IADIS International Journal on Computer Science and Information Systems
Vol. 18, No. 1, pp. 1-17
ISSN: 1646-3692

DIGITAL PARTICIPATION FOR DATA LITERATE CITIZENS – A QUALITATIVE ANALYSIS OF THE DESIGN OF MULTI-PROJECT CITIZEN SCIENCE PLATFORMS

Carolin Stein, Jonas Fegert, Alicia Wittmer and Christof Weinhardt
FZI Research Center for Information Technology, KIT Karlsruhe Institute of Technology,
Haid-und-Neu-Straße 10-14, 76131 Karlsruhe, Germany

ABSTRACT

In the digitized world, data literacy becomes a skillset that enables participation in transforming societies and job markets. This motivates the exploration of how creative teaching concepts might look like in order to form those relevant competencies. Citizen science, as a collaborative approach to scientific research, engages citizens in almost all forms in working with data that is meaningful to them. As such, this paper seeks to examine the potential of digital citizen science in enabling to practice and gain data literacy knowledge. To do so, this paper is focusing on the design of multi-project citizen science platforms through a structured artifact review. Furthermore, we gain insight through a qualitative interview study from 14 interviews (citizens, researchers, and technical experts), drawing conclusions on current platform designs as well as potentials and challenges for their further development. Our results show, on the one hand, that platforms provide a variety of ways to interact with different kinds of data while offering a simple and user-friendly design. On the other hand, functions for communication and support could be further developed and learning opportunities for citizens comprise several, but not all aspects of data literacy.

KEYWORDS

Digital Citizen Science, Platforms, Data Literacy

1. INTRODUCTION

In the digitized world, the ability to navigate in a data-based environment is a prerequisite for taking part in societal discussions and decision-making (Bhargava et al., 2015; Debruyne et al., 2021; D’Ignazio, 2022). The discourse on the Covid-19 pandemic, involving statistical metrics like the R-value, infection rates, and occupancy rates, exemplifies the importance of data literacy as a crucial skill, not only for professionals but also for the general public (Schüller et al., 2019; Debruyne et al., 2021). However, many citizens are missing basic data skills without the opportunities for continuing education in adulthood (Bhargava et al., 2016; Debruyne et al., 2021), motivating the exploration of creative strategies to involve them in data processes (D'Ignazio, 2022). While in several countries the public interest in science and research has increased during the pandemic (Jensen et al., 2021), the field of Citizen Science (CS), as a collaborative approach to undertaking scientific research, continues to gain increasing attention (Vohland et al., 2021). CS can involve non-professionals in different research activities, from the development of research questions to the collection or analysis of research data (Shirk et al., 2012). Citizens have contributed to scientific achievements in a variety of disciplines, such as collecting and sharing geophysical data for earth observation research, or engaging in collective problem-solving and symptom or treatment surveillance in biomedicine (Shirk and Bonney, 2018). While the academic discourse recognizes the ability of CS to generate unique insights and power up research workforce (Shirk and Bonney, 2018), it also discusses its potential to improve participants' knowledge and skills, such as scientific literacy or domain knowledge (Jennett et al., 2016; National Academies of Sciences and Medicine, 2018). In most forms, CS includes the engagement of citizens with data (National Academies of Sciences and Medicine, 2018; Bowser et al., 2020), which makes it an interesting use case in the debate on promoting data literacy. In the CS literature, it has been shown that learning opportunities depend on a multitude of project characteristics (National Academies of Sciences and Medicine, 2018), which complicates broad claims for learning and transformative effects within the variety of CS (De Albuquerque and Almeida, 2020; Bela et al., 2016). In this context, digital CS, also referred to as online or virtual CS, involving citizens by means of information and communication technology (ICT), is interesting for multiple reasons. On the one hand, it allows for large-scale participation and thus potentially the learning experience of a multitude of hobby citizen scientists (Jennett et al., 2016; Aristeidou and Herodotou, 2020). On the other hand, lately a shift from individual to generic infrastructure can be observed (Baudry et al. 2022), with some platforms evolving into integrated platforms for the conduction of generic projects (Liu et al., 2021). These multi-project platforms standardize the way digital CS projects are designed and conducted (Baudry et al., 2022; Liu et al., 2021), what in turn determines certain project outcomes such as democratic impact or learning (Bela et al., 2016; De Albuquerque and Almeida, 2020). As such, rather than evaluating opportunities for learning and applying data literacy in individual projects, the functionality and design of multi-project CS platforms can be an alternative starting point for exploring and shaping potentials and obstacles by design. Thus, we define our research question as follows:

RQ: *What are the potential and challenges for promoting citizens' data literacy through digital CS?*

- *RQ1: What multi-project CS platforms currently exist?*
- *RQ2: How do current multi-project CS platforms support the application and learning of data literacy in CS projects?*
- *RQ3: What potentials and challenges do researchers and citizens identify to further develop multi-project CS platforms?*

To answer our research questions, we draw on the conduction of two independent qualitative studies. Through a structured artifact review, we assess the status quo of the platform landscape, providing detailed insights into the functionality and design of 16 multi-project CS platforms. Additionally, we draw from a qualitative interview study with citizens, researchers, and technical experts that elaborates on their perspective on the utilization and design of multi-project CS. By combining the results of both studies, we are able to provide a comprehensive assessment of the potential and challenges for citizens to apply and learn data literacy on digital CS platforms from the angle of platform design.

2. RELATED WORK

In this chapter, the concepts of data literacy and (digital) CS are introduced and brought together to provide a theoretical basis for investigating data literacy promotion in digital CS.

