Form@re
Open Journal per la formazione in rete

articoli

# Datathons across the globe: Potential for the development of (Open) Data Literacy

# Dathatons nel mondo: Potenziale per lo sviluppo di una alfabetizzazione ai dati (aperti)

Eugenia Loría-Solano[a], Juliana E. Raffaghelli[b], Montse Guitert-Catasús[c]

[a] *Universitat Oberta de Catalunya, eloria@uoc.edu*
[b] *Università degli Studi di Padova, juliana.raffaghelli@unipd.it*
[c] *Universitat Oberta de Catalunya, mguitert@uoc.edu*

## Abstract

This paper introduces a study aimed at understanding informal learning to develop open data literacy. Based on the open datasets and reports published by the Open Knowledge Foundation (OKF) on the globally organised event Open Data Day 2021, we extracted 20 records and applied to them quantitative and thematic analysis. Convergent and diversified patterns of engagement with the Open Data Day across the five continents were observed. On this basis, some formal and particularly, informal learning activities arise as a collaborative and participatory work of data communities seeking to work with open data. Along the several experiences there is concern about the technical skills to search, find and use Open Data, including *data storytelling* abilities. However, this set of abilities was not connected with a set of critical and holistic skills, required to promote some of the goals expressed by the ODDs. Overall, integrating formal and informal learning into a situated context should be considered to promote open data literacy.

Keywords: open data; dathatons; informal learning; data literacy.

## Sintesi

Questo documento introduce uno studio volto a comprendere l'apprendimento informale per sviluppare l'alfabetizzazione dei dati aperti. Sulla base dei set di dati aperti e dei rapporti pubblicati dalla Open Knowledge Foundation (OKF) sull'evento Open Data Day 2021 organizzato a livello globale, abbiamo estratto 20 record e applicato loro analisi quantitative e tematiche. Sono stati osservati modelli di impegno convergenti e diversificati con l'Open Data Day nei cinque continenti. Su queste basi, alcune attività di apprendimento formale e in particolare informale nascono come lavoro collaborativo e partecipativo delle comunità di dati che cercano di lavorare con i dati aperti. Lungo le diverse esperienze c'è preoccupazione per le capacità tecniche per cercare, trovare e utilizzare gli Open Data, comprese le capacità di *data storytelling*. Tuttavia, questo insieme di abilità non era connesso con un insieme di abilità critiche e olistiche, necessarie per promuovere alcuni degli obiettivi espressi dagli ODD. Nel complesso, l'integrazione dell'apprendimento formale e informale in un contesto situato dovrebbe essere presa in considerazione per promuovere l'alfabetizzazione dei dati aperti.

Parole chiave: dati aperti; datathons; apprendimento informale; alfabetizzazione ai dati.

## 1. Introduction

If citizens are to navigate the turbulent waters of data and algorithms, then data literacy must be featured not only in academic programmes. Fostering reflection on how data are constructed and operationalized across societies and providing opportunities to learn from the analysis of data from a range of sources, from the classroom to community environments, is crucial. Recent research has identified a lack of data literacy as one of the primary obstacles to the use of open data for empowerment, engagement, and an overall critical approach to data in our society (Boychuk, Cousins, Lloyd, & MacKeigan, 2016; Raffaghelli, 2020). Literature reveals a dearth of systemic interventions to build what we could call *open data literacy,* as well as research on the design and implementation of relevant techniques to promote open data literacy (Coughlan, 2019).

Citizens' knowledge, attitudes, and abilities to make targeted and efficient use of open data for their own lives and socio-cultural situations appear to be crucial (Boychuk et al. 2016). Consequently, it might be hypothesised that the development of abilities to deal with data infrastructures and domain knowledge for their application are factors that enable citizens and professionals to make more specific and effective use of data (Gil-Garcia et al., 2020). This attitude towards data cannot be purely theoretical or attained solely through formal education. Occurring in situated social and political contexts, it requires active participation and political, environmental, personal, and professional interests as the drivers of informal learning to acquire skills and knowledge useful for the transformational processes in which citizens are engaged (D'Ignazio & Klein, 2020). In addition, technical data literacy alone would not suffice; open government data has been identified as a driver of transparency and transformation based on a holistic, contextual, and critical understanding of data in their political contexts of production (Baack, 2015). There is a body of literature that discusses the role of critical approaches to data and advocates for the necessity of *hacking* open data to promote the empowerment of citizens and the visibility of hidden collectives (Pybus et al., 2015).

Hence, a broader picture of existing practices and activities supporting the development of open data literacy in contextual contexts should be useful for the development of systematic educational interventions with a later influence on data-based citizen empowerment. Specifically, it would target higher education and school initiatives aimed at aiding other collectives engaged in civic participation, thus promoting a more complex and sustainable conception of education and learning (Atenas et al., 2020).

## 2. Background

Participation in democracy in today's digital and data-driven society necessitates the development of several transversal skills, which should be fostered in higher education (HE) through critically oriented pedagogies that interweave technical data skills and practises with information and media literacies (Hobbs, 2020). If students are to navigate the choppy waters of data and algorithms, data literacies must be incorporated into academic development programmes, enabling higher education to take the lead in the development of approaches to understanding and analysing data in order to foster reflection on how data are constructed and operationalized across societies and provide opportunities to learn from the analysis of data from a variety of sources (Bhargava et al., 2015; OECD, 2018). Several solutions rely on the utilisation of open data as open educational resources within the context of problem- and research-based learning activities (Coughlan, 2019;

Raffaghelli, 2018). It has also been argued that academic training and curriculum design programmes in HE should be established to promote advanced practices with open data that contribute to civic participation and empowerment (Atenas et al., 2020).

In a preliminary review of the literature by the authors (Loría-Soriano, Guitert & Raffaghelli, 2023), 66 articles indicate that active engagement and situated usage of open government data for significant causes can lead to the development of the necessary capacities or skills required for a critical approach to open data.

This preliminary research revealed pertinent information about the need for data literacy, but it also highlighted pertinent questions about how citizens interacted with public activities and the extent to which these assisted their learning.

