

Along with learning about students' awareness of how and when they are creating data, we are also interested in understanding their perceptions of who is collecting their data and for what purpose. Informed by the personal data literacies framework proposed by Pangrazio and Selwyn [42], we wanted to examine how the domain of *Data Understanding* is manifested in participants' answers and whether they identify and comprehend the entities collecting their personal data and those entities' goals. During the PD activities, students referred to various stakeholders who collect their data for different purposes. Among these stakeholders, corporations were mentioned the greatest number of times (17), followed by government agencies (11), social media platforms (8), educational institutions (6), medical entities and institutions (4), financial institutions (1), and even malicious agents (4) such as hackers. This diversity of stakeholders was summed up by one participant who said: *"Pretty much everyone who can profit off your data wants your data"*, while other students were less sure who wanted their data, as one participant said: *"That is a good question. I don't know!"*. These responses capture the wide range of awareness students have about who is interested in their data.

As for the reasons entities collect data, 15 out of 37 (41%) answers given by the students in the Stakeholders Board activity and the discussion that followed referred to their data being collected and used to improve the user experience. That is, to gather data to customize the services according to the users' needs or interests and to provide a personalized user experience. For example, one participant noted that *"TikTok collects data based on what you like, and it will show you videos based on that"*. Another participant wrote on the Stakeholders' Board that companies collect data *"to know what consumers like/don't like"*. The second most common reason given for stakeholders wanting to collect their data was that data can be used to inform decision-making. For example, one participant stated: *"You know, [companies] analyze it and make decisions based on it. It doesn't have to be political or economical"*. Another participant referred to decision-making and wrote on the Stakeholders' Board that *"The different uses of data are to make decisions about their company and to be able to solve their problems"*. This motivation of collecting data to inform decisions was mentioned six times during the sessions. Two other participants referred in their answers to colleges that utilize data to determine the acceptance of candidates and award scholarships. Given that the participants are high school students, college admission is a topic of interest to them, so it was interesting to see them connect these larger discussions on data to important decisions that will impact their futures. Six responses referred to companies' profit from data collection. For example, one participant wrote explicitly that *"They* [companies] *also use it for money"*. Another continued this argument and elaborated that *"It's important because they need to keep track of what's going on with their company because that's where most of their money comes from"*. It is evident from these responses that the students understand the fact that data is driven by business and that companies utilize data towards the goal of financial gains.

Three answers referred to hackers using data maliciously, such as stealing personal information. One of the references in the discussion was that *"With hackers and phishers, it's all a lot simpler. They just want your data so they can get all of your other stuff"*. Three participant responses indicated that they themselves had collected data to track the whereabouts of others. For example, one participant mentioned, "*I use the Find My Friend app to keep track of people's location; I use that every day.*" Lastly, another reason mentioned for using data was associated with regulatory matters. Two participants referred in their answers to the collection of data related to credit scores and social security numbers. For example, one responded, *"Your credit score keeps track of if your credit is better or worth over the years all the time. It keeps track of what you save"*. It is evident from the answers students provided in the various design activities that most of them demonstrated a broad understanding of the reasons entities collect data and who those entities are.

This analysis shows that most of the participants possess what Pangrazio and Selwyn [42] describe as *Data Understandings*. This means that they can recognize the origins and objectives of personal data collection and comprehend the consequences and potential risks of data sharing. However, it is worth noting that many of the responses here are general, lacking specifics of how particular data might be used. In the discussion of companies collecting data, for example, we note that students simply say companies use this data for "decisions" or "money." No comments indicated a specific understanding of the process by which data might produce greater revenue or profits for the company, such as connecting sales metrics to specific decisions and using that research to guide future decision-making. Discussion of "hackers" shared this generality: students indicated that hackers want access to their "stuff" but did not have specific concerns about what could be done with it or even which private information the hackers attempt to access.