2.1 Concepts

Data Literacy is a wide-ranging concept that is difficult to define and separate from other literacies such as digital or statistical literacy (Bhargava et al., 2015; Schüller et al., 2019; Gould, 2021). It can be described as an objective set of skills, such as the "ability to read, work with, analyze, and argue with data as part of a larger inquiry process" (D'Ignazio and Bhargava, 2016, p. 84), but also more generally as the individuals' empowerment to navigate and engage in their data-based environment and society (Bhargava et al., 2015; Schüller et al., 2019). Data, as collected or generated information used to infer about different phenomena (Wise, 2020), is situated in an ecosystem that includes organizational elements such as its producers and consumers or the infrastructure and tools used (Bhargava et al., 2015). Data literacy therefore encompasses different competencies (Pedersen and Caviglia, 2019; Schüller et al., 2019; Debruyne et al., 2021), that can be modeled around the pyramid model of data value creation. To guide the education of data literacy, Schüller et al. (2019) developed a framework that describes relevant skills allocated to six core competencies (establish a data culture, provision data, analyze data, interpret results, interpret data, and derive actions). Furthermore, multiple educators have defined best practices for teaching data literacy, emphasizing more than just skill acquisition (Ridsdale et al., 2015; Bhargava et al., 2016; D'Ignazio and Bhargava, 2016; Wise, 2020; Debruyne et al., 2021; D'Ignazio, 2022). They place value on contextualization, such as drawing attention to the relationship between data and its production context (Wise, 2020), working with real-world community data (Ridsdale et al., 2015; D'Ignazio, 2022), or empowering learners to apply skills in their own contexts (Bhargava et al., 2015). Learning in interdisciplinary problems (Pedersen and Caviglia, 2019) and end-to-end data processes is encouraged to explore the impact of inquiry goals or stakeholder interests (D'Ignazio, 2022). Additionally, effective learning tools should be focused, guided, and inviting (D'Ignazio and Bhargava, 2016). Given its complexity, data literacy cannot be taught in a single initiative (Debruyne et al., 2021), so a modular approach should be chosen (Ridsdale et al., 2015).

Citizen Science describes an umbrella term for civic engagement in research that is yet lacking a uniform definition (Haklay et al., 2021). The Socientize project characterizes CS as "the general public engagement in scientific research activities when citizens actively contribute to science either with their intellectual effort or surrounding knowledge or with their tools and resources" (Socientize, 2014). While CS has been used originally, especially in natural science, nowadays it is a useful means to conduct research in various research domains (Levy and Germonprez, 2017; Pettibone et al., 2017). With the expansion of CS into a variety of fields of application, the heterogeneity of approaches has increased (Spasiano et al., 2021). The potential involvement can cover different parts of the research process, from the creation of the research question and hypotheses through data collection to analysis and publication (Shirk et al., 2012).

This participation can be organized through CS platforms, describing web-based infrastructures that are used to support CS initiatives (Liu et al., 2021). Platforms can comprise a variety of infrastructures that either present CS activities, display project information, provide material, offer tools, or a combination of the aforementioned (Liu et al., 2021). The emergence of multi-project CS platforms offers project initiators the opportunity to set up and conduct their project with its distinct goals while sharing digital infrastructure with other projects instead of developing their own (Baudry et al., 2022). They thus standardize functions, as well as aspects such as community management (Baudry et al., 2022), which makes their design relevant for achieving results such as learning outcomes (Bela et al., 2016; National Academies of Sciences and Medicine, 2018; De Albuquerque and Almeida, 2020).

2.2 Digital Citizen Science for Data Literacy Promotion

Educational initiatives can be assessed within four levels of the Kirkpatrick Model, being the learner's response to the initiative, the learning, the performance, and the result (Horton, 2001). While the first level investigates whether a learner generally enjoys participation, the second and third levels focus on the knowledge and skills acquired that can be applied in the learning setting or transferred to real life problems (Horton, 2001). The fourth level underlines the focus on the initial goals of the learning, which can be a certain business goal or an impact on the participants' attitudes (Horton, 2001; Schüller et al., 2019). In the case of promoting citizens' data literacy, it is therefore decisive whether participants enjoy their participation in a CS project and what data knowledge and skills they gain in participating in the project, which is usually not a learning setting but a real-life problem. Likewise, it should be assessed whether the initiative empowers them to navigate their own data-based environment, which can be seen as an overarching goal of data literacy for citizens (Bhargava et al., 2015; Schüller et al., 2019). In CS, a considerable corpus of literature focuses on evaluating learning or learning opportunities in offline or online projects: They often assess general motivations (Jennett et al., 2016; Phillips et al., 2018); however, instead of focusing on data literacy, mostly domain-specific learning and digital or scientific literacy are evaluated (Jennett et al., 2016; National Academies of Sciences and Medicine, 2018; Phillips et al., 2018; Aristeidou and Herodotou, 2020; Herodotou et al., 2021). Although these competencies overlap with data literacy (Bhargava et al., 2015), only some research explicitly focuses on the acquisition of data skills and potential opportunities (Radchenko and Maksimenkova, 2016; Bowser et al., 2020; Golumbic et al., 2020). Other literature investigates the projects resulting effects and the empowerment of citizens (Haklay, 2013; Bela et al., 2016; Phillips et al., 2018; De Albuquerque and Almeida, 2020). Overall educational opportunities seem to be project-specific, depending on the content and design of the projects (National Academies of Sciences and Medicine, 2018; Aristeidou and Herodotou, 2020). A different angle for analyzing CS projects can be the analysis of underlying digital infrastructure. Many best practices have been developed for CS platforms to optimize their design for participation (Newman et al., 2010; Jennett and Cox 2016; Wald et al., 2016; Yadav and Darlington, 2016; Sturm et al., 2018; Skarlatidou et al., 2019; Musto and Dahanayake, 2021). Linking these principles to their relevance for learning and data literacy education allows us to assess platform designs according to their potential to promote citizens' data literacy. An overview of the identified design principles can be found in Table 1.

Table 1. Key design dimensions identified in the digital CS literature

	Platform Design Concept	Meaning for Data Literacy
Aesthetics and Usability (AU)	Provide simple and clear project main pages (Skarlatidou et al., 2019)	Provide an enjoyable education initiative (Horton, 2001) Provide a focused and inviting tool (D'Ignazio and Bhargava, 2016)
	Ease entry barriers (Jennett and Cox, 2016; Sturm et al., 2018)	
	Reduce information on the platform (Jennett and Cox, 2016; Skarlatidou et al., 2019)	
	Standardize naming and navigation (Sturm et al., 2018; Skarlatidou et al., 2019)	
	Communicate project goals (Newman et al., 2010; Jennett and Cox, 2016)	Create understanding for inquiry goals and data production context (Wise, 2020; D'Ignazio, 2022)
Data Standards (DS)	Validate user-generated data (Skarlatidou et al., 2019; Musto and Dahanayake, 2021)	Enable informal learning through the execution of tasks (Jennett et al., 2016) Empower learners to practically apply their skills to real-world data (Bhargava et al., 2015; Ridsdale et al., 2015; D'Ignazio, 2022)
	Facilitate entering user-generated data (Newman et al., 2010; Sturm et al., 2018; Skarlatidou et al., 2019)	
	Enable data analysis and visualization (Newman et al., 2010; Wald et al., 2016; Skarlatidou et al., 2019; Musto and Dahanayake, 2021)	
Support (SU)	Provide separate support pages (Skarlatidou et al., 2019)	Enable formal learning (Jennett et al., 2016) Provide a guided tool (D'Ignazio and Bhargava, 2016)
	Provide educational material (Wald et al., 2016)	
	Provide interactive tutorials and information (Jennett and Cox, 2016; Skarlatidou et al., 2019)	
Communi cation (CO)	Enable communication between participants (Newman et al., 2010; Wald et al., 2016; Sturm et al., 2018; Skarlatidou et al., 2019)	Enable informal learning through interaction (Jennett et al., 2016) Enable two-sided learning for an empowering effect (De Albuquerque and Almeida, 2020)
	Enable communication between participants and researchers (Newman et al., 2010; Yadav and Darlington, 2016; Sturm et al., 2018; Skarlatidou et al., 2019)	