Beyond the review of the literature, mapping practices related to the development of open data literacies and the usage of open government data (OGD) could support a better understanding of contextual motivations, the forms of informal and non-formal learning, and the citizens' learning needs. In this regard, interviews, case studies, and ethnographic observation can be deemed relevant as a methodological approach to exploring the development of data literacy through informal and non-formal adults' learning. Nonetheless, such idiographic methods are extremely focused on specific contexts and cross-cultural, global comparisons (Okoko, Tunison, & Walker, 2023). Though nomothetic, quantitative studies cover broader samples and allow comparative international studies, they can also be considered schematic or impose a top-down (researcher-respondent) approach (Babones, 2016). In this regard, the analysis of ongoing participatory events through extractive methods on the web is a new frontier of educational research (Kimmons & Veletsianos, 2018). Therefore, this research builds on existing rich open data, namely the reports resulting from a specific type of data day (the Open Data Day, or ODD), gathered on one page by the organisers (the Open Knowledge Foundation). These datasets are relevant, easily available, and abundantly connected with reports and open pages gathering participants' narratives. Therefore, our research is observational and non-invasive, building on the wealth of existing public information. The relevance of our work could be deemed twofold: from one side, to make sense and enhance the wealth of public information and existing open data; from the other, to shed light on informal educational dynamics that could be promoted, re-organised, or triggered to expand citizens' abilities and awareness about data as an instrument for empowerment.

## 3. Methods

The following research question was formulated over the basis of the assumptions made conceptually: How could we characterise the contribution of datathons in the context of the ODD, as informal learning spaces with focuses on situated problems and topics, to promote the usage of open data as final outcome?

To respond to the research question, a preliminary study based on mapping informal learning activities (within datathons) around Open Data Day was carried out. In fact, we investigated how such datathons were diversified across the globe, which relevant topics and problems were considered, and which types of data literacy emerged through data appropriation and usage.

In this regard, the practices mapped came from a convenient but relevant sample: activities scheduled and documented at the Open Data Day 2021, as a relevant, global event

organised by the Open Knowledge Foundation (OKF) (https://okfn.org/). The ODD is organised by the Open Knowledge Foundation in synergy with other organizations that promote the use of open data and or training to achieve such goal. This organisation also supports participation by providing mini grants for the organization of open data use and learning events around the world. In a nutshell, the Open Data Day is an annual celebration of open data all over the world. Groups from around the world create local events on the day where they will use open data in their communities. It is an opportunity to show the benefits of open data and encourage the adoption of open data policies in government, business, and civil society. All outputs are open for everyone to use and re-use (https://opendataday.org/). On the Open data day webpage, the data of the events held on the open data day held in 2021 can be found. There is available a list of event reports can be obtained, which include aspects such as type of activity, country, and city where it was held, whether it was held virtually or in person, event theme, methodology used, types of data as well as lessons learned, among others. The list of reports is available at https://opendataday.org/events/2021/. As expressed before, using these available datasets is a great opportunity to better understand the practical phenomenon of open data usage and how knowledge is built at the community level through collaborative and participatory ways at a global level. In the same vein, this database can be considered a frame of reference for the identification of subject matter experts who have organised such events and who are located around the world.

The activities were randomly selected from the overall list of data and reports published in their website as open datasets by the OKF, yielding a total of 20 events with a full report with descriptions from all continents. Specifically, from the following countries: Somalia, Niger, Cameroon, US, Mexico, Costa Rica, Colombia, Brazil, Russia, Afghanistan, Georgia, India, Thailand, Japan, Australia, United Kingdom, Switzerland, Hungary, Italy. The data extracted consisted in 327 events held in 2021 Open Data Day which is a file with quantitative information, such as the following variables: *country event_date, event_name, event_photo_url, event_purpose, event_report_url, event_time, event_tweet_url, event_video_url, has_event_report, latitude, longitude, mini_grant_winner, num_participants, online, online_event_url, organisers, place, report_question_1, report_question_2, report_question_3, slug,timezone, url, world_region_code and world_region_text*. Also, a qualitative dataset was extracted. It was composed by a corpus of documents (from the 327 events, 85 of them have report on "Lessons Learned"). The extension in words of the text gathered was of 39756 words.

To approach such analysis, we adopted a mixed method, data-driven approach (Kimmons, 2022; Kimmons, R., & Veletsianos, 2018).

Initially, we carried out a quantitative analysis over the datasets published by the OKF. Firstly, the quantitative data yielded through open datasets was analysed for a better understanding of the scope of ODD events held in 2021. Also, the number of participants by location was considered to understand the magnitude of the events in terms of reaching participants from different sectors and stakeholders around open data. As for the text analysis, we adopted the thematic Analysis method frequently adopted in qualitative research (Braun & Clarke, 2021) was adopted to identify patterns of meaning within the participants' discourses and documents. Through this method, we specifically, observed:

- the diversified geopolitical contexts;
- the topics and interests revealed by the local populations;
- the several types of data literacy emerged;

- the types of emotional engagement (sentiments) around data in the several contexts studied.

We analysed the text in search of expressions taking into consideration skills, knowledge, attitudes to be promoted through the activities, namely, as expressions of informal learning. In fact, informal learning is not an initial purpose of ODDs, but as referred in the literature, it can be captured either through observation (with pedagogical instruments such the idea of literacies) or through self-verbalisation (http://www.eucen.eu/sites/default/files/OECD_RNFIFL2010_Werquin.pdf). We were particularly interested in the diversification of learning experiences leading to open data skills. In this regard and basing on a prior analysis of the literature (Loría et al., forthcoming), we divided the types of data literacies in *technical*, namely, abilities of data disaggregation, data visualization, statistical methods and reasoning; and *critical*, as skills connected with civic participation, cultivating diversity, civic empowerment, and visibility of vulnerable groups through the use of data. Moreover, the section "Lessons Learned" from the ODD publication, supported our search about the ODD's implications for the citizenship.

Ultimately, and as further text-mining technique, we used Sentiment Analysis. This technique consists in the analysis of text by identifying, extracting, and quantifying systematically the words that have emotional connotations. In this regard, there are different "lexicons" or dictionaries that are adopted by automated systems embedded into research software to process text corpora (a collection of texts). In this case, we adopted NVIVO's tool to perform the Sentiment Analysis (https://help-nv.qsrinternational.com/12/win/v12.1.110-d3ea61/Content/coding/auto-detect-code-sentiment.htm). Overall, the analysis was made comparing the texts as separated corpora, therefore, aiming at comparisons between continents.

## 4. Results

Results of the quantitative analysis, the qualitative thematic analysis, the keyword mapping analysis and the sentiment analysis are presented in the following. The quantitative analysis shed light on very general issues, while the meaning making processes around the events and the people engagement emerged more clearly through the qualitative (manual and automated) analysis. 4.1. Quantitative descriptive analysis.