4.4 RQ4: Impacts of Data and Students' Concerns

To better understand how students perceive the impact of data on their personal lives and their concerns about using data, we analyzed their responses in the opening discussion along with the persona created in the Empathy Map activity and the discussion around the "Selfiecity" Data Exploration activity. These activities provided

opportunities for them to reflect on their data-sharing practices and test their understanding of their implications. The analysis of these activities helped assess how Pangrazio and Elwyn's [42] *Data Reflexivity* domain was reflected in their perceptions. Also, the analysis helped to assess whether and how the domain of *Data Uses* was expressed. When participants were asked if they think data impacts their lives, six out of 28 immediately responded "Yes" without further explanation, while others provided concrete examples of situations where data impacts their lives. For instance, one participant said: *"You go to the doctor. The doctor is going to collect information on you... the doctor is trying to decide what to prescribe based on data that's been collected"*. Another participant gave an example related to the effect of data on receiving scholarships: *"Colleges use data, financial data, to see whether you should get a scholarship"*. A third response referenced the potential occupational aspect of data impact: *"You must have a certain amount of knowledge to get a job"*. These answers indicate that they have a certain degree of understanding that data has an impact on different areas of their lives but might not understand the full extent and impact of data in their lives.

During the various activities, participants also raised some concerns about the role that data collection and use play in their lives. Our analysis identified three themes of concerns that emerged from the personas: privacy and security, data collection methods, and unauthorized usage. Out of the 21 concerns raised, 12 were related to privacy and security and the possibility of information being leaked and used by other people. For example, as part of the persona activity, participants listed the following concerns: *"His data being used by someone who is not him"*, *"His data being publicly released"*, and *"Having little knowledge that he is being recorded without his consent"*. Five concerns raised by the participants were related to data collection. A second persona listed the following concerns: *"How many people actually collect data?"*, *"What do people do with our data?"* and *"How many people watch our data?"* Lastly, in four cases, the students provided more concrete concerns regarding unauthorized access, saying things like the persona: "*Feel like I'm being tracked by the apps used*", "*Information will get leaked b/c of the apps*", and "*Scared of other apps because they may take more information from you*". These statements suggest that students are aware and concerned about potential privacy issues with the use and collection of data.

The analysis of the discussion with the students about the "Selfiecity" data exploration activity also shed light on their concerns about sharing and collecting personal data and their doubts about potential biases. Several of them raised the question of whether anyone can use the photos they publish freely. This question reflects their concerns regarding the collection methods and the use of their personal data by other parties. In addition, the students raised questions regarding subjectivity in interpreting the collected data, i.e., the selfies. Interestingly, one of the participants commented that *"you don't really know if it* [the data] *is accurate"*. He claimed that sometimes teenagers use apps but don't report their real names, gender, or age, and therefore, it is possible that the information provided by "Selfiecity" is not reliable. Another participant referred to the data analysis and claimed the interpretation could be subjective*: "If somebody just wasn't smiling in the picture, you can't really determine their mood"*. Additionally, when they were asked in the discussion whether the data in "Selfiecity" was representative, one of the participants claimed: *"I feel like they did okay with the different cultures, but I feel that inside these countries, it's different. They don't represent all of America"*. The concerns raised here about the study largely focused on whether it was accurate and representative.

5 DISCUSSION

This research aimed to understand students' perceptions of and experiences with data. To answer our questions, we utilized participatory design activities to provide a platform for students to express their thoughts on the data in their lives. We analyzed these perceptions through various analytical lenses while evaluating their personal data literacies using the domains suggested by Pangrazio and Selwyn [42]. Our findings indicate that students possess most of the domains, but there are gaps, which we will detail below. First, we analyzed the students' understanding of the term "data", finding there to be two distinct conceptualizations of the term, with the difference being whether or not analysis had yet taken place. These findings resemble a previous study in which students conceptualized data by its structure, graphic configuration, digital traces, and the tangible output of a scientific investigation [9]. These findings demonstrate the importance of data science instruction in K-12 contexts to help students more fully understand what data is and the process of moving from raw data to information and insight. Data literacy and data science learning opportunities should help students understand what data is and is not, as well as how to go from data to information, knowledge, and, ultimately, wisdom [47]. Given the frequency with which students encounter data and the importance it plays in their lives, these aspects should be a core goal of K-12 data science education.