3. METHODOLOGY

Our research approach comprised the conduction of two independent qualitative studies with the aim of shedding light on both present and future potentials for promoting citizens' data literacy in digital CS from the perspective of platform design. First, we conducted a structured artifact review to identify and assess current multi-project CS platforms according to their functionality and key dimensions identified in the literature. For the review, we followed a 7-step methodology for reviewing real-world software artifacts that aims to provide methodological guidance to the review in a systematic way, including problem formulation, artifact search and screening, assessment of practical quality, data extraction, artifact documentation and archiving, and data analysis (Gnewuch and Maedche, 2022). While steps 1-6 have been conducted by one reviewer, for step 7, two independent reviewers have been engaged in coding the documentation. For data extraction and coding, we followed best practices for structured-content analysis of web pages (Saraswat, 1999). Second, based on this depiction of the status quo, we aim to contextualize results with the interests and needs of CS stakeholders and uncover potential further opportunities or threats to promoting citizens' data literacy in digital CS. As such, we draw from the conduct of semi-structured interviews with citizens, researchers, and technical experts. For the interview preparation, conduct, transcription, and structured content analysis, we followed Kaiser‘s method for qualitative interviews (Kaiser, 2014). This included the design of an interview guideline and the conduct of a pretest, which was used to test and refine the interview guideline. The transcription and coding were undertaken by a single researcher with the tool MAXQDA, while in the analysis of the results two researchers were involved.

4. A REVIEW OF CS PLATFORMS

In this chapter, we present the implementation and results of the structured artifact review.

4.1 Implementation

The structured artifact review was implemented from July to September 2022. It aims to identify the current functionality and design of digital platforms to derive their potential and challenges for promoting citizens' data literacy. Especially interesting to us are multi-project CS platforms, as they standardize opportunities for a multitude of projects (Baudry et al., 2022). Therefore, we defined active participation opportunities for citizens (a) and the possibility for generic project creation (b) as inclusion criteria to exclude other platform types (Brenton et al., 2018). In terms of assessing practical usability for the review, we defined as criteria availability in either English or German (c) and the possibility to review them free of charge (d), meaning either the platform must be free of charge or it must offer a demo or link to project examples that can be reviewed. For the software artifact search, we utilized three search directions being: overviews in the CS platform literature (Liu et al., 2021; Brenton et al., 2018; Luna et al., 2018; Yadav and Darlington, 2016; Aristeidou and Herodotou, 2020; Skarlatidou et al., 2019) (43 artifacts); the commercial database provider *Crunchbase* with the filter option 'Citizen Science' (14 additional artifacts); and the EU's and Austria's CS information webpages (18 additional artifacts). The initial sample of 75 artifacts included next to multi-project platforms, 32 project overview platforms, five community exchange hubs for educational material and workshops, and 12 single-project platforms. Additionally, six platforms presented several projects but did not allow for the creation of new generic initiatives. Thus, after screening for (a) and (b), 20 multi-project platforms were further reviewed for quality. Excluding four platforms due to language barriers or accessibility issues, 16 platforms for the final review were archived in *Archive.Today* and can be seen in Table 2. For data extraction and analysis, the platforms were assessed according to their participatory functionality based on the concept degree of participation by Shirk et al. (2012) and their design, based on the criteria in Table 1, and documented in a concept matrix.

Table 2. List of multi-project platforms for the review

Platform ID and Name

A	Biocollect-Atlas of Living Australia	I	Ispot
B	CS Center Zürich	J	Just One Giant Lab (JOGL)
C	CitSci	K	nQuire
D	conserve.Io*	L	Pybossa
E	CyberTracker	M	SciStarter
F	DataCertus	N	Spotteron
G	Epicollect5	O	World Community Grid
H	Inaturalist	P	Zooniverse

* The review of this platform is based on two freely viewable projects and thus limited.

4.2 Results

In the following, the assessment results of the 16 multi-project CS platforms are reported in detail by dimension. Additionally, a brief overview of the results can be found in Table 3.

Table 3. Summary of review results

Dim.	**Platforms**															
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**
Participatory Functions (PF): computational power (PF1) data collection (PF2), data analysis (PF3), task assignment (PF4) research questions and approach (PF5)																
PF	PF 2,3	PF 2,3	PF 2,3	PF 2	PF 2	PF 2,3	PF 2	PF 2	PF 2	PF 4	PF 5	PF 2,3	PF 2,3	PF 2,3	PF 1	PF 2,3
Aesthetics & Usability (AU): similar project main pages (PA1) eased entry barriers (PA2) reduced information (PA3) easy naming conventions (PA4) search options (PA5) project goal communication on project page (PA6)																
PA	AU 1-6	AU 1-6	AU 1-6	AU 2,3 4,6	AU 3-5	AU 1, 3-5	AU 1, 3-6	AU 1-6	AU 1-5	AU 1-6	AU 1-3 5,6	AU 1,3 4,6	AU 1-6	AU 2-6	AU 1,2 4-6	AU 1-6
Data Standards (DS): eased data entering (DS1) data analysis and visualization (DS2) data validation (DS3)																
DS	DS 1,2	DS 1,2	DS 1,2	DS 1,2	DS 1,2	DS 1,2	DS 1,2	DS 1-3	DS 1-3	-	DS 1,2	DS 1,2	-	DS 1-3	DS 2	DS 1-3
Support (SU): support page (SI1) educational material (SI2) interactive information and tutorials (SI3)																
SI	SI 1-3	SI 3	SI 1-3	SI 1,2	SI 2	-	-	SI 1-3	SI 1,2	SI 1-3	SI 3	SI 1	SI 1-3	SI 1,3	SI 1-3	SI 1-3
Communication (CO): communication between participants (CO1), communication between participants and researchers (CO2)																
CO	CO 2	CO 2	CO 1,2	-	-	CO 1,2	CO 1,2	CO 1,2	CO 1,2	CO 1,2	CO 1,2	-	CO 1,2	CO 1,2	CO 1,2	CO 1,2