Figure 1 presents the number of participants per continent and subregion, as extracted from the database available in the OKF. We observe that overall, 1681 participants were engaged in the ODD events across the globe.

| Continent | Sub-Region | Mean of participants |
|---|---|---|
| Africa | Central Africa | 50 |
| | East Africa | 55 |
| | North Africa | 30 |
| | Southern Africa | 43 |
| | West Africa | 83 |
| America | Central America + C | 87 |
| | North America | 93 |
| | South America | 105 |
| Asia | Central Asia | 20 |

| | | |
|---|---|---|
| | East Asia | 34 |
| | Middle East | 190 |
| | North Asia | 80 |
| | South Asia | 68 |
| | Southeast Asia | 186 |
| Europe | East Europe | 58 |
| | North Europe | 64 |
| | Southern Europe | 81 |
| | West Europe | 79 |
| Oceania | Australia | 275 |

Figure 1. Distribution of participants by continent and sub-region.

The continent with the highest massive event is Oceania because in this region a webinar with high audience was organised . Other events with high attendance took place in Middle East and Southeast Asia. The boxplot in Figure 2 highlights that most events across the globe had an audience ranging from one to fifty participants.

The variability between events in terms of participants, with a median of fifty participants and a mean of seventy-eight participants become evident. Some few massive events with more than two hundred attendees took place through the online format due to the COVID-19 restrictions. This was an element of advantage, letting the participation of a higher number of participants from different locations, even reaching distant rural communities.

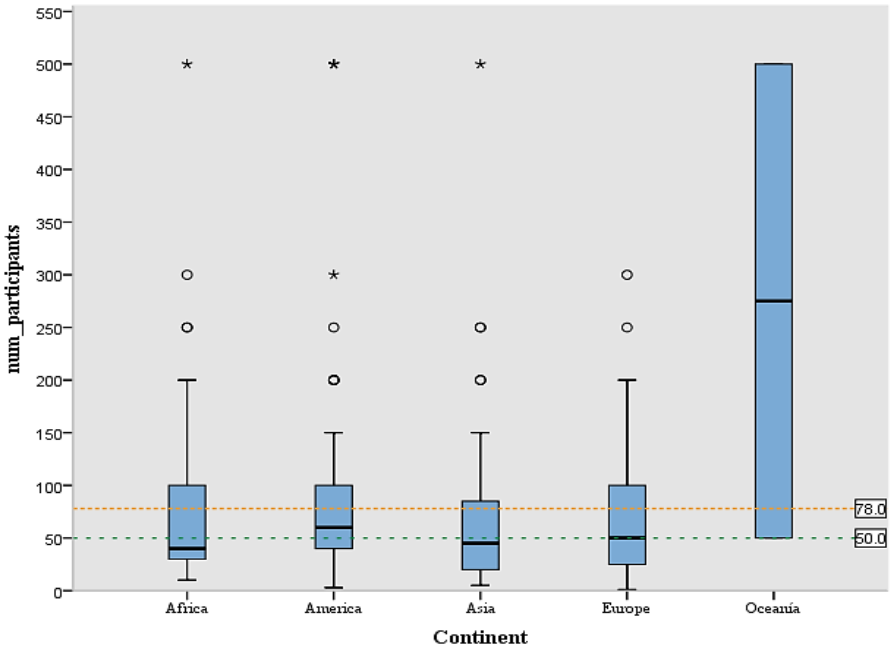


Figure 2. Boxplot of number of Open Data Day events' participants by continent.

### 4.1. Thematic Analysis

Figure 3 introduces a tree of themes and subthemes relating to learning, pedagogical practices and the use of open data in general.

| Name | Documents | References |
|---|---|---|
| Activism | 11 | 14 |
| Challenges | 3 | 5 |
| Data Literacy | 0 | 0 |
| Critical data literacy | 2 | 2 |
| Civic participation | 10 | 14 |
| Diversity | 9 | 13 |
| Empowerment | 5 | 7 |
| Visibility | 4 | 6 |
| Tecnical Data Literacy | 5 | 5 |
| Data disaggregation | 5 | 7 |
| Data visualization | 7 | 13 |
| Statistical methods | 2 | 3 |
| Formal Learning | 9 | 13 |
| Training hours | 4 | 4 |
| Glocal | 11 | 12 |
| Informal learning | 3 | 3 |
| Interactive discussion | 4 | 5 |
| Massive online | 6 | 8 |
| Online | 14 | 18 |
| Outputs | 9 | 19 |
| Skills Requirements | 1 | 1 |
| Critical | 6 | 7 |
| Technical | 9 | 15 |

Figure 3. Code-tree extracted from the NVIVO corpus. Documents are text units from the overall corpus, references are the text bracketed by the researcher, representing the theme. themes and subthemes.

We can observe the topics emerged through an inductive codification. Particularly, we can grasp the relevance given to civic participation (14 references in ten documents) and diversity (nine reference sin 13 documents), as elements of a critical data literacy. Nonetheless, technical data literacy is also emphasised, particularly in relation with data visualisation. As for the learning methods, training hours appear to have the same relevance than interactive discussions. More importantly, massive participation through online environments is evident. As for the skills requirements, we observe the prevalence of focus on technical skills in comparison with critical approaches to data (15 references for the first in comparison with the seven references for the second).

As for the themes identified, overall, the analysis yielded that globally, the importance of open data for community work supporting public policy-making, solidarity, and transparency were valued. Moreover, that this work should be knowledgeable, making the citizens to feel closer to data. Nevertheless, specific themes by Continent were identified.

The main themes addressed in the continents were equal development, gender, and democracy for Asia; labour market, and digital citizenships for the Americas; data science for innovation for Oceania; COVID-19 for Africa; and for Europe topics such as data-driven investigation for transparency and detecting potential frauds and injustice. Across continents, we noticed indeed clear differences between the several geopolitical areas. For

the Americas, more importance of open data for public policy and transparency was given, from the micro to the macro, i.e., from the personal level to the city and country level. As for Asia the focus was put on the availability and openness of public information used in international open data day events through technology for solidarity. In the case of Australia, the focus was put on the promotion of open data use in engaging way for the World Open Data Day. Europe centred its attention instead on awareness of the importance of everyone's open data ideas and projects as part of a joint transnational community.

