ODECO

Towards a sustainable Open Data ECOsystem

D2.1
Open data user needs: seven flavours

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955569.
The opinions expressed in this document reflect only the author's view and in no way reflect the European Commission's opinions. The European Commission is not responsible for any use that may be made of the information it contains.

Project Acronym	ODECO
Project Title	Towards a sustainable Open Data ECOsystem
Grant Agreement No.	955569
Start date of Project	01-10-2021
Duration of the Project	48 months
Deliverable Number	D2.1
Deliverable Title	Open data user needs: seven flavours
Dissemination Level	Public
Deliverable Leader	TU Delft (TUD)
Submission Date	31-05-2023
Editor	Ingrid Mulder, TUD
Co-author(s)	Davide Di Staso, María Elena López Reyes, Georgios Papageorgiou, Alejandra Celis Vargas, Liubov Pilshchikova, Caterina Santoro, Héctor Ochoa Ortiz, Umair Ahmed, Ahmad Ashraf Ahmad Shaharudin

Version #	Date	Description (Section, page number)	Author & Organisation
V0.0	29-03-2023	Plenary workshop during training week on understanding users (needs) & providers (needs) from multiple perspectives. Discussion with all ESRs facilitated on a MIRO board	Ingrid Mulder, TUD, Glenn Vancauwenberghe, KU Leuven, Silvia Cazacu, KU Leuven; All ESRs participated in the workshop and created a shared understanding of the various roles of users and providers in a data ecosystem
V0.1	28-04-2023	First setup with contributions from ESR perspectives	ESR01: Davide Di Staso, TUD; ESR06: María Elena López Reyes, AAU; ESR09: Georgios Papageorgiou, ESR10: Alejandra Celis Vargas, AAU; Liubov Pilshchikova, TUD; ESR12: Caterina Santoro, KU Leuven; ESR13: Héctor Ochoa Ortiz, UNICAM; ESR14: Umair Ahmed, UNICAM; ESR15: Ahmad Ashraf Ahmad Shaharudin, TUD
V0.1a	03-05-2023	Feedback and alignment meeting to discuss structure and contribution of the deliverable	Ingrid Mulder, TU Delft; Francisco J. Lopez-Pellicer, UNIZAR; Andrea Polini, UNICAM; Bastiaan van Loenen, TU Delft together with all ESR contributors
V0.1b	08-05-2023	Restructure of Introduction and Conclusions	ESR15: Ahmad Ashraf Ahmad Shaharudin, TUD; ESR12: Caterina Santoro, KU Leuven.

Version #	Date	Description (Section, page number)	Author & Organisation
	10-05-2023	Collaborate peer-review meeting	Ingrid Mulder, TU Delft; Francisco J. Lopez-Pellicer, UNIZAR; Andrea Polini, UNICAM; Bastiaan van Loenen, TU Delft together with all ESR contributors
V0.2	17-05-2023	Second draft. Processed feedback obtained from the peer review meeting on May 10, 2023	ESR01: Davide Di Staso, TUD; ESR06: María Elena López Reyes, AAU; ESR09: Georgios Papageorgiou, ESR10: Alejandra Celis Vargas, AAU; ESR11: Liubov Pilshchikova, TUD; ESR12: Caterina Santoro, KU Leuven; ESR13: Héctor Ochoa Ortiz, UNICAM; ESR14: Umair Ahmed, UNICAM; ESR15: Ahmad Ashraf Ahmad Shaharudin, TUD
V0.2a	26-05-2023	Review	Francisco J. Lopez-Pellicer, UNIZAR; Andrea Polini, UNICAM; Loukis Euripides, UAEGEAN
V0.3	30-05-2023	Third draft. Integration of feedback received from QA peer review	Ingrid Mulder, TUD
V0.4	30-05-2023	Final report	Ingrid Mulder, TUD
V0.5	31-05-2023	Approval	Bastiaan van Loenen, TUD
V1.0	31-05-2023	Final editing	Danitsja van Heusden, TUD

Table of Contents

List of figures

List of tables

Abbreviations

AI	Artificial Intelligence
CARE	Collective benefit, Authority to control, Responsibility, Ethics
CBO	Community-Based Organisations
D	Deliverable
EC	European Commission
ESR	Early Stage Researcher
EU	European Union
M	Milestone
NGO	Non-Government Organisation
NPO	Non-Profit Organisation
OD	Open Data
ODECO	Open Data ECOsystem
WoS	Web of Science
WP	Work Package