5.1 Students' Perceptions of Data Science

Our findings also shed light on how students see themselves as participating in a data-driven world. From the design activities and discussions, high school students mainly saw their creation of data through their engagements on social networks, mobile applications, and online searches. Interestingly, absent from these discussions on their role in the data ecosystem were activities related to their own data gathering and knowledge building. For example, participants never mentioned how or when they gather or analyze data in their own lives. Students also did not mention data practices associated with school, like doing research for a school project or conducting an experiment in science class, as data science practices. In this way, they viewed their role as passive data creators rather than active agents who contribute data and engage in data collection and analysis.

This disconnect between students' perceptions of their role in a data-driven world and the actual, meaningful data science practices they engage in reveals a significant opportunity for data science educators to draw on. Creating and using data is an important part of students' educational experiences. For example, conducting experiments in science classes or analyzing historical data in social studies or civics classes has students engaging in data practices. Further, by helping students see those activities like comparing prices

across different sites or looking up stats about their favorite athletes as meaningful data science engagements, we can help develop their identities as confident, active participants in a data-driven world while also potentially serving as important on-ramps to the larger field of data science. Developing and providing learning experiences that connect these data practices from academic and non-academic contexts is important. By gathering data that aligns with their interests, data science curricula can provide relevant and engaging content and increase student involvement in the field [37].

5.2 Students Perceptions of Who Does Data Science

At its core, data science aims to gather insights and derive new knowledge from data and has a pivotal role in advancing scientific understanding and discovery. However, our findings suggest that students' understanding of the key players and their reasons for collecting data are insufficient, non-specific, and largely focused on commercial purposes. First, students' perceptions of why companies collect data were overly general. Comments were made about companies collecting data for the purposes of greater profit, but there was little sense of how, specifically, the data collection might lead to gains for the company. Also absent from the discussion was the importance of data collection and analysis for non-commercial purposes, whether democratic, academic, artistic, political, interest-driven, or personally meaningful. The use of data and data practices to advance social and democratic causes and improve the lives and well-being of people is critically important and seemingly absent from the perceptions of data science we found. Chou et al. [12] noted that data science does *not "intrinsically improve social good. . . it is merely a force multiplier"* (p. 24), indicating that students must be actively taught ways that data science can be used to help improve society. Moreover, pedagogical support must be provided to create sustained engagement with data practices that will link the personal, social, and political aspects [51]. It is, therefore, important to include the larger perspective of the use of data and data science to help students understand the full breadth and importance of these ideas beyond capitalistic ends.

5.3 Students Perceptions of the Dangers of Data Science

While we were surprised by the students not identifying the data science practices they engage in and having a relatively narrow and commercially focused view of data science, we were not surprised by their ability to articulate various concerns and dangers related to data and data science. Our findings show that students have a relatively sophisticated and savvy perspective on concerns regarding data. Their perceptions revolve around privacy and security, data collection, and unauthorized usage. These concerns primarily pertain to the raw configuration of the data rather than the analysis conducted with it. Few concerns were voiced about disinformation, data accuracy, and bias; the only exception was in the Selfiecity activity, where students indicated that accuracy might be a concern but did not offer specifics on how the data might be inaccurate. It is, therefore, crucial to include the social, ethical, and equity aspects of data usage in data science curricula to help students understand these important dimensions of data collection and analysis [5].

6 CONCLUSIONS AND FUTURE DIRECTIONS

In a world awash with data, understanding, analyzing, and interpreting data is a fundamental skill for everyone. Thus, data literacy has gained prominence as a critical skill set for students to develop, and there are calls to instill it from an early age. Students equipped with a strong foundation in data literacy are better positioned to thrive academically and professionally and as engaged and informed citizens. Data literacy helps understand the methods of collecting and using data for various needs, from research to personally adapting content, analyzing and interpreting different forms of data, identifying misleading information, and evaluating the validity of claims. It improves critical thinking and fosters evidence-based decision-making. Therefore, education systems must adapt to incorporate instruction for data literacy into curricula, addressing both technical proficiency and ethical considerations.