The above-mentioned contexts set the bases to support formal, non-formal and informal learning activities with impact on data literacy The events took generally the form of webinars and workshops, which embedded components aimed at learning how to explore data and how to use it for specific activities such as crowded sourced maps (Mapathons) or problem solving/creative activities based on open data (dathatons). Such activities could be deemed formal or non-formal. Also, community. interactive discussions about topics such as: data gaps, weaknesses, and the need to strengthen sex-disaggregated data and social and economic issues in their communities could be considered spaces promoting informal learning. According to the reports the above-mentioned activities took around 3 to 4 hours, and in some cases inclusive weeks. Figure 4 presents specific text extracts from the reports referring to learning activities performed by continent.

Diving further in the type of data literacy promoted in the events, *technical data literacy* focused more on the promotion and spread the use of open data and data driven initiatives, as presented the texts found in the reports and classified under the main three technical data literacy thematic nodes which are stated in the columns of Figure 5.

| Continent | Activity | Type of literacy |
|---|---|---|
| Africa | Open data training with the OpenStreetMap platform | TDL |
| | Mapping techniques using OpenStreetMap | TDL |
| | Examples of accessible data (citizen engagement; access to lifesaving techniques in government intermediaries, education, and healthcare) improve life | TDL & CDL |
| | Guidelines, talk/seminar, digital demonstrations, and data training | TDL |
| | Interactive discussions on data gaps, weaknesses, and the need to strengthen sex-disaggregated data | TDL |
| America | Workshop | TDL |
| | 10 Data Science Workshops facilitated by INTEL specialists | TDL |
| | Community | CDL |
| | Tutorials and demo days at http://GreenMap.org/stories/ogm2 | TDL |
| | Skills improvement by taking part and learning in other events held on the day | TDL |
| Asia | Open discussions | CDL |
| | Openness and knowledge sharing for innovation | TDL |
| Australia | Webinar | TDL |
| Europe | Collective/Critical Cartography online workshop | TDL & CDL |
| | Peer-reviewed by a wider audience | TDL |
| | Workshop | TDL |

| | |
|---|---|
| Short tutorial to learn how to map the simplest features, namely buildings, with iDeditor | TDL |
| Collaborative mapping through an overview of the OpenStreetMap project | TDL & CDL |

Figure 4. Informal learning activities mapped by continent [Caption: Technical Data Literacy= TDL // Critical Data Literacy=CDL].

| Data Disaggregation | Data Visualization | Statistical Methods |
|---|---|---|
| Our website organises and structures all our data. Gênero e Número's Open Data session [USA] | Mapathon. [Africa] 'Collaborative and Humanitarian Mapping', green living icons and engagement practices. Datathons [Europe] | Analysing the first dataset published in Costa Rica under the Open Contracting Data Standard (OCDS) [America] |
| The dataset includes 12,491 contracting processes managed by the Supreme Court of Justice covering the period from January 2016 to October 2020 [America] | We learn about mapping techniques using OpenStreetMap [Africa] Acquiring mapping skills, such mapping practices in most of the cases also reveal new data about the world around us [Europe] | Notebook where they documented their analysis procedures [America] |
| Machine-readable format [Europe] | Resourcing technologies that could be used for smart visualization of data [Africa] | Synthesize data out of our user's activity [Europe] |
| Opener application [Europe] | Encourage the re-use of payment data for the early detection of potential conflicts of interest and raise the participants' data visualizations skills [Europe] | Exploratory data converting/sharing scripts in Grafoscopio interactive notebooks [America] |
| Accessible [Europe] | Collective/critical cartography can help reveal social processes from macro-level processes on the global scale to micro-level processes on the scale of the human body [Europe] | Sex-disaggregated data policies in government, business, and civil society [Africa] |

Figure 5. Texts extracts clustered by Sub-nodes mapped under the node of technical data literacies.

Furthermore, the Critical data literacy nodes (Figure 6) classified texts which focus were: civic participation, diversity, empowerment, and visibility of socioeconomic issues in the communities, covering topics such as sex-disaggregated data policies in government, business, and civil society for achieving gender equality and the empowerment of all women and girls, collaborative and humanitarian mapping and increasing awareness of the problems of their communities through open data analysis results. The specific nodes are expressed in the columns and the texts extracts reveals the topics and activities.

| Civic participation | Diversity | Empowerment | Visibility |
|---|---|---|---|
| It's a chance to map our region's entities for social improvement decision-making. [Africa] | Including local views and topics. Diverse backgrounds, gender, and age made discussions interesting. [America] | To empower them to advance open data in this rural dynamic region. [Africa] | Mapping drought cycles plaguing communities in Gambos. Difficult access to water, and main public services [Africa] |

| | | | |
|---|---|---|---|
| Open Data Development in rural Zimbabwe amid social divisions, instability, and COVID19 [Africa] | Many don't know how to find the pandemic's women's labour market. [America] | A database for everyone to use to encourage development, social cohesion, and innovative collaboration [Africa] | Open Data Day is a great chance to learn, share, create, and envisage a data-driven world. [America] |
| Government oversight and reducing corruption through local openness [Africa] | To bring together diverse people to learn, share, create, and dream a better data-driven future. [America] | Discovered the enormous potential of open data for developing educational, social, professional projects [America] | Women who lead in mapping and responding in times of pandemic, and trans women in tech [Asia] |
| Publicizing citizen-government interactions. Increasing governmental procurement competition and transparency [America] | International Women's Day and Trans-Visibility Week: #ChooseToChallenge Gender Inequality in Open Mapping [Asia] | | The places that are less mapped are less accessed [Asia] |
| Data-driven investigation for transparency and detecting potential frauds [Europe] | Aim to increase representation and participation of women and LGBTQIA+ in the OpenStreetMap [Asia] | Importance of women's perspective in mapping [Asia] | Housing injustices in different regions of the world [Europe] |
| Digital citizenship and progressive change. Re-using publicly available payment data of the public sector [America] | Everyone shares and makes it better for everyone. Collective cartography projects show perspectives, contexts, designs, uses and tools different from the dominant ones [Europe] | Co-created maps to develop collective action against dominant narratives & existing power structures. Gaining ownership over data for humanitarian purposes [Europe] | Work with children and marginalised groups underlined their experiences, feelings, and personal stories. [Europe] |

Figure 6. Texts extracts clustered by Sub-nodes mapped under the node of critical data literacies.