Nr	Partner	Partner short name	Country
Beneficiary			
1	Technische Universiteit Delft	TU Delft	Netherlands
2	Katholieke Universiteit Leuven	KUL	Belgium
3	Centre National de la Recherche Scientifique	CNRS	France
4	Universidad de Zaragoza	UNIZAR	Spain
5	Panepistimio Aigaiou	UAEGEAN	Greece
6	Aalborg Universitet	AAU	Denmark
7	Università degli Studi di Camerino	UNICAM	Italy
8	Farosnet S.A.	FAROSNET S.A.	Greece
Partner organisations			
1	7EDATA	7EDATA	Spain
2	Digitaal Vlaanderen	DV	Belgium
3	City of Copenhagen	COP	Denmark
4	City of Rotterdam	RDAM	Netherlands
5	CoC Playful Minds	CoC	Denmark
6	Derilinx	DERI	Ireland
7	ESRI	ESRI	Netherlands
8	Maggioli S.p.A	MAG	Italy
9	National Centre of Geographic Information	CNIG	Spain
10	Open Knowledge Belgium	OKB	Belgium
11	Spatineo	SPATI	Finland
12	SWECO	SWECO	Netherlands
13	The government lab	GLAB	
14	Agency for Data Supply and Infrastructure	ADSI	

1. Introduction

1.1. Background

The European Commission (EC) made a projection that by 2025, the net worth of the European Union (EU)'s data economy is predicted to be €829 billion and is going to increase significantly over the following few years (Kumpula-Natri, 2021). Open data, which is data that is made available free of charge, with an open license, and in an open, machine-readable format, is expected to generate even more value (European Commission, 2011). It can enhance the effectiveness and efficiency of public services, increase institutional accountability, boost citizen participation, accelerate scientific progress, and foster the creation of other economic and social values (Hossain et al., 2016; Janssen et al., 2012; Zhu et al., 2019). In line with this enormous potential, the EC has published the directive on open data and the re-use of public sector information (Open Data Directive) that entered into force on 16 July 2019.

Despite the benefits of open data, its full potential has not yet been realised due to shortcomings in the current open data systems. Current developments in the field of open data are characterised as highly fragmented. Experts suggest that collaboration and coordination are necessary and that government efforts alone are not enough to make open data available and valuable (Harrison et al., 2012; Pollock, 2011; Zuiderwijk et al., 2014). Open data is often developed in different domains in isolation and with little involvement of potential users, resulting in approaches that significantly limit open data reusability for users. The current "one-way street" and top-down approach to open data is not ideal, as it does not consider the user needs and may lead to limited use and value generation. Therefore, a more collaborative and user-centric approach is essential for the optimal use and value of open data (Pollock, 2011; Van Loenen et al., 2021). To address this, many researchers have advocated for the "open data ecosystem" approach (Davies, 2011; Poikola et al., 2011; Pollock, 2011; van Loenen et al., 2018; van Loenen et al., 2021).

While this concept is still developing, at the heart of the ecosystem metaphor is the recognition of the diversity of data users' needs that emphasizes the importance of not only data suppliers but also users (Davies & Edwards, 2012; Pollock, 2011; van Loenen et al., 2021). To achieve this, an important step is shifting from a supplier-driven to a user-driven perspective. ODECO is working towards creating a user-driven ecosystem to better match the demand and supply of open data. Previous studies have suggested ways to improve the use of open data (Olausson, 2016; Ruijer et al., 2017; Susha et al., 2015), but more research is needed to understand the needs of different user types throughout the open data lifecycle[1]. Only then can appropriate governance and technical measures be implemented.

1.2. Role of this deliverable in the ODECO project

In this deliverable, we aim to address the first knowledge gap of ODECO, which is understanding the diverse needs of different types of users (Task 2.1). This will help us create a more user-driven open data ecosystem. At the ODECO project's proposal stage, seven major categories of open data users were identified: non-specialist data users, government, intermediaries & companies, journalists, students, non-government organisations (NGOs), and artificial users. Hence, the proposed title of this deliverable is "Open data user needs: seven flavours". These seven user types are then translated into nine ESRs in the project, looking into user types, with local and regional/central governments and intermediaries and companies being divided into four separate user types. Hence, the resulting nine types of open data users are studied in the current report.

[1] Open data lifecycle is "the process and practices around handling data, starting from its creation, through the provision of open data to its use by various parties" (Charalabidis et al., 2018b)

The structure of this report is as follows. Developing a user-driven open data ecosystem starts with identifying the needs of users to create value from open data. Chapter 2 elaborates upon the methodology to identify user needs and presents the agreed-upon research approach. The following chapters present the user needs of the user types according to the nine different contexts:

- Chapter 3: Non-specialist data users (ESR1)
- Chapter 4: Local government (ESR6)
- Chapter 5: Journalists (ESR9)
- Chapter 6: Students (ESR10)
- Chapter 7: NGOs (ESR11)
- Chapter 8: Central/regional government (ESR12)
- Chapter 9: Companies (ESR13)
- Chapter 10: Artificial users (ESR14)
- Chapter 11: Open data intermediaries (ESR15)

Chapter 12 concludes with a categorisation of user needs that brings our findings together. We close the report with a research agenda helpful for our ongoing research. However, in the first place, this report is the basis for developing technological and governance measures to satisfy user needs in Task 2.2 and 2.3.