Platforms that meet all criteria in one dimension are shaded

1) Participatory functions: To evaluate what aspects of data literacy citizens may apply and improve through their participation in digital CS platforms, projects are reviewed for their offer of participation activities. The lowest degree of participation was found for platform O, where citizens can only contribute to research projects through computational power. On the other hand, many platforms, focus on data collection and analysis: Four platforms (D, E, H, I) provide functionality to upload and classify observations, while one platform (F) allows for the upload of complete data sets. Other platforms extend the possibilities for data collection, e.g., to surveys (B, G, P), digital diaries (B), or different types of media (B, G, L). The opportunities for participatory data analysis range from functionalities for transcriptions (B, L, P), mapping or classification of images (B, L, P) to pattern recognition in sound or video material (L). Additionally, one platform offers no code toolboxes for machine learning algorithms (F). Four data collection and analysis platforms (A, C, M, N) stood out through their flexibility, which would potentially allow their usage for other participatory activities. This includes one platform (M) offering a range of third-party tool integrations, another platform (C) with customizable data sheets and possible integrations, a further platform (A) providing multiple individual applications, e.g., for evaluation or learning games, and a platform (N) with functionalities that can be customized on demand. Next to platforms for data collection and analysis, one platform (J) in the sample supports the participatory assignment of tasks, although tasks themselves are not conducted on the platform. This platform enables initiators to structure the project into several phases, including phases such as project conception or prototyping, with the offer of partly challenging tasks. Another platform (K) enables users to define research questions, design a methodological approach, and collect data; however, except for data collection, activities are not undertaken participatory.

2) Aesthetics and usability: The platforms' aesthetics and usability play an important role in who can access and enjoy participation in CS projects (Skarlatidou et al., 2019), as an important prerequisite to a learning initiative. As such, we reviewed five subdimensions identified in the literature. First of all, we reviewed whether platforms have a consistent concept for their project main pages. Most platforms follow the principle of deploying the same design for every project. Only two platforms (D, N) vary their designs throughout different projects, and one platform (E) has no project overview page by design. Five projects (B, H, N, O, P) give an overview of the projects' progress on the main page, while some platforms include information on contributing participants (B, C, E, H, J, M, N, P) or contributed data (A, C, D, E, G, H, I, K, L, N, P). Additionally, outstanding features that were noted were an overview of currently active participants in the project (I, N, P) or the proposition of similar projects (M). Second, concepts for easing entry barriers to platform usage were reviewed. For this subdimension, some platforms enable participation via the web browser (A, B, C, H, I, J, K, M, P), while others require the installation of software. Above that, three platforms (D, J, N) include either pop-up explanations upon registration or entering a project, and ten platforms provide example pictures or explanations for project tasks (A, B, D, I, J, K, M, N, O, P). Additionally, four platforms (A, J, M, P) indicate the level of difficulty for tasks or indicate the required skills for participating. As a third and fourth subdimension, concepts for reducing text and standardizing naming and navigation conventions were evaluated. To reduce the information load, all platforms utilize pictures or icons, except for platform O, which is primarily text-based. In particular, one platform (E) stood out by replacing texts comprehensively through icons. Additionally, seven platforms (A, H, J, L, M, N, P) use expandable and collapsible text to avoid information overload at first sight. In terms of naming conventions, most platforms try to avoid technical terms and follow easy naming conventions such as 'Project or Community', 'Add' or 'Contribute'. An exception to this was found in the platform J, which uses individual names e.g. 'Spaces', 'Needs', 'Programs' or 'Claps' and technical terms, for instance, in search options. For the platform's navigation, three platforms (B, C, F) allow searching for specific project names and one (E) for a specific web address. Besides that, three platforms enable users to search for names or categories (G, O, P) and three (D, K, L) platforms do not employ any search options. The majority of platforms (A, H, I, J, M, N) however, allow for a broad search according to different characteristics such as age, activity, or organization. As creating an understanding of inquiry goals is important for teaching data literacy, we lastly evaluated the communication of project goals, identifying two different approaches. Three platforms (E, F, I) include general goals on the platform's main homepage, while other platforms include an explanation of project-specific goals on the individual project pages.

3) Data standards: In terms of working with scientific data, platforms have been reviewed for their concepts to support entering, validating, visualizing, or analyzing data, as mentioned in the CS literature. For collecting project data, most platforms facilitate participation through specified entry masks that include, for example, drop-down menus, checkboxes, or drag-and-drop (A, B, C, D, E, F, G, H, I, K, L, N, P). Additionally, some platforms provide extra help for the classification of data (E, H, I, N, P) in form of common mistakes, help information, or automated proposals. Two platforms do not support data contributions (J, O) and one platform does not have standardized data entry options (M). In terms of data validation, only four platforms present validation concepts: This comprised checking whether all necessary entries were filled (P), flagging contributions through the community (N), a like and reputation system (I), or the differentiation between quality levels (H). For platform D, data validation could not be reviewed. Regarding the analysis of project data, some platforms have integrated data analysis tools (A, B, C, E, F, K, N, O, P), while others focus only on data visualization (D, G, H, I). Analysis tools include data tables (P), sampling options (A), statistics (C, N, P, M), or predictions (A, N). Furthermore, shared codebases for researchers (C, P), machine learning classifications (F), options to compare specific data points or groups (A, C, N), or to follow the total numbers of classifications or solved tasks (C, N, P, B, O), are included in platforms. In terms of visualizations, platforms offer scatter plots (A, N, P), histograms (C, N, P, D, E, G, F, K, M), map visualization (A, C, N, P, D, E, G, F, H, I, M) or plots visualizing data trends over time (C, N). One platform (L) allows for flexible analysis through code integrations without predefining tools automatically, and two platforms (J, M) do not enable data analysis. Besides project data, five projects allow the analysis of data about the projects' progress (B, H, N, O, P), and five projects analyze meta-data about the community (L, M, N, O, P).