Despite the prevalence of technical data literacy interests, across the events critical data literacy emerges and it is cultivated through the citizen participation, the value of community empowerment, the minorities' rights and visibility, and political participation.

Other dimensions of the analysis are thematic nodes of topics such as Outputs, Activism, and Challenges. These nodes can be understood as lessons learned too. Specifically for the Outputs, the main elements from the events are collective, humanitarian, and critical mapping and creation of geographical information, co-creation and release of new databases and coding, presentations and showcases to government stakeholders to increase awareness about local communities and what can be done with open data, connections and conversations with community, collective learning. The node of Activism shows that it is associated with the intention to increase visibility of economic and social needs such as social fissures, insecurity and covid 19 pandemic, to improve innovation, transparency and

collaboration in procurement processes, awareness of gender and race issues, also to increase public institutions awareness about local communities' requirements and actions. Green Map movement, collective/critical cartography and 'Collaborative and Humanitarian Mapping' can help reveal social processes, progressive change. Finally, the node of the challenges explored through the analysis of the corpus, the focus was put on machine-readable data, availability, capacity and understand, mostly related to a technical ability to deal with data. Also, other highlighted challenges are around community synergies around data, what makes emerge the relevance of societal goals behind data, but in this case, the types of skills needed are less clearly focused. Allegedly, this would make it difficult to work out educational and lifelong learning strategies.

### 4.2. Sentiment Analysis

Figure 7 introduces the sentiment analysis performed using NVIVO. As it can be seen, there is a prevalence of positive sentiments about the activities, though "moderately positive" is the most frequent labelling (78 of 89 references, in 19 documents of 19). Less negative expressions can be found overall (32 with regard to 89 overall positive references). Nevertheless, 22 moderately negative references, and 10 highly negative references were found, a number that overcome the highly positive expressions. We might assume that event hough there was a general positive impact relating the datathons, negative elements might have emerged, probably connected with the difficult topics and situations dealt with.

| Name | Documents | References |
|---|---|---|
| Positive | 19 | 89 |
| Highly positive | 6 | 11 |
| Moderately positive | 19 | 78 |
| Negative | 14 | 32 |
| Moderately negative | 12 | 22 |
| Highly negative | 8 | 10 |

Figure 7. Sentiment analysis

Figure 8 presents sentiment analysis levels by continent, also shows the titles of the sampled informs under study.

| Continent | Report | [- -] | [-] | [+] | [++] |
|---|---|---|---|---|---|
| Africa | 1: Advancing the production, dissemination and openness of sex disaggregated data in Somalia - Event Report | 19% | 47% | 34% | 0% |
| | 2: Angola Gambos Mapathon - Event Report | 0% | 100% | 0% | 0% |
| | 3: Cameroon Geoevangelisation workshop in the use of JOSM and GeOsm, the first 100% African open source geolocation platform - Event Report | 0% | 0% | 100% | 0% |
| | 4: Nigeria Abuja Open Data Day - Event Report | 22% | 9% | 69% | 0% |

| | | | | | |
|---|---|---|---|---|---|
| | 5: Open Data Development in an age of social fissures, insecurity and COVID19 in rural Zimbabwe - Event Report | 6% | 32% | 37% | 25% |
| America | 1: Brazil Gênero e dados o mercado de trabalho para mulheres na pandemia - Event Report | 23% | 10% | 66% | 0% |
| | 2: Colombia Cajicá Ciudadanías y Digitales Gobierno Abierto - Event Report | 0% | 0% | 100% | 0% |
| | 3: CRC Data Challenge take advantage of the first dataset published under the OCDS in the country - Event Report | 0% | 2% | 98% | 0% |
| | 4: US-Open Green Map 2 - OGM2 and YOU! - Event Report | 0% | 0% | 77% | 23% |
| Asia | 1: Afghanistan Celebrating Open Data Day (Promoting Open Data in Afghanistan & Beyond) - Event Report | 0% | 29% | 15% | 56% |
| | 2: Georgia PechaKucha Night Tbilisi Open Data for Equal Development - Event Report | 0% | 40% | 60% | 0% |
| | 3: India Open Environment Data - 2021 - Event Report | 0% | 36% | 64% | 0% |
| | 4: Open Data Day in Moscow - Event Report | 0% | 0% | 100% | 0% |
| | 5: Philippines MAPAbabae 2021 ChooseToChallenge Gender Inequality in Open Mapping - Event Report | 15% | 18% | 67% | 0% |
| | 6: Thailand Open Data for Democracy Emphasise what would be lost if we don't have open data in our country - Event Report | 28% | 0% | 72% | 0% |
| Australia | 1: AUSTRALIA International open Data Day Webinar | 31% | 0% | 69% | 0% |
| Europe | 1: Estonia OKEE Andmeklubi - Event Report | 26% | 15% | 34% | 26% |
| | 2: Germany | 0% | 0% | 38% | 62% |
| | 3: Hungary Open data for critical mapping of gentrification in Hungary - Event Report | 0% | 15% | 85% | 0% |
| | 4: Italy - Collaborative and Humanitarian Mapping 1 OpenStreetMap introduction (Open Data Day) - Event Report | 0% | 0% | 78% | 22% |

Figure 8. Sentiment analysis relative distribution by report and continent. [Caption: [--] Very Negative Sentiment, [-] Moderately Negative Sentiment, [+] Moderately Positive Sentiment, [++] Very Positive Sentiment].

The percentages are estimated by row to highlight the importance of every level of the sentiment analysis unveiling a moderately positive as the category with the highest percentages in almost every row. The sentiment analysis yielded relevant elements around the importance of some topics and the areas where problems are still very present. Almost all comments are related to the importance of open data for the communities for the search of solutions for problems such as lack of transparency, lack of access, lack of knowledge

on how to use the data, the lack of disaggregated data by gender, the scepticism of relevant stakeholders and decision makers about the impact of open data usage. Moreover, most of the negative feelings are about the current problems that communities face and the lack of data usage due to access problems or lack of the know-how to interact with a real set of data. On the other hand, the positive feelings are about how the usage of open data available can be key for data driven policy and visibility of problems at the same time searching for the solutions in a fun way and in a community of open data enthusiast. The positive comments reflect the importance of projects to promote, connect and learn about the very different ways people put open data to good use worldwide and the importance of the engagement of different governmental stakeholders and civil society to encourage the re-use of public data to raise the participants' skills while detection of potential conflicts of interest in the public procurement process.