2. Methodology

2.1. Identifying user needs

The first challenge to answering the question of what the needs of open data users are is to define "open data users" and "user needs". A user is anyone who interacts with a system to perform a task, regardless of their level of expertise or familiarity with the technology (Norman, 2013). One way to identify open data users is by grouping them according to characteristics they share, such as their purpose of open data use (e.g., aggregator, enabler, enricher, developer, end-user), the nature of the user (e.g., commercial, government, scientific, citizen), their capabilities (e.g., technical, creative, domain, business skills), their access to resources, and their network (van Loenen, 2018). However, even within these groups of similar characteristics, there may be differences in their open data needs (e.g., small/medium-sized companies versus large companies, citizens with data skills versus citizens without data skills) (van Loenen, 2018). These differences mean that the list of possible open data users is endless, hence identifying them and their needs is a challenging task.

Now, after we have established how diverse open data users and their needs are, another layer of complexity is needed to conceptualise and operationalise "user needs". Table 1 shows various definitions of "user needs" gathered from different fields. However, it may not be possible to produce a single definition of "user needs", especially for a multidisciplinary research endeavour like ODECO. Nevertheless, it is useful to list common notions around "user needs". From the sources in Table 1, user needs are often associated with (i) the user's goals, values, and aspirations and their (ii) current problems and (iii) activities/tasks to achieve them within (iv) their context, space, or environment.

Table 1: Several definitions of user needs

Source	Definition	Field
(Kujala et al., 2001)	"difference between users' goals and the present condition, which is manifested by user problems and possibilities, and the context of use, which includes the characteristics of the intended users, users' present tasks and environment"	Human-computer interaction
(Kujala, 2002)	"problems that hinder users in achieving their goals, or opportunities to improve the likelihood of users' achieving their goals. An important factor affecting on user needs is the context of use"	Human-computer interaction
(Zickler et al., 2009)	"person's wants and necessities with respect to different aspects of independence"	Assistive technology
(Government of UK, 2017)	"the needs that a user has of a service, and which that service must satisfy for the user to get the right outcome for them"	Public administration
(Government of Scotland, n.d.)	"people's goals, values, and aspirations. They are the things people need from a product or service to do something"	Public administration
(Tuunanen, 2021)	"requirements that add value to the user (i.e., the reuser) of a license compliance tool"	Open source
(Heijs, 2022)	"physiological, social or psychological states of the users of an object (e.g., of a building), or activities to reach those states, linked to the process of use of the object, that contribute to the physiological, social or psychological well-being of the users in the process of use"	Built environment

Source	Definition	Field
(Lee et al., 2022)	"results of identifying people's usability and user experience goals, by exploring the problem space, and investigating who the users are and their activities to see what can be improved"	Human-computer interaction

The notions gathered are also consistent with the four-layer model of human needs and aspirations by (Van Der Bijl Brouwer & Dorst, 2014). The four levels are (i) the solution level, which describes what people want or need; (ii) the scenario level, which describes how people want to interact with a solution in a specific context of use; and the (iii) goals and (iv) themes levels that together describe why people want or need a particular solution and scenario (goals describe what people want to achieve within the context of a certain design problem and themes describe the underlying needs and aspirations that can be analysed independently of the context).

2.2. Research approach

Guided by the notions of user needs from the literature, as described in the sub-section, we conduct literature reviews to identify user needs for these user types: non-specialist data users, local government, journalists, students, NGOs, central/regional government, companies, artificial users, and open data intermediaries.

There are a couple of limitations worth mentioning. Firstly, while this report seeks to answer a single research question that is "What are the user needs of open data users?", the diversity of users as well as interpretations of "user needs" means that it is not feasible nor necessary to gather and pay attention to only a single variable/indicator of user needs. In fact, the diverse lenses to capture "user needs", inspired by different disciplines of ESRs, can serve as an exercise to learn from each other on what we miss by looking at "needs" through a certain lens.

Secondly, the endless list of possible open data users and ways of grouping them also means that there are various ways of defining user types. In some instances, it is also more practical to scope certain user types narrowly to identify certain needs that may otherwise be overlooked if the user type is defined broadly, especially among disadvantaged users. It is also worth highlighting that the nine user types studied in ODECO are by no means an exhaustive list of user types.