4) Support: Next to informal learning, digital platforms were reviewed for three aspects of supporting formal learning. The first one comprised a concept for a separate support page. Most platforms have help pages for several topics, such as guidelines on how to build a project (A, B, C, E, F, G, H, I, J, L, M, N, P), or general support pages for citizens (A, D, H, I, J, L, N, O). Additionally, most platforms employ an FAQ page for citizens (A, C, H, I, J, L, M, N, O, P). Five platforms (B, E, F, G, K) do not provide any separate help pages for participants. As a second aspect, platforms were reviewed for their provision of educational material. Implementations of this feature comprised workshops (A, M, P), training sections (A, C, D, H, I, M, P), blog or newsletter articles (C, E, H, I, O, P), and community spaces for learning (J). Six platforms did not integrate educational resources by design, however, project-based integrations are still possible. Third, concepts for interactive tutorials or help information were reviewed. Design features that were employed in this category were pop-up tutorials leading through the platform (J, N, P), tutorials and videos on project pages (H, J, M, N, O, P), or additional information based on hovering over content (A, B, C, K, N, O). Six platforms do not include interactive help information.

5) Communication: As a central aspect of empowerment and learning through interaction, concepts for communication between participants, and between participants and researchers were reviewed. Five platforms (A, B, D, E, L) have no concept for a project-based debate of participants, while three platforms (D, E, L) additionally have no possibilities to contact researchers. Five platforms (H, J, M, N, P) enable exchange within the community and with researchers via direct chat options. Nine platforms (C, G, H, I, J, K, N, O, P), allow for communication via forums, either project-based or general forums. Other concepts for community exchange comprised messages on newsfeeds (J, N), comment options (H, I, J, K, M, N, P) or project-based question and answer options (M). For contact with researchers or project creators additionally, many projects (A, B, C, F, G, K) offered an e-mail option. One platform in the sample (N) especially stood out due to its broad range of options for community exchange.

5. INTERVIEWS ON POTENTIALS OF CS PLATFORMS

In the following chapter, we report on the conduct of a qualitative interview study and a subset of its results, which are used to investigate the research object. The interviews aimed to identify potential and challenges for engaging citizens in scientific research projects via digital platforms and were conducted earlier and independently of the artifact analysis. By revising aspects of the study relevant to our findings from the structured artifact review and the context of applying and learning data literacy, we can contextualize our findings about the current platform landscape and provide guidance for its further development.

5.1 Implementation

The interviews were implemented from April to June 2021 using a virtual meeting tool. They included 14 participants, who were between 16 and 64 years old and composed of six citizens (B1-B6) and eight professional researchers (F1-F8) out of whom two can be considered technical experts for platform design (F7, F8). We ensured equal gender representation and diversity in terms of prior knowledge regarding the participation in or realization of CS projects. The interviewed researchers had diverse backgrounds, such as medicine, sports science, or information systems, to cover different potential applications of CS. After a brief introduction to the topic of digital CS, the interviewees were consulted for their perspectives on the offer of different participatory activities according to Shirk et al. (2012), support needs to conduct these activities on a CS platform, challenges in working with scientific data, especially with regard to data quality according to Parsons and Lukyanenko (2011), and needs for communication and feedback. Additionally, the interview guideline included the topics of incentives, quality of participation and aspects of anonymity, which are outside the scope of this paper.

5.2 Results

The results of the structured content analysis are reported by dimension, with the respective interview source indicated in brackets (Citizens C1-C6, and Researchers (R1-R8) including two technical experts (R7, R8)).

A) Participatory Functions: To determine potentials for participation and engagement with scientific data, we asked the interviewees whether they could imagine themselves or citizens participating in *asking research questions* (A1), *collecting information and resources* (A2), *developing hypotheses* (A3), *developing research methodologies* (A4), *collecting data* (A5), *analyzing and interpreting data* (A6), and *publishing results* (A7) via a digital platform (Shirk et al., 2012). For every research step, at least half of the participants expressed interest in participating. While citizens could imagine their participation best in the step of *developing hypotheses* (n=6), more challenging to them appeared the steps of *data analysis and* *interpretation* as well as *publishing results* (n=3). Researchers could imagine citizens' participation best in *developing research questions* and *data collection* (n=8) and the least in *developing hypotheses* and *publishing results* (n=4). Obstacles on the side of participants were entitlement (C4, (A1)), adhering to formal structures and wording (C5, (A7)), or motivation and time (C1, C3, (A2); C1, C6 (A4); C5 (A5); C3 (A7)). Additionally, they claimed that they might miss important knowledge (C6, (A1); C4 (A4); C4 (A7)), or feared contributing incomplete or wrong contributions (C3, (A5); C5 (A6)). For researchers, major challenges included the utilization of domain literature (R4 (A1)), containing accessing, evaluating, and reading scientific contributions (R1, R2, R4, R8 (A2)). Additionally, they remarked on the challenge of adhering to formal structures and logical reasoning, thereby ensuring quality (R4 (A1); R2, R5, R6 (A3)), and of missing knowledge necessary for the conduct of tasks (R8 (A1); R4, R7 (A4); R5, R6, R8 (A6)). One researcher added the challenge of coordinating work while being exposed to deadlines (R8 (A7)). On the positive side, participants remarked that enabling participation in more challenging parts of the research process could be interesting and valuable for citizens and researchers (C1, C2, R4, R7 (A6)), generating potentially unconventional and interesting results (R5, R8 (A4), R2, R4, R5 (A6)), while being a great source to impart knowledge (R1, R8 (A4)) and give credit (R5, R8 (7)).

B) Data Tasks: In terms of working with scientific data, the interviewees were asked about their perception of challenges, especially concerning data quality. Researchers and citizens equally expressed concerns about the quality of data regarding intentional or unintentional manipulations by participants. Citizens could try to adapt their contributions to fit socially accepted perceptions (R2), their own ideas and expectations (C1, C3, C4), or to build up certain political topics (C4, R6). They could also lose interest in working conscientiously over time (R2, R4). The lack of control over whether participants adhere to scientific standards (R4, R5, R8) and the potential incompleteness of contributions (C2) are named as challenges for citizens contributing scientific data. In this context, two researchers noted that challenges might be dependent on the type of data and measurement (R1, R6). Additionally, quality issues might also comprise the scientific relevance of contributed ideas or data, such as generated research questions (R1, R6).