## 5. Discussion

In this paper, we aimed at answering the following research question:

How could we characterise the contribution of datathons in the context of the ODD, as formal, non-formal and informal learning spaces with focuses on situated problems and topics, to promote open data literacy as outcome?

Our preliminary study based on 20 existing open datasets and reports published by the Open Knowledge Foundation (OKF) on the globally organised event Open Data Day 2021, yielded relevant information. Indeed the richness of the information extracted, not purposely collected through an a priori design required the triangulation of methods (Flik, 2004) in order to explore informal learning as complex construct.

We performed a number of a mixed-methods, expert-led and data-driven analysis consisting of in a quantitative descriptive analysis of the number of participants by event, continent, and subregion; followed by a thematic and sentiment analysis of the type of learning activities classified by tree of nodes thematic analysis. Through this work we came up with an interesting global picture on citizen engagement with open data. Moreover, we observed how this could liaise with data literacy skills development by informal learning events.

In the international context there is attention to the development of data literacy frameworks, also appliable to deal with open research and government data (Maybee & Zilinski, 2015), their application seems to be connected with formal learning settings and to disciplinary compartments, particularly relating statistical literacy and data science (Smolnikova, 2020). However, contextualised adult learning could not be covered by such approaches. Therefore, community learning, as part as situated strategies of lifelong learning, should be carefully considered to achieve the necessary literacies to deal with data, and particularly with open data. This is in line with the findings made by Zuiderwijk, Janssen, and Dwivedi (2015) who highlighted the factors encompassing acceptance and engagement with open data. Not only understanding and accessible data infrastructures (like retrievable open datasets and open data portals) is relevant. Also, the social relevance given by close others (neighbours, peers, collectives with similar ideals) to data and data practices support the motivations to engage with them. In any case, we must consider that community learning has been deemed a relevant source of adults' education in other fields (Foley, 2020). However the community context offers and requires forms of learning that go well beyond the technical (Biesta & Tedder, 2007). In the specific case of data literacy,

as Jansen (2021) purports, there is need of "re-politicise" data literacy, "reflecting on the question of literacy, asking toward what end these approaches aim to build critically informed users or critically informed citizens" (p. 5). This researcher has elaborated several resources to support a critical vision of data literacy, and basing on her experience, the ultimate goal of scrutinising data is always political. Indeed, as she claims, "those resources that have been created within the third sector, and are aimed at increasing literacy of human rights defenders, civil society organisations, journalists and activists […] are designed to be problem-oriented, experiential, actionable and relevant for the learners' context" (p. 6) what evidently means social and cultural contexts of life based on political actions and debates. The prevalence of technical data literacy skills requirements to take part in the datathons, as emerged from the thematic analysis, might invisibilise the relevance of critical data literacy. Eventhough, this last emerge across the documents analysed and it appears to be cultivated through the citizen participation, the value of community empowerment, the minorities' rights, and visibility, and the several forms of political participation. Technical data literacy might be deemed essential, since the first activity with open data is to approach them, understand where or how they are configured, linked, extracted, analised and/or mapped. But awareness on the skills to critically engage and scrutinise data is also necessary. A relevant element to discuss, in this regard, is the need of integrating informal learning and formal education to support civic participation in relation with open data. Becoming active around data, in a datafied society, means also to cultivate data justice (Dencik & Sanchez-Monedero, 2022). Therefore, data literacy must be cultivated throughout lifelong learning, from experiences and activities in the schooling systems, towards civic and political participation around data (Sander, 2020). Moreover, data justice could be achieved only through diversified forms of educational interventions, not only through traditional, formal education promoting statistical literacy or data science (Raffaghelli, 2023). As out analysis shown, civic engagement happens around topics that are crucial at a situated level. Therefore, the encapsulated learning in formal education cannot be sufficient to cover the data literacy needs for this form of participation. So far, the sentiment analysis reveals that the participants are enthusiastic about having access to open data and are aware of the need to acquire skills to access, utilise, and produce results based on open data. However, they are ambivalent (showing moderate sentiment and some negative sentiments) about their ability to apply open data to community issues. Nonetheless, the type of sentiments also pointed out the relevance of issues such as representation, visibility, access, and citizen rights around data. The engagement with such topics across datathons plausibly generates spaces to discuss and focus problems, supporting the development of understanding and the identification of key skills needed for socio-cultural processes of transformation. Problems that clearly require critical understandings of data. This is consistent with the prior findings at the thematic analysis level. Technical data literacy is necessary, though not sufficient to become a catalyst for civic participation. In this regard, the quality of education has frequently been evaluated based on its impact on the development of skills that enable individuals to obtain better employment and, consequently, to live better lives within a given status quo. Such discourses, connected to the theory of cultural capital have plagued the entire literature from the 1960s on (Lareau & Weininger, 2003). In the more technocratic tradition of education systems, education has also been criticised for their lack of effectiveness in meeting the skills required in the labour market (Carey, 2015). Educational researchers have addressed such criticisms by emphasising the relevance of lifelong learning as a continuum between formal, non-formal and informal learning experiences as a personal, educational pathway (Blaschke, 2012). But also this last approach was contested from the philosophy of education, through the conceptual lenses of what was called the

*learnification* of education systems, i.e. an overemphasis on the controlled design of learning processes that determine the formation of competencies recognised in the labour market (Biesta, 2020). The analysis of datathons displays formal education is relevant to support citizen's engagement in datathons. However, we also verified that learning (particularly informal learning) takes place along ODDs' deployment.

In light of this result, the relationship between formal and informal education should be reconsidered. As a matter of fact, the schools and higher education might participate as part of formal training at datathons, whereas the community might request that the students provide support for more technical data science tasks. Aligning D'Ignazio and Klein (2020), data scientists should not be considered "unicorns" or experts working in elitist groups; instead, they must engage with communities, interpreting social and political needs in order to use data. Therefore, students from secondary and higher education could play a role in supporting community learning about data. We insist, at this point, on the relevance of an enabling context that equips participants to achieve the symbolic, emotional, behavioural, and material means to realise their own social justice needs. As we observed in this paper, community, and government efforts, but also educational research and practice, will have to explore to what extent it is possible to generate an open curriculum and authentic educational tasks capable of embedding students' engagement in Open Data Days events (or similar). Clearly, formal education could also support processes of skill recognition achieved through ODDs. The focus of such recognition should be not only technical but also critical, embracing the understanding of data infrastructures and practices and their unintended consequences.