3. Non-specialist users

3.1. Introduction

"Non-specialist users" are users who lack data skills that allow for the cleaning, formatting, analysis, and visualisation of datasets. However, even without these data skills, non-specialist users can contribute by bringing in "thick data" (Wang, 2016) or contextual knowledge about the domain being studied. Non-specialist users, therefore, fit the role of "user" in the sense that they can benefit from the value of certain open datasets but can also be "providers" when offering their observations and lived experience in the domain of the datasets. This knowledge is essential in order to contextualise the data and appropriately frame problems that open data can address. (boyd & Crawford, 2012) argue that "computer scientists and social scientists both have valuable perspectives to offer" when working with data and that it is unfair to set up hierarchies based on who possesses the most data skills.

To bring the potential of open data to fruition, both specialists (developers) and non-specialists (domain experts) are needed, as the former can provide technical solutions to work with data, and the latter can help to contextualise datasets and insights resulting from the analysis. One environment in which specialist and non-specialist users can collaborate and bring their knowledge together is data hackathons (or datathons), which often involve the reuse of open datasets found online or brought in by partner organisations. These are events where attendees with different backgrounds can interact and collaborate in analysing datasets. The difference in background, skills, and knowledge of specialist and non-specialist users poses a challenge to their successful collaboration and to the fruitful reuse of open datasets.

3.2. Method

Hackathons typically involve technical users and a limited number of non-specialist users (Briscoe & Mulligan, 2014). Literature explicitly discussing the needs of non-specialist users is therefore limited, especially in the specific context of open data hackathons. As a result, we decided to analyse three cases of hackathons from the literature in which non-specialist users worked with open datasets. We review these cases, looking for the challenges to the engagement of non-specialist users, to discuss their needs. We selected cases that provide a detailed account of the hackathon event, and in which open data was used by participants.

3.3. Results

Case 1 - Issue-oriented hackathons (Lodato & DiSalvo, 2016)

Lodato & DiSalvo (2016) describe two cases of issue-oriented hackathons in New York City in 2012: EcoHack3, which focused on environmental issues, and Hack//Meat, which focused on issues related to the meat industry. At these events, open datasets were introduced by partner organisations or by the participants themselves, who brought in data from public reports or from their professional backgrounds. When data was yet to be collected, participants created fake data points to imagine possible future visualisations. For example, the solutions presented at EcoHack3 used geographic, satellite data and data related to the wetlands.

In discussing participants' motivations to attend, Lodato & DiSalvo (2016) noted that participants are driven both by the social nature of these events and by their professional interests. One participant, for example, was interested in connecting with developers who could help her create basic data visualizations that would help her in a professional project.

Lodato & DiSalvo (2016) do not provide a full account of the attendees' background and of the group composition, as this was not the focus of the study. However, there seems to be a tendency for participants to gather based on their skill set. In Eco-Hack 3, participants formed groups based on their technical expertise and often worked separately. One group working on bike sharing split into

three subgroups, with the authors noting that "the split followed skill areas, where, for example, many of the designers, feeling excluded from the technical tasks, worked on information design" (Lodato & DiSalvo, 2016, p. 547). The result was a subgroup "composed primarily of self-identified nontechnical designers" (Lodato & DiSalvo, 2016, p. 551). A similar division based on skills was also observed at Hack//Meat, where the researcher joined a subgroup "composed of three other members, all of whom self-identified as developers" (Lodato & DiSalvo, 2016, p. 547).

While we lack a full breakdown of each team's composition, it seems that, unless specific group formation methods are used, hackathon participants tend to gravitate around other people with a similar skill set and avoid interaction with other groups. As a result, we identify the following user need presented in Table 2.

Table 2: First identified non-specialist user need

	User group	User need
Need 1	Non-specialist	Non-specialist users participating in open data hackathons or similar design engagements need team formation methods that allow for the mixing and interaction with specialist users

Case 2 - Hacking with NPOs (Hou & Wang, 2017)
Hou & Wang (2017) collected data from two "civic data hackathons" held in a Midwestern city in the U.S. in 2015 and 2016. NPOs often lack the internal resources to analyse their own data and realise its potential (ibid.). Data hackathons are one way in which NPOs can get access to data specialists who can help them answer questions by analysing both internal NPO data and open datasets (ibid.). NPOs participating in the two events analysed included local libraries, nature preservation groups, and organisations in the fields of education, literacy, and arts.

Hou & Wang (2017) used surveys to analyse, among other things, the attendees' data expertise and obtained a sample of 40 respondents (the same proportion of men and women). Almost half of the respondents reported having passing data knowledge, 10% reported "expert" knowledge, and 15% reported having no knowledge. Survey results showed that participants were mostly interested in practicing data skills (65%), learning new data skills (60%), helping non-profit organisations (60%), and networking with others (42.5%). However, we are missing a cross-tabulation between data expertise and motivation to attend, which would have provided better insights into the motivations of non-specialist users.