C) Support: In A) and B), citizens were additionally asked for needed support to overcome obstacles in participation and data engagement. Participants repeatedly stressed the importance of collaboration and exchange. This could include a direct point of contact in case of questions (C5, R3, R7), the organization of workshops (R1, R5, R6), or functionality for a joint development of hypotheses or analyses both between participants and researchers but also the community itself (C1, C3, C5, C6, R4-R8). Furthermore, feedback was frequently mentioned as an important factor (C1, C3, C5, C6, R1, R3-R7) which could be linked to the possibility of reminding participants (R2). When it comes to giving feedback, the participants suggested that it could be either manual or automated (R1, R3) or elaborated through a collaborative review process (R3, R7); however, it should be personal (C3, R5) and transparent (R7). In order to support formal learning, participants ideated that platforms might include tutorial videos (C2-6, R1, R2, R4, R5, R8), short manuals and checklists (C1, C3, C4, R5), or best practices (F7). Materials could be collected in a database (R3, R8), potentially with a recommender system (C4, R3, R8) or a translator, thereby helping with complex scientific language (R8). Additionally, respective knowledge could be transferred by introducing a toolbox to explain, e.g., scientific methods and what they are for (C4, R1, R4) or providing templates (C2, R7). Participants also ideated on the use of chatbots for answering questions (R5, R7) or supporting certain research steps (R1, R3). To help with quality concerns, platforms could introduce control questions (R4), enable comparison with other data (C2, R8), or introduce the need to upload explanations and supporting documents for contributions (C1, C5). However, ethical education might also be important, such as stressing the importance of the validity of contributions and the neutrality of science (C1, C4, R1).

D) Communication: For the communication, participants were asked about several communication mechanisms (voting, discussion forum, comment option, chat function) and their suitability for using them on digital CS platforms. Overall, all participants showed great support for all different communication mechanisms, underscoring that communication means are indispensable to enable meaningful participation (C2, R4) and that it would be helpful to have everything integrated into a single platform (R5). One citizen highlighted that offering a variety of different communicative means would be important as participants have different preferences (C3), and two interviewees added that it would give important flexibility to the participants to contribute their opinion (C1, R8). The highest endorsement was given to the discussion forum, up-and-down voting, and chat functions. The up and down voting would allow for quick participation with minimal effort (C2, C5, R4, R5) that would help to structure content and enable easy overviews of the opinions and interests of citizens (R2, R4). However, it could also lead to polarization, group building, or the suppression of certain views (R6, R8). A discussion forum was seen as an important means to facilitate collaboration, and the collective creation of innovative ideas (R5, R8), while a direct chat function would improve the building of a community and mutual, quick exchange (R1, R2, R8). Additional participants mentioned video conferences (C1, C3, R4), and one participant noted that audio comments could enable freer participation (C1).

6. DISCUSSION

Through the conduct of two qualitative studies, we aimed at exploring the potential and challenges of digital CS to enable citizens to use and extend their data literacy. While the review of 16 multi-project CS platforms enables us to draw an extensive picture of the current digital CS landscape, the conduct of 14 expert interviews helps us contextualize these findings and elaborate on future directions for multi-project CS platforms. As such, we arrive at three main positions:

A) *Current digital CS projects provide opportunities to engage citizens in collecting and analyzing different types of data while largely neglecting other competence fields in data literacy, although they could potentially be more integrated.*

Our review found that all but one platform offered participatory tasks where citizens could engage with data, with a focus on data collection and analysis. For these research steps, most platforms offered extensive functionality to facilitate data contribution or classification and to allow citizens to access some form of data analysis or visualization. Only mechanisms to automate data verification were not common across the platforms, which our interviews suggested could be a challenge given that feedback and quality control are seen as important factors in improving citizens' ability to contribute qualitative data. The review showed that engagement with different types and formats of data is encouraged in many projects and is not limited to research data but also data collected on the projects and its participants. Such engagement can lead to informal learning (Jennett et al., 2016) of data literacy, including aspects of data measurements, variability, and visualization (National Academies of Sciences and Medicine, 2018; Kermish-Allen et al., 2019) or interpreting data products to derive recommendations (Golumbic et al., 2020). Data literacy, however, also encompasses skills such as identifying data use cases, measurable objects, and hypothesizing about their relationships (Schüller et al., 2019), which might not all be covered by contributory CS projects. To this end, platforms with community inquiry approaches might be promising (Herodotou et al., 2021), of which we could only find one example in our review. Likewise, no platform could be found that formally engaged citizens in final research steps such as interpreting and disseminating the results of the inquiry, which would allow for the application of data literacy aspects such as verbalizing data and data products or identifying data-driven actions (Schüller et al., 2019). From our qualitative interview study, both citizens and researchers believe that collaboration is feasible and interesting in other research activities as well. In particular, for developing research questions, hypotheses, and research methods, citizens showed great interest in collaborating with researchers. This could be an interesting starting point to expand opportunities for citizens to also apply and learn important aspects of data culture competencies, aspects of which were frequently mentioned in our interviews as challenges to participation and data quality.

B) Current digital CS platforms are advanced in their aesthetic and user-friendly design while providing diversity for different target groups.

The review suggests that, while varying in their design most platforms provide concepts for different aspects of usability and aesthetics. This can be seen as an important precondition for providing a pleasant learning opportunity and be open to a large group of participants (Skarlatidou et al., 2019). The review showed, that platforms for more literate citizens exist, while others enable working with data almost without any text. This is especially interesting as creative and art-based approaches to conveying data literacy more inclusively are stressed in the literature (Bhargava et al., 2016; D'Ignazio, 2022). Information on the projects' goals is present on almost all platforms' projects' main pages. To further strengthen a positive influence on citizens' data literacy, platforms could additionally think about making them more present during the execution of tasks and highlighting their influence on the data production context or resulting requirements the task's fulfillment.

C) Current digital CS platforms provide limited support and communication mechanisms that could be enriched with a variety of innovative functionality.