## 6. Conclusions

Our findings emphasised the relevance of contexts to support open data literacy, as part of civic competences to take part in a datafied society. Transnational and national policy recommendations emphasise the need to promote Open Data usage. In our research, we observed that these recommendations can be implemented in diversified ways. Equal development, gender representation, and building a democratic society were relevant for Asia; labour market, and digital citizenship was deemed important in the Americas; data science for innovation and funny/knowledgeable approaches to data was a focus for Oceania; COVID-19 and gender issues for Africa; humanitarian topics were central for Europe. Therefore, despite open data for civic engagement through solidarity, transparency, and access to data were topics valued across the all events, the specificities signalled situated social contexts of learning. We also shed light on the relevance of informal learning activities as a collaborative and participatory work around data. The several reports analysed supported the idea that there is need of integrating skills achieved through formal learning, with informal learning. In fact, while formal learning could be a source of technical data, political engagement and participation in topics that are central for the community development bring to the fore the relevance of critical data literacy as set of holistic and reflexive skills. We conclude that open data literacy requires lifelong learning settings connected to active citizenship, considering local, informal spaces and instruments that facilitate understanding, questioning and expression about data.

The first implication of our findings is that educational approaches to develop open data literacy cannot be just *taken out of the shelve* and universally applied through formal

education. However, the adult citizens willing to operate with open data might need new skills, given the fast development of datafication in the society (Van Dijck, 2014).

Informal learning within communities or citizen activities around open data should be further explored, being a relevant challenge for both policymakers and educators to create the settings for informal learning to happen. Most importantly, ways to activate the recognition of prior learning (RPL) could play a relevant role for citizens engaged with open data practices, particularly in developing countries (Andersson, Fejes, & Sandberg, 2013). However, to recognise learning, the sets of skills and knowledge should also be categorised, with relevant supporting theories relating to what data literacy is overall and which areas of data literacy can be considered. For example, in our work, we grasped that technical data skills were deemed central and frequently cited as keys to engaging with open data. However, critical data abilities were less apparent in the discourse of the monitored reports. We could just infer their relevance from the connection of such abilities with the ultimate goals of open *data*: to become empowered, to express or show community problems (like gender issues or inequities in the distribution of services).

Despite the limitations of a study based on second-hand open data, where data collection is not designed *ad hoc,* the findings in this paper can be deemed insightful given the global perspective and the richness of experiences explored. Therefore, the results might be relevant for policymakers as well as for educators when planning the development of open data literacy. Nonetheless, our study spots only superficial trends, making it necessary to deepen the dynamics observed through direct questions to the participants. Several methodological approaches could serve this purpose, spanning from short surveys and quantitative studies to interviews displaying the participants' needs for skills and knowledge to better engage with open data. This information should also be correlated with other factors triggering their acceptance or possibilities to engage with open data, not the least, the quality of the published open data, a fact that not only includes basic citizens' engagement with government data but also the researchers' open data practices (Quarati & Raffaghelli, 2020; Ruijer et al., 2016).

Further research must shed light on the way open data literacies can be cultivated, beyond academic or professional learning. Datathons can be a flexible and creative space for community learning, but their educational design and settings could be a focus of deeper analysis in the future. , As we emphasised, the informal learning activities attempting to develop technical skills to search, find, use data, and also, to generate spaces for *data storytelling* should be explored in connection with critical approaches that discuss data justice and ethics. In this regard, *datathons* as source of informal learning could be integrated with spaces to reflect about the skills achieved, through self-assessment and participatory evaluation. As a matter of fact, D'Ignazio & Bhargava (2015) refer to forms of popular education where data is assessed considering the needs of communities from participatory frameworks and socio-cultural reflection.

Also, specific topics relating to collectives struggling to make their voice heard should be promoted as cases to support community reflection, embracing feminist, eco-critical and emancipatory paradigms. In such contexts, recognition should go beyond just *acknowledging* the achievement of technical skills. By supporting integrated groups of stakeholders (the same community, local government, the formal system, and the private sector) a participant collection of narratives about the learned lessons could become a record of capabilities, or could be further connected to the participant's identities as lifelong learners (Biesta & Tedder, 2007). Methodological approaches, instruments, implementation, immediate and mid-term effects will require specific studies.

Open data literacy, indeed, cannot be seen through the lens of rigid, international frameworks, but requires situated approaches and local awareness. These approaches are yet to be conceptualised and piloted within relevant socio-cultural contexts, from the Global South, from indigenous populations, from hidden collectives, across the globe.

## Disclaimer

The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Institutions they work for.

## Reference list

Andersson, P., Fejes, A., & Sandberg, F. (2013). Introducing research on recognition of prior learning. *International Journal of Lifelong Education*, *32*(4), 405–411.

Atenas, J., Havemann, L., & Timmermann, C. (2020). Critical literacies for a datafied society: Academic development and curriculum design in higher education. *Research in Learning Technology*, *28*, 2020. https://doi.org/10.25304/rlt.v28.2468

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society, 2*(2). http://dx.doi.org/10.1177/2053951715594634 (ver. 28.11.2023).

Babones, S. (2016). Interpretive Quantitative Methods for the Social Sciences. *Sociology*, *50*(3), 453–469. http://dx.doi.org/10.1177/0038038515583637 (ver. 28.11.2023).

Bhargava, R. (2023). Teaching Data That Matters: History and Practice. In J. E. Raffaghelli & A. Sangrà (Eds.), *Data Cultures in Higher Education: Emergent Practices and the Challenge Ahead* (pp. 267–292). Springer International Publishing.

Biesta, G. (2020). What constitutes the good of education? Reflections on the possibility of educational critique. *Educational Philosophy and Theory*, *52*(10), 1023–1027. http://dx.doi.org/10.1080/00131857.2020.1723468 (ver. 28.11.2023).

Biesta, G., & Tedder, M. (2007). Agency and learning in the lifecourse: Towards an ecological perspective. *Studies in the Education of Adults*, *39*(2), 132–149.

Blaschke, L. M. (2012). Heutagogy and lifelong learning: A review of heutagogical practice and self determined learning. *International Review of Research in Open and Distance Learning*, *13*(1), 56–71. http://dx.doi.org/10.1016/j.system.2004.09.015 (ver. 28.11.2023).