Organisers put in place a specific strategy to facilitate collaboration between NPOs and data volunteers. The strategy consisted in recruiting "client teams", each assigned to a specific NPO, that would act as "brokers" between the NPO and the data volunteers. Specialist data users were present both in the client teams and data volunteer teams; Hou & Wang (2017) noted several interesting tensions in the interaction between specialist and non-specialist users. For example, non-specialist users chose data tools that they were not familiar with to learn how to use them. This is beneficial for individual learning but takes away from the goal of helping NPOs. On a similar note, specialist users spent significant time helping non-specialists in getting up to speed with data analysis tools, taking time away from actual data analysis work. Hou & Wang (2017) found that "good civic data hackathons should address the tensions between benefiting NPOs' data-driven work and helping volunteers learning expectations". Based on this case we define the need for an environment that facilitates learning opportunities for non-specialist users as shown in Table 3.

Table 3: Second identified non-specialist user need

	User group	User need
Need 2	Non-specialist	Non-specialist users participating in open data hackathons or similar design engagements need the freedom and support to learn new data skills

3.4. Conclusions

Non-specialist users participating in open data hackathons or similar design engagements need team formation methods that allow for the mixing and interaction with specialist users. Moreover, they need the freedom and support to learn new data skills. The literature on open data sprint events (hackathons and similar) rarely provides a detailed account of the event structure and scaffolding. Future research should investigate in greater detail the needs of non-specialist users and how these needs can be addressed by different event structures and activities.

4. Local government

4.1. Introduction

Although local governments are often portrayed as open government data suppliers, the existing literature fails to understand their role across the open data processing cycle (Conradie & Choenni, 2014; Dawes et al., 2016; Johnson, 2016; Kleiman et al., 2023); especially if we think of it through a value-creating open data ecosystem perspective (van Loenen et al., 2021). In many cases, local governments are the primary users of this data as they need government data to be open to run their public tasks, such as driving participation in policymaking, identifying societal needs, or improving the delivery and efficiency of public services (Abella et al., 2019; Foulonneau et al., 2014; Walravens et al., 2021). Research conducted in the Netherlands demonstrated that how data is stored, obtained, and used varies in local government departments regarding policies and practices toward data release (Conradie & Choenni, 2014). Following the research approach of this report, we seek to better understand the local government's role within the open data ecosystem by identifying local governments' needs as open data users. In this section, we will refer to local governments' efforts to use open government data as open data initiatives.

4.2. Method

We conducted a literature review on open government data initiatives and local governments' needs to use open government data. Literature reviews are valuable for justifying research approaches, identifying gaps, and supporting theory development (Danson & Arshad, 2014). The review process involved defining the research question, developing a search strategy, and selecting relevant databases and keywords. The following bibliographic databases were searched to explore the literature in the field of OGD: Scopus Database; Web of Science; and EBSCOhost. The search for the studies was conducted from October 14th, 2022, to January 10th, 2023. Iterations of searches and screenings of paper titles with various combinations of keywords were conducted to ensure the identification of the maximum number of papers that meet the inclusion and exclusion criteria.

After performing some initial searches to understand the domain, the following keywords were chosen: "social" OR "public", AND "value" OR "benefit" OR "impact", AND "local" OR "city" OR "municipal" OR "community", AND "open government data." Of 2568 records initially identified, 158 were included after the search was narrowed down to only peer-reviewed records and where the keywords were mentioned in the abstract; 83 records were identified after duplicates were removed. Abstracts of these records were screened for eligibility, and 15 studies were selected for qualitative analysis. The study only includes primary research studies with empirical approaches published bet between 2018 and 2023. We used the four-layer model of human needs and aspirations (Van Der Bijl Brouwer & Dorst, 2014) a theoretical model to survey and analyse the studies. The results are presented in the following section.

4.3. Results

Following the theoretical approach, the following categories were analysed: solutions, scenarios, goals, and themes. After the analysis, four main themes were identified: (1) reliability and meaningfulness of data; (2) communication and coordination across and within open data systems, (3) data protection, representation, and validity; and (4) legal, technical, and operational adaptability and efficiency. These themes encapsulate the needs and aspirations of local governments concerning the use of open data. Table 4 presents the categorization of themes and the references included in each theme. Next, we present the descriptions for each of the identified themes.

Table 44: Needs categories of local governments concerning the use of open data and their respective sources.