The conducted review revealed that current multi-project CS platforms severely vary in opportunities for support and communication on the platform. Some platforms even present no concept for either support or the communication especially, between participants. This is concerning, as contributory projects can be instrumental, postponing citizens to being data providers incapable of influencing the ways their data is used (De Albuquerque and Almeida, 2020), and studies found that there can exist a large fraction of participants contributing to digital initiatives while not feeling empowered through them (Haklay, 2013). The close relationship between communicative means and learning opportunities was stressed in the interviews, where the participants frequently mentioned collaboration and feedback as important support aspects. Although there are notable platform examples presenting several support and communication opportunities, the interviews suggested that there is room for improvement to add innovative concepts that are already discussed in other educational settings. This includes, for example, collaboration workflows (Zagalsky et al., 2015), recommender systems (Deschênes, 2020), or chatbots (Pérez et al., 2020; Okonkwo and Ade-Ibijola, 2021).

As concerns about poor data quality are one of the key obstacles for policymakers and scientists to support citizen science (Bowser et al., 2020), investing more in functionality to support data literacy education would not only be beneficial for citizens but also for researchers themselves.

7. CONCLUSION

In this work, we presented insights from a structured artifact review and a qualitative interview study investigating what potentials and challenges arise for applying and extending data literacy through digital CS projects. Grounded in the literature of CS and data literacy, we identified and analyzed 16 multi-project CS platforms according to their functionality and four design dimensions and contrasted the current conditions with the perspectives of 14 citizens and researchers. By conducting our assessment from the angle of platform design rather than individual projects, we have provided a new view on educational opportunities in digital CS, deriving three main positions for data literacy. Additionally, the two studies presented in this paper can be used in multiple ways by researchers and practitioners. As a theoretical contribution, through conducting a structured artifact review, our work allows IS researchers to compare CS platforms and identify structural gaps in terms of research or technology development for digital CS (Gnewuch and Maedche, 2022). From a practical point of view, CS project initiators and educators can use our work for guidance on choosing a platform that is most suitable for their needs. Although different classifications for CS platforms already exist (Brenton et al., 2018; Liu et al., 2021), navigating the large number of digital platforms and determining whether there is an appropriate solution for their project is a challenge to practitioners (Brenton et al., 2018). Additionally, the insights gained through the qualitative study can guide future developments by discussing and contrasting the needs and considerations of both citizens and researchers.

When utilizing the results of our work, the qualitative nature of the studies must be noted, which brings limitations in terms of the generalization of the research results. First of all, the structured artifact review is subject to natural limits in the search process as well as the data extraction undertaken by only one reviewer, which implies that some platforms or functionality or characteristics of the reviewed platforms might not have been detected. Additionally, the platform landscape is constantly evolving and progressing (Brenton et al., 2018; Liu et al., 2021), which implies that the review can only serve as a momentary snapshot that should be enriched in the future with additional platforms or review dimensions. As a next step, additionally identified platforms characterized as community hubs might be interesting to investigate. Although they do not enable active participation, they might hold great potential for data literacy education. Secondly, the qualitative interview study that was used to contextualize findings and give ideas for potential future developments, although including citizens and researchers of different ages, professions and pre-knowledge, did not strive for a representative sample. As the underrepresentation of certain cultural groups in CS projects is seen as a weakness and challenge (National Academies of Sciences and Medicine, 2018; Sorensen et al., 2019), it could be valuable to extend the interview study to include a larger cultural variety of citizens within a quantitative research setting. Finally, the studies only investigate the potentials, opportunities, and challenges of data literacy learning rather than actual effects. It is likely that not all learning opportunities are equally effective, nor will they be used equally by all participants. As such, future research is needed to quantify the actual effects on learning and empowerment on the platforms based on different competencies and opportunities to promote them. By providing a starting point for this research, we hope to initiate a more profound discussion on the design of CS platforms and related opportunities as well as challenges for promoting citizens' data literacy.

REFERENCES

Aristeidou, M. and Herodotou, C. (2020). Online Citizen Science: A Systematic Review of Effects on Learning and Scientific Literacy, *Citizen Science: Theory and Practice*, Vol. 5, No. 1, pp. 1–12.

Baudry, J. et al. (2022). Turning crowds into communities: The collectives of online citizen science, *Social Studies of Science*, Vol. 52, No. 3, pp. 399–424.

Bela, G. et al. (2016). Learning and the transformative potential of citizen science. *Conservation Biology*, Vol. 30, No. 5, pp. 990–999.

Bhargava, R. et al. (2015). *Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data*, *Data-Pop Alliance*. Available at: https://datapopalliance.org/item/ beyond-data-literacy-reinventing-community-engagement-and-empowerment-in-the-age-of-data/ (Accessed: 14 May 2023).

Bhargava, R. et al. (2016). Data Murals: Using the Arts to Build Data Literacy. *The Journal of Community Informatics*, Vol. 12, No. 3.

Bowser, A. et al. (2020). Still in Need of Norms: The State of the Data in Citizen Science, *Citizen Science: Theory and Practice*, Vol. 5, No. 1, pp. 18.

Brenton, P. et al. (2018). Technology Infrastructure for Citizen Science. *Citizen Science: Innovation in Open Science, Society and Policy*. London, England, pp. 63-80.

De Albuquerque, J., and Almeida, A. (2020). Modes of Engagement: Reframing "Sensing" and Data Generation in Citizen Science for Empowering Relationships. *Toxic truths: Environmental justice and citizen science in a post-truth age.* Manchester, England, pp. 267-281.

Debruyne, C. et al. (2021). DALIDA: Data Literacy Discussion Workshops for Adults, *ACM International Conference Proceeding Series*, pp. 23–25.

Deschênes, M. (2020). Recommender systems to support learners - Agency in a Learning Context: a systematic review. *International Journal of Educational Technology in Higher Education*, Vol. 17, No.1, p. 50.

D'Ignazio, C. (2022). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, Vol. 23, No. 1, pp. 6–18.

D'Ignazio, C. and Bhargava, R. (2016). DataBasic: Design Principles, Tools and Activities for Data Literacy Learners. *The Journal of Community Informatics*, Vol.12, No.3, pp. 83-107.

Gnewuch, U., and Maedche, A. (2022). Toward a Method for Reviewing Software Artifacts from Practice. *International Conference on Design Science Research in Information Systems and Technology*. St. Petersburg, USA, pp. 337-350.

Golumbic, Y.N. et al. (2020). Science literacy in action: understanding scientific data presented in a citizen science platform by non-expert adults. *International Journal of Science Education, Part B*, Vol. 10 No. 3, pp. 232–247.