This study can help inform the design and development of a data science learning experience that aims to introduce students to foundational ideas of data science. Our findings shed light on students' existing perceptions and knowledge about the role of data and its impact on their personal lives. They indicate which personal data literacy skills they possess and the gaps that remain. Acknowledging these gaps can greatly assist in adapting tools and curricula and improving learning experiences. Indeed, this is our intention: We hope that by directly addressing areas in which students are missing data literacies in our data science curriculum and doing it in a context that is personally relevant to their actual lives, they may not only be able to understand key data science concepts but also apply them to more than just capitalistic concerns. Designers, educators, and policymakers must recognize the significance of data literacy and work towards overcoming the challenges that hinder its effective implementation. Central to this effort is understanding what students understand about the role of data in their lives and designing and developing ways to help prepare them to be data-literate citizens. Furthermore, as data sources become more complex and the impact of data becomes even more impactful, it becomes increasingly important that students are prepared to navigate and excel in our data-driven world.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (Award # 2141655). Any opinions, conclusions, and/or recommendations are those of the investigators and do not necessarily reflect the views of the [Funding Agency].

REFERENCES

[1] [1]Acker, A. and Bowler, L. 2018. Youth data literacy: Teen perspectives on data created with social media and mobile devices. *Proceedings of the 51st Hawaii International Conference on System Sciences* (Hawaii, 2018), 1923–1932.

[2] [2]Bang, M. and Vossoughi, S. 2016. Participatory design research and educational justice: Studying learning and relations Within social change making. *Cognition and Instruction*. 34, 3 (Jul. 2016), 173–193. DOI:https://doi.org/10.1080/07370008.2016.1181879.

[3] [3]Bargagliotti, A. *et al.* 2020. *Pre-K-12 guidelines for assessment and instruction in statistics education II (GAISE II).* American Statistical Association.

[4] [4]Belitz, C. *et al.* 2023. Constructing categories: Moving beyond protected classes in algorithmic fairness. *Journal of the Association for Information Science and*