Category	References	Count
Reliable and meaningful data	(Gao & Janssen, 2022a; Golub & Lund, 2021; Milojevic-Dupont et al., 2020; Najafabadi & Cronemberger, 2022; Runeson et al., 2021; Schrotter & Hürzeler, 2020; B. Wilson & Cong, 2021a; J. Zhang et al., 2021)	8
Communication and coordination across and within open data systems	(Gao & Janssen, 2020; Hlabano & Van Belle, 2019; Kassen, 2013; Meng & DiSalvo, 2018; Najafabadi & Cronemberger, 2022; Runeson et al., 2021; Schrotter & Hürzeler, 2020; B. Wilson & Cong, 2021; J. Zhang et al., 2021)	9
Legal, technical, and operational adaptability and efficiency	(Cabitza et al., 2020; Cantador et al., 2020; Gao & Janssen, 2022; Golub & Lund, 2021; Hlabano & Van Belle, 2019; Jarke, 2021; Meng & DiSalvo, 2018; Najafabadi & Cronemberger, 2022; Runeson et al., 2021; Schrotter & Hürzeler, 2020; B. Wilson & Cong, 2021; J. Zhang et al., 2021)	12
Data protection, representation, and validity	(Cantador et al., 2020; Gao & Janssen, 2020; Golub & Lund, 2021; Hlabano & Van Belle, 2019; Jarke, 2021; Meng & DiSalvo, 2018; Milojevic-Dupont et al., 2020; Najafabadi & Cronemberger, 2022; Runeson et al., 2021; Schrotter & Hürzeler, 2020; B. Wilson & Cong, 2021)	11

Reliable and meaningful data

When local governments act as implementers of open data initiatives and are tasked to capture, publish, update, and aggregate data, they need to ensure reliable data, which means that the data being used need to be accurate, consistent, and integrated (Gao & Janssen, 2022; Milojevic-Dupont et al., 2020; Najafabadi & Cronemberger, 2022; Runeson et al., 2021; Schrotter & Hürzeler, 2020; B. Wilson & Cong, 2021; J. Zhang et al., 2021). For that to happen, one of the critical needs is having the capacity to capture data users' input (B. Wilson & Cong, 2021; J. Zhang et al., 2021), which can also help governments to be informed about the needs and priorities of non-government users (B. Wilson & Cong, 2021). Some other consideration when publishing and updating data is the ability to keep the timeliness (Gao & Janssen, 2020), contextualise the data (J. Zhang et al., 2021), build large datasets, and enable bulk access (Runeson et al., 2021). Taking care of these aspects will enable the publication of meaningful data that can drive innovation for activities such as machine learning (Runeson et al., 2021) or research (Najafabadi & Cronemberger, 2022).

Communication and coordination across and within open data systems

When local governments act as initiators of open data initiatives, they need proper communication and coordination so that the initiatives have the support and engagement of local businesses and communities and align standards, business rules, and architecture. These will enable data flow and drive innovation and value creation (Hlabano & Van Belle, 2019; Kassen, 2013; Najafabadi & Cronemberger, 2022; B. Wilson & Cong, 2021). It is also necessary to have the economic resources and the appropriate business models to drive the commitment of the different participants (Meng & DiSalvo, 2018; B. Wilson & Cong, 2021). Furthermore, local governments need to coordinate with other participants to ensure robust metadata and satisfy the increase in data demand, for example, when data is used for prediction models (Gao & Janssen, 2020; Runeson et al., 2021; Schrotter & Hürzeler, 2020; B. Wilson & Cong, 2021). Additionally, accessing end-users' feedback is crucial for evaluating and prioritizing the data availability (B. Wilson & Cong, 2021). Finally, as implementers of open data initiatives, local governments need trust and coordination mechanisms so that they can prevent distrust and competition between public and private actors (Runeson et al., 2021; Schrotter & Hürzeler, 2020; J. Zhang et al., 2021).

Legal, technical, and operational adaptability and efficiency

When local governments design open data initiatives, they need to increase legal, technical, and operational efficiency. Some of the considerations to achieve it are the awareness of digital literacy gaps (Jarke, 2021; Meng & DiSalvo, 2018) when designing open data solutions and having the legal and technical tools to be able to integrate different licensing frameworks (Hlabano & Van Belle, 2019; Runeson et al., 2021; Schrotter & Hürzeler, 2020; B. Wilson & Cong, 2021) and technologies (Gao & Janssen, 2022; B. Wilson & Cong, 2021; J. Zhang et al., 2021); for example, enabling the convergence of Building Information Modelling (BIM) and Geographic Information System (GIS) technologies (Runeson et al., 2021; Schrotter & Hürzeler, 2020). The technological considerations should also include upgrading tools and systems (Schrotter & Hürzeler, 2020; J. Zhang et al., 2021), which is necessary to cover the end-user demands. To have enough resources, they also need to sustain the interest of elected officials (Hlabano & Van Belle, 2019; Najafabadi & Cronemberger, 2022; B. Wilson & Cong, 2021). The operational considerations should address time efficiency by, for example, making data publishable through fewer steps (Najafabadi & Cronemberger, 2022) and having the training, specialised teams, and software so that they can automate processes (Gao & Janssen, 2022; B. Wilson & Cong, 2021; J. Zhang et al., 2021) and quickly identify insights primarily when the solutions use crowdsourcing as data collection method (Cabitza et al., 2020; Cantador et al., 2020; Golub & Lund, 2021; Hlabano & Van Belle, 2019; B. Wilson & Cong, 2021). They also need to have enough employees to properly follow up on data requirements, evaluate and prioritise data provisions, and generate evidence of the impact of open data in policymaking (Gao & Janssen, 2022; B. Wilson & Cong, 2021).