Boychuk, M., Cousins, M., Lloyd, A., & MacKeigan, C. (2016). Do We need Data Literacy? Public Perceptions Regarding Canada's Open Data Initiative. *Dalhousie Journal of Interdisciplinary Management*, *12*(1). http://dx.doi.org/10.5931/DJIM.V12I1.6449 (ver. 28.11.2023).

Braun, V. & Clarke, V. (2021). *Thematic Analysis: A practical guide*. Sage.

Carey, K. (2015). *The End of College: Creating the Future of Learning and the University of Everywhere*. Penguin Publishing Group.

Coughlan, T. (2019). The use of open data as a material for learning. *Educational Technology Research and Development*, 1–28. http://dx.doi.org/10.1007/s11423-019-09706-y (ver. 28.11.2023).

Dencik, L., & Sanchez-Monedero, J. (2022). Data justice. *Internet Policy Review*, *11*(1). http://dx.doi.org/10.14763/2022.1.1615 (ver. 28.11.2023).

D'Ignazio, C., & Bhargava, R. (2015). *Approaches to Building Big Data Literacy*. Bloomberg Data for Good Exchange. New York, NY, USA.

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. MIT Press. http://dx.doi.org/10.7551/mitpress/11805.001.0001 (ver. 28.11.2023).

Flik, U. (2004). Triangulation in Qualitative Research. In U. Flick, E. von Kardoff, & S. Ines (Eds.), *A Companion to Qualitative Research* (p. 432). SAGE.

Foley, G. (2020). *Dimensions of Adult Learning: Adult education and training in a global era*. Routledge.

Gil-Garcia, J. R., Gasco-Hernandez, M., & Pardo, T. A. (2020). Beyond Transparency, Participation, and Collaboration? A Reflection on the Dimensions of Open Government. *Public Performance & Management Review*, *43*(3), 483–502.

Hobbs, R. (2020). Propaganda in an Age of Algorithmic Personalization: Expanding Literacy Research and Practice. *Reading Research Quarterly*, *55*(3), 521–533.

Kimmons, R. (2022). Mixed Methods: How does one go about doing good mixed methods research? In *Education Research.* EdTech Books.

Kimmons, R., & Veletsianos, G. (2018). Public Internet Data Mining Methods in Instructional Design, Educational Technology, and Online Learning Research. *TechTrends*, *62*(5), 492–500. http://dx.doi.org/10.1007/s11528-018-0307-4 (ver. 28.11.2023).

Lareau, A., & Weininger, E. B. (2003). Cultural capital in educational research: A critical assessment. *Theory and Society*, *32*, 567–606.

Loría-Solano, E., Guitert Catasús, M., & Raffaghelli, J. E. (2023). (Open) Data literacy: Which relationships with open data adoption? A systematic review of the literature. Italian Journal of Educational Technology. Accepted Manuscript Online. http://dx.doi.org/10.17471/24994324/1303 (ver. 28.11.2023).

Maybee, C., & Zilinski, L. (2015). Data informed learning: A next phase data literacy framework for higher education. *Proceedings of the Association for Information Science and Technology*, *52*(1), 1–4. http://dx.doi.org/10.1002/pra2.2015.1450520100108 (ver. 28.11.2023).

Okoko, J. M., Tunison, S., & Walker, K. D.. (2023). *Varieties of Qualitative Research Methods*. Springer.

Pangrazio, L., & Selwyn, N. (2019). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*, *21*(2), 419–437. http://dx.doi.org/10.1177/1461444818799523 (ver. 28.11.2023).

Pybus, J., Coté, M., & Blanke, T. (2015). Hacking the social life of Big Data. *Big Data and Society*, *2*(2). http://dx.doi.org/10.1177/2053951715616649 (ver. 28.11.2023).

Quarati, A., & Raffaghelli, J. E. (2020). Do researchers use open research data? Exploring the relationships between usage trends and metadata quality across scientific disciplines from the Figshare case. *Journal of Information Science*, *48*(4), 423–448. http://dx.doi.org/10.1177/0165551520961048 (ver. 28.11.2023).

Raffaghelli, J. E. (2018). Open Data for Learning: A case study in Higher Education. In A. Volungeviciene & A. Szűcs (Eds.), *Exploring the Micro, Meso and Macro Navigating between dimensions in the digital learning landscape. Proceedings of the EDEN Annual Conference, 2018* (pp. 178–190). European Distance and E-Learning Network.

Raffaghelli, J. E. (2020). Is Data Literacy a Catalyst of Social Justice? A Response from Nine Data Literacy Initiatives in Higher Education. *Education Sciences*, *10*(9), 233. http://dx.doi.org/10.3390/educsci10090233 (ver. 28.11.2023).

Raffaghelli, J. E. (2023). Pathways for Social Justice in the Datafied Society: Reconsidering the educational response. *Media Education*, *14*(1), Article 1. http://dx.doi.org/10.36253/me-13383 (ver. 28.11.2023).

Ruijer, E., Grimmelikhuijsen, S., van den Berg, J., Meijer, A., Quarati, A., Raffaghelli, J. E., Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B. O. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., … Rahm, E. (2016). On the barriers for local government releasing open data. *Government Information Quarterly*, *31*(1), 267. http://dx.doi.org/10.3390/info9110267 (ver. 28.11.2023).

Sander, I. (2020). Critical big data literacy tools—Engaging citizens and promoting empowered internet usage. *Data & Policy*, *2*. http://dx.doi.org/10.1017/dap.2020.5 (ver. 28.11.2023).

Santos-Hermosa, G., Quarati, A., Loría-Soriano, E., & Raffaghelli, J. E. (2023). Why Does Open Data Get Underused? A Focus on the Role of (Open) Data Literacy. In J. E. Raffaghelli & A. Sangrà (Eds.), *Data Cultures in Higher Education: Emergent Practices and the Challenge Ahead* (pp. 145–177). Springer International Publishing.

Smolnikova, M. (2020). Next Step: Data Literacy Measurement. *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 234–240. http://dx.doi.org/10.5220/0010146402340240 (ver. 28.11.2023).

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance and Society*, *12*(2), 197–208. http://dx.doi.org/10.24908/ss.v12i2.4776 (ver. 28.11.2023).

Zuiderwijk, A., Janssen, M., & Dwivedi, Y. K. (2015). Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology. *Government Information Quarterly*, *32*(4), 429–440. http://dx.doi.org/10.1016/j.giq.2015.09.005 (ver. 28.11.2023).