Technology. 74, 6 (2023), 663–668. DOI:https://doi.org/10.1002/asi.24643.
[5] [5]Benjamin, R. 2019. *Race after technology: Abolitionist tools for the new Jim code*. John Wiley & Sons.
[6] [6]Berman, F. *et al.* 2018. Realizing the potential of data science. *Communications of the ACM*. 61, 4 (2018), 67–72.
[7] [7]Biehler, R. *et al.* 2022. Research on data science education. *Statistics Education Research Journal*. 21, 2 (Jul. 2022), 1–1. DOI:https://doi.org/10.52041/serj.v21i2.606.
[8] [8]Bødker, S. *et al.* 2022. What is participatory design? *Participatory Design*. S. Bødker *et al.*, eds. Springer International Publishing. 5–13.
[9] [9]Bowler, L. *et al.* 2017. "It lives all around us": Aspects of data literacy in teen's lives. *Proceedings of the Association for Information Science and Technology*. 54, 1 (2017), 27–35. DOI:https://doi.org/10.1002/pra2.2017.14505401004.
[10] [10]Carlson, J. and Johnston, L. eds. 2015. *Data information literacy: Librarians, data, and the education of a new generation of researchers*. Purdue University Press.
[11] [11]Carmi, E. *et al.* 2020. Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*. 9, 2 (2020), 1–22. DOI:https://doi.org/10.14763/2020.2.1481.
[12] [12]Chou, S. *et al.* 2014. Democratizing data science. *Proceedings of the KDD 2014 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2014), 24–27.
[13] [13]Coenraad, M. *et al.* 2019. Enacting identities: Participatory design as a context for youth to reflect, project, and apply their emerging identities. *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (New York, NY, USA, Jun. 2019), 185–196.
[14] [14]Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*. 20, 1 (1960), 37–46.
[15] [15]Craig, T. and Ludloff, M.E. 2011. *Privacy and big data: The players, regulators, and stakeholders*. O'Reilly Media, Inc.
[16] [16]Deahl, E. 2014. Better the data you know: Developing youth data literacy in schools and informal learning environments.
[17] [17]DiSalvo, B. 2016. Participatory design through a learning science lens. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, May 2016), 4459–4463.
[18] [18]Druin, A. 1999. Cooperative inquiry: Developing new technologies for children with children. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (New York, NY, USA, May 1999), 592–599.
[19] [19]Engel, J. 2017. Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal*. 16, 1 (2017), 6.
[20] [20]Engels, B. 2019. Digital first, privacy second? Digital natives and privacy concerns. (Apr. 2019), 209.
[21] [21]Finzer, W. 2013. The data science education dilemma. *Technology Innovations in Statistics Education*. 7, 2 (2013). DOI:https://doi.org/10.5070/T572013891.
[22] [22]Fontichiaro, K. 2020. Data literacy and civic engagement: Polling and election data. *Teacher Librarian*. 47, 5 (2020), 51–53.
[23] [23]Franklin, C. and Bargagliotti, A. 2020. Introducing GAISE II: A guideline for precollege statistics and data science education. *Harvard Data Science Review*. 2, 4 (Oct. 2020). DOI:https://doi.org/10.1162/99608f92.246107bb.
[24] [24]Gebre, E. 2022. Conceptions and perspectives of data literacy in secondary education. *British Journal of Educational Technology*. 53, 5 (Sep. 2022), 1080–1095. DOI:https://doi.org/10.1111/bjet.13246.
[25] [25]Gould, R. 2017. Data literacy is statistical literacy. *Statistics Education Research Journal*. 16, 1 (2017), 22–25.
[26] [26]Investigating the style of self-portraits (selfies) in five cities across the world.: 2023. https://selfiecity.net/. Accessed: 2023-08-09.
[27] [27]Jenkins, H. 2009. Confronting the challenges of participatory culture: Media education for the 21st century. The MIT Press.
[28] [28]Kumar, P. *et al.* 2018. Co-designing online privacy-related games and stories with children. *Proceedings of the 17th ACM Conference on Interaction Design and Children* (New York, NY, USA, Jun. 2018), 67–79.
[29] [29]LaMar, T. and Boaler, J. 2021. The importance and emergence of K-12 data science. *Phi Delta Kappan*. 103, 1 (Sep. 2021), 49–53. DOI:https://doi.org/10.1177/00317217211043627.
[30] [30]Lee, V.R. *et al.* 2021. A call for a humanistic stance toward K–12 data science education. *Educational Researcher*. 50, 9 (Dec. 2021), 664–672. DOI:https://doi.org/10.3102/0013189X211048810.
[31] [31]Lee, V.R. *et al.* 2022. Taking data feminism to school: A synthesis and review of pre-collegiate data science education projects. *British Journal of Educational Technology*. 53, 5 (Sep. 2022), 1096–1113. DOI:https://doi.org/10.1111/bjet.13251.