Data protection, representation, and validity

When local governments act as initiators of open data initiatives, they need ethical representation and personal data protection mechanisms to prevent pre-existing software biases, potential mistreatment, and misinterpretation of the data (Najafabadi & Cronemberger, 2022). Data protection of personal data and validation of data is particularly important when initiatives include AI technologies to ensure minority groups can access the benefits of open data and prevent harm to individuals, such as social exclusion (Cantador et al., 2020; Gao & Janssen, 2020; Meng & DiSalvo, 2018; Runeson et al., 2021; B. Wilson & Cong, 2021). When local governments publish data, they also need to consider different contexts and languages to ensure the representation of local communities (Golub & Lund, 2021; Jarke, 2021; Meng & DiSalvo, 2018; Milojevic-Dupont et al., 2020; Najafabadi & Cronemberger, 2022; Runeson et al., 2021; B. Wilson & Cong, 2021). The open-data solutions owned by local governments also need to consider biases in data results presentations, for which it is important to offer adaptable communication and visualisation tools (Cantador et al., 2020; Schrotter & Hürzeler, 2020). Finally, when evaluating the impact of open government initiatives, local governments want input from a diverse group of users, not only specialists such as developers, to include them in impact measures (Golub & Lund, 2021; Hlabano & Van Belle, 2019).

4.4. Conclusions

The results show that the roles and needs or desires that local governments can acquire within the open data processing cycle can vary according to the goals to be achieved and the context of the initiative. With this study, the understanding of the role of local municipalities is broadened. More than acting solely as a supplier within the open data ecosystem, local governments tend to act as re-users as they are the initiators, promoters, coordinators, implementers, evaluators, and/or owners of the open data initiatives. According to the results, there are four themes that act as research directions to investigate how local government needs can be achieved: (1) reliability and meaningfulness of data, (2) communication and coordination across and within the open data systems, (3) data protection, representation, and validity, and (4) legal, technical, and operational adaptability, and efficiency. However, there is an opportunity to take more empirical approaches to create more accurate evidence and verify the results. Further research is needed to understand the links within the stages in the open data processing cycle, the roles of local governments, and the needs that arise. This study also has some limitations. The first one is related to the coverage of the study, as the scope of the literature

was narrow; what that means is that the number of studies included in the review is limited and does not consider a long period limiting the validity and generalizability of the results. Finally, more descriptive methods could also help to understand better the context in which local governments act from different roles; for example, by focusing on the domains of application or describing the maturity level of the open government data ecosystem in a particular context.

5. Data journalism

5.1. Introduction

The importance of data in journalism is becoming increasingly prominent over the years. There have been great steps in the adoption of data-related technologies in journalism since Philip Meyer established Computer-Assisted Reporting (CAR) in the 1960s (Gray et al., 2012). Since that time, the use of data in journalism has evolved, with the contemporary term defined as data journalism (Rogers, 2008). This new trend has some significant differences from CAR. While CAR is focused on the collection and analysis of data, data journalism covers all the stages of the process, from the collection and the writing of the article to the visualisation of the data (Veglis & Bratsas, 2017). As journalism was evolving, there were also developments in the field of data. The beginning of the open data movement is considered the publication "On the Full and Open Exchange of scientific data" (*On the Full and Open Exchange of Scientific Data*, 1995) when the term open data was quoted for the first time. Since then, the open data movement has grown, and governments have started opening their data to increase transparency, accountability and innovation (Publications Office of the European Union, 2022).

Open data movement and data journalism have the same objectives at heart. They both aim to increase transparency and accountability. As the open data movement provides more information, journalists can use them and not only increase transparency and accountability but also prove the value of open data. That creates a synergy that should be promoted as it is beneficial to our society. Therefore, it is important to identify the needs and requirements journalists must guide the policies and the technological intonation in the field of open data to our collective advantage. To discover these needs and requirements, a research question was formulated: *What are the most significant user needs and requirements that are encountered in published papers?*

5.2. Methods

The method used for the analysis of the literature and to identify the topics is a systematic literature review (SLR) (Okoli, 2015; Xiao & Watson, 2019). The initial results from the search for abstract and keywords in four scientific databases, IEEE Xplore, Web of Science, Science Direct, and Scopus, for user needs and requirements were not sufficient and, therefore, we broadened the research by removing the terms "User needs" and "user requirements" and we deducted them during the analysis of the papers. The keywords that were finally selected are divided into the two main focuses of the research, open data, and journalism. To identify the journalistic aspect of the publications, "Journalism", "journalists", and "journalist" have been used. Other keywords, such as "media" or "reporting", were deemed unsuitable since the results were not related to journalism.