Gould, R. (2021). Toward data-scientific thinking. *Teaching Statistics*, Vol. 43, pp. 11–22.

Haklay, M. (2013). Neogeography and the delusion of democratisation. *Environment and Planning A*, Vol. 45, No. 1, pp. 55–69.

Haklay, M. et al. (2021). What Is Citizen Science? The Challenges of Definition. *The Science of Citizen Science*. Cham, Switzerland. pp. 13–33.

Herodotou, C. et al. (2021). Methods of Promoting Learning and Data Quality in Citizen and Community Science. *Frontiers in Climate*, Vol. 3.

Horton, W. (2001). *Evaluating E-learning*. American Society for Training and Development, Alexandria, USA.

Jennett, C., and Cox, A. (2016). Eight guidelines for designing virtual citizen science projects. *Second AAAI Conference on Human Computation and Crowdsourcing*. Pittsburgh, USA, pp. 16-17.

Jennett, C. et al. (2016). Motivations, learning and creativity in online citizen science. *Journal of Science Communication*, Vol. 15, No. 3.

Jensen, E. A. et al. (2021). Pandemic: public feeling more positive about science. *Nature*, Vol. 591 No. 7848, pp. 34–34.

Kaiser, R. (2014). *Qualitative Experteninterviews*. Springer Fachmedien Wiesbaden, Germany.

Levy, M. and Germonprez, M. (2017). The potential for citizen science in information systems research. *Communications of the Association for Information Systems*, Vol. 40, No. 1, pp. 22–39.

Liu, H.-Y. et al. (2021). Citizen Science Platforms. *The Science of Citizen Science*. Cham, Switzerland, pp. 439–459.

Luna, S. et al. (2018). Developing Mobile Applications for Environmental and Biodiversity Citizen Science: Considerations and Recommendations. *Multimedia tools and applications for environmental & biodiversity informatics*. Cham, Switzerland, pp. 9-30.

Musto, J. and Dahanayake, A. (2021). An Approach to Improve the Quality of User-Generated Content of Citizen Science Platforms. *ISPRS International Journal of Geo-Information*, Vol. 10, No. 7, p. 434.

National Academies of Sciences, Engineering and Medicine (2018). *Learning Through Citizen Science: Enhancing Opportunities by Design*. The National Academies Press, Washington, DC, USA.

Newman, G., et al. (2010). User-Friendly Web Mapping: Lessons from a Citizen Science Website. *International Journal of Geographical Information Science*, Vol. 24, No. 12, pp. 1851–69.

Parsons, J. and Lukyanenko, R. (2011). Rethinking Data Quality as an Outcome of Conceptual Modeling Choices. *Proceedings of the 16th International Conference on Information Quality*, Adelaide, Australia.

Pedersen, A.Y. and Caviglia, F. (2019). Data Literacy as a Compound Competence. *Digital Science*. Cham, Switzerland, pp. 166–173.

Pettibone, L. et al. (2017). Understanding the (inter)disciplinary and institutional diversity of citizen science: A survey of current practice in Germany and Austria. *PLoS ONE*, Vol. 12, No. 6, pp. 1–6.

Phillips, T. et al. (2018). A Framework for Articulating and Measuring Individual Learning Outcomes from Participation in Citizen Science. *Citizen Science: Theory and Practice*, Vol. 3, No. 2, p. 3.

Radchenko, I. and Maksimenkova, O. (2016). Principles of Citizen Science in Open Educational Projects Based on Open Data. *Proceedings of the 12th Central and Eastern European Software Engineering Conference in Russia*. Moscow, Russia, pp. 1–5.

Ridsdale, C. et al. (2015). *Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report*. Available at: https://dalspace.library.dal.ca/handle/10222/64578 (Accessed: 14 May 2023).

Saraswat, S. (1999). Structured Content Analysis of Corporate Web-Sites: A Methodological Perspective. *AMCIS 1999 Proceedings*. Milwaukee, USA, pp. 206-208.

Schüller, K. et al. (2019). *Future Skills: a Framework for Data Literacy*. Available at: https://hochschulforumdigitalisierung.de/de/future-skills-ein-framework-f\%C3\%BCr-data-literacy. (Accessed: 14 May 2023)

Shirk, J. L. et al. (2012). Public Participation in Scientific Research: A Framework for Deliberate Design *Ecology and Society*, Vol. 17, pp. 29–48.

Skarlatidou, A. et al. (2019). What do volunteers want from citizen science technologies? A systematic literature review and best practice guidelines. *Journal of Science Communication*, Vol. 18, No. 1,

p. A02.

Socientize (2014). *Green paper on Citizen Science for Europe: Towards a society of empowered citizens and enhanced research*. Available at: https://digital-strategy.ec.europa.eu/en/library/green-paper-citizen-science-europe-towards-society-empowered-citizens-and-enhanced-research (Accessed: 14 May 2023).

Sorensen, A. E. et al. (2019). Reflecting on efforts to design an inclusive citizen science project in West Baltimore. *Citizen Science: Theory and Practice*, Vol. 4, No. 1., p. 13.

Spasiano, A. et al. (2021). Towards a Transdisciplinary Theoretical Framework of Citizen Science: Results from a Meta-Review Analysis. *Sustainability*, Vol. 13, No. 14, p. 7904.

Sturm, U. et al. (2018). Defining principles for mobile apps and platforms development in citizen science. *Research Ideas and Outcomes*, Vol. 4.

Vohland, K. et al. (2021). Editorial: The Science of Citizen Science Evolves. *The Science of Citizen Science*. Cham, Switzerland, pp. 1–12.

Wald, D. M., et al. (2016). Design principles for engaging and retaining virtual citizen scientists. *Conservation Biology*, Vol. 30, No. 3, pp. 562–570.

Wise, A. F. (2020). Educating Data Scientists and Data Literate Citizens for a New Generation of Data. *Journal of the Learning Sciences*, Vol. 29, No. 1, pp. 165–181.

Yadav, P. and Darlington, J. (2016). Design Guidelines for the User-Centred Collaborative Citizen Science Platforms. *Human Computation*, Vol. 3, No. 1, pp. 205-211.

Zagalsky, A. et al. (2015). The Emergence of GitHub as a Collaborative Platform for Education. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. New York, NY, USA pp. 1906–1917.