[32] [32]Levitt, S. 2019. *America's math curriculum doesn't add up [Audio podcast]*. Technical Report #391.
[33] [33]Livingstone, S. *et al.* 2019. *Children's data and privacy online: Growing up in a digital age. An evidence review*. London School of Economics and Political Science.
[34] [34]Martinez, W. and LaLonde, D. 2020. Data science for everyone starts in kindergarten: Strategies and initiatives from the American Statistical Association. *Harvard Data Science Review*. 2, 3 (Sep. 2020). DOI:https://doi.org/10.1162/99608f92.7a9f2f4d.
[35] [35]Marwick, A.E. and boyd, danah 2014. Networked privacy: How teenagers negotiate context in social media. *New Media & Society*. 16, 7 (Nov. 2014), 1051–1067. DOI:https://doi.org/10.1177/1461444814543995.
[36] [36]Miaskiewicz, T. and Kozar, K.A. 2011. Personas and user-centered design: How can personas benefit product design processes? *Design Studies*. 32, 5 (Sep. 2011), 417–430. DOI:https://doi.org/10.1016/j.destud.2011.03.003.
[37] [37]Moon, P.F. *et al.* 2023. The tools being used to introduce youth to data science. *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (Chicago IL USA, Jun. 2023), 150–159.
[38] [38]National Academies of Sciences, Engineering, and Medicine 2023. *Foundations of Data Science for Students in Grades K-12: Proceedings of a Workshop*. National Academies of Sciences, Engineering, and Medicine.
[39] [39]Ng, W. 2012. Can we teach digital natives digital literacy? *Computers & Education*. 59, 3 (Nov. 2012), 1065–1078. DOI:https://doi.org/10.1016/j.compedu.2012.04.016.
[40] [40]Noble, S.U. 2020. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press.
[41] [41]O'Neil, C. 2016. Weapons of math destruction: How big data Increases inequality and threatens democracy. Crown.
[42] [42]Pangrazio, L. and Selwyn, N. 2019. 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*. 21, 2 (Feb. 2019), 419–437. DOI:https://doi.org/10.1177/1461444818799523.
[43] [43]Pangrazio, L. and Selwyn, N. 2021. Towards a school-based "critical data education." *Pedagogy, Culture & Society*. 29, 3 (Jun. 2021), 431–448. DOI:https://doi.org/10.1080/14681366.2020.1747527.
[44] [44]Panke 2019. Design thinking in education: Perspectives, opportunities and challenges. *Open Education Studies*. 1, 1 (2019), 281–306.
[45] [45]Pariser, E. 2011. The filter bubble: How the new personalized web Is changing what we read and how we think. Penguin Books.
[46] [46]Rivas, C. 2012. Coding qualitative data. *Researching Society and Culture*. C. Seale, ed. SAGE Publications. 367–392.
[47] [47]Rowley, J. 2007. The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*. 33, 2 (Apr. 2007), 163–180. DOI:https://doi.org/10.1177/0165551506070706.
[48] [48]Saldaña, J. 2016. *The coding manual for qualitative researchers*. SAGE.
[49] [49]Schanzer, E. *et al.* 2022. Integrated data science for secondary schools: Design and assessment of a curriculum. *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education* (Providence RI USA, Feb. 2022), 22–28.
[50] [50]Stein, J. 2011. Data mining: How companies now know everything about you. *Time*.
[51] [51]Stornaiuolo, A. 2020. Authoring data stories in a media makerspace: Adolescents developing critical data literacies. *Journal of the Learning Sciences*. 29, 1 (Jan. 2020), 81–103. DOI:https://doi.org/10.1080/10508406.2019.1689365.
[52] [52]Tifentale, A. and Manovich, L. 2015. Selfiecity: Exploring photography and self-fashioning in social media. *Postdigital Aesthetics: Art, Computation and Design*. D.M. Berry and M. Dieter, eds. Palgrave Macmillan UK. 109–122.
[53] [53]Weiland, T. and Engledowl, C. 2022. Transforming curriculum and building capacity in K–12 data science education. *Harvard Data Science Review*. 4, 4 (Oct. 2022). DOI:https://doi.org/10.1162/99608f92.7fea779a.
[54] [54]Wilkerson, M.H. and Polman, J.L. 2020. Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences*. 29, 1 (Jan. 2020), 1–10. DOI:https://doi.org/10.1080/10508406.2019.1705664.
[55] [55]Wise, A.F. 2022. Educating data scientists and data literate citizens for a new generation of data. *Situating Data Science: Exploring How Relationships to Data Shape Learning*. M.H. Wilkerson and J.L. Polman, eds. Routledge.
[56] [56]Wolff, A. *et al.* 2016. Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*. 12, 3 (Aug. 2016). DOI:https://doi.org/10.15353/joci.v12i3.3275.