For the open data, the term "Open government data" was also used along with "open data", although it is a subset in the domain of open data. The reason for the use of it is that since the terms were quoted in the search, the databases were returning only exact matches. Therefore, during the use of the "open data" keyword, all the publications with the term "open government data" were not included. The goal of the inclusion and exclusion criteria is to ensure that only publications that are relevant to the topic of the research will be analysed. The criteria in this research are making sure that the main focus of the literature is the use of open data in journalism.

For the SLR, four scientific databases were searched to discover publications that are relevant to the research questions. Based on the selected keywords, a query for each database was formed and after the use of the queries in the databases, 131 publications were extracted. The next step was to remove the duplicates that were 34 and to filter out the ones that were not in the English language. After the language filtering, we ended up with 82 publications, there were excluded eleven that were in Spanish, two in Portuguese, one in Turkish and one in French.

During the screening, based on the titles and the abstracts of the papers, the remaining exclusion criteria were applied to the 82 papers we ended up with 45. From the initial 82 papers, two records were excluded for not having access to them, five for not being research papers, and 30 for not being focused on open data or journalism. The next step was to identify the papers in which it was possible the identification of user needs and requirements, and the result was 36 publications.

5.3. Results

To organise the user needs that emerged from the analysis of the papers, we categorise them into four distinct steps. These steps are part of open data and data journalism life cycles. These life cycles are models describing the process of handling data from creation to consumption and are described in detail by open data (Charalabidis et al., 2018) and data journalism (Gray et al., 2012) experts. Four categories are Discoverability, Quality, Analysis and Communication. Although these categories were based on the life cycle of open data and data journalism, they were not enough to accommodate all the user needs that were discovered, so we had to add two more categories: Skills and Ethics. The skills incorporate the need for journalists to be accustomed to new technologies so they can effectively use open data sources. Ethics are about the need for assurance that the use of open data cannot harm the citizens. The research results are presented in detail below in the corresponding paragraph of each need. Table 5 presents the needs with their literature source and the number of times each need was encountered.

Discoverability

Most of the publications that mention the need for improved discoverability of data are technological tools that provide the functionality to explore datasets, but they can be split into two distinct categories with different needs. In the first category, they exist tools that are used for investigative journalism, these are designed to search data sets and topics (Böhm et al., 2010; Klímek et al., 2018; Paley et al., 2021). The second category is used for immediate journalism, which focuses on searching open data sources for new information as fast as possible (Gottron et al., 2015; Ocaña et al., 2021). Finally, two publications are about the difficulty journalists encounter in the discovery of open data (Bozsik et al., 2022; Martin et al., 2022).

Quality

There were also identified publications that mention the need for quality. There are cases where the need for better quality data is just mentioned vaguely (Porlezza & Splendore, 2019), but also more specific needs for quality are detected, like datasets that require cleaning or are missing data (Bozsik et al., 2022). Also, there are mentioned cases that do not have to do with the technical quality of the data, cases where the journalists do not trust the released data sets (Camaj et al., 2022) or they find them to be of low importance for their work (Faini & Palmirani, 2016).

Analysis

Interestingly, all the publications that mention the analysis of data are focused on the processing of big datasets. Of the five publications that are focused on data analysis, four advocate the use of technical tools to help journalists with their work (Andrews & Da Silva, 2013; Le Borgne et al., 2016; Sandoval-Martín & La-Rosa, 2018; Shehu et al., 2016). The fifth publication supports crowdsourcing and the use of games to make the public participate in the data analysis process (Handler & Ferrer Conill, 2016). From that finding, we identify a need for a way to handle large volumes of information.

Communication

The communication need is specialised since the topic of the research is journalism, and there is a great need for the users to be able to convey their findings in an interesting and understandable way for the public. In the publications, there is a great focus on visualisations and compelling storytelling. In many publications, the introduced need is to find a way to present complicated data in a form that can be easily understandable (Rind et al., 2016; Windhager et al., 2016). Another important need is for

compelling storytelling. Very interesting is the case of the NHS winter crisis (Lawson, 2022), where the published articles did not have the expected impact on the public although they were accompanied by extensive data, the main problem was that they were not accommodated by interesting personal stories. So, the outcome of the communications needs is not only the adoption of technological tools but also proper storytelling techniques that can elevate the published articles.

Skills

The lack of digital skills of journalists and their limitation in the use of the available open data sources is a theme that is mentioned in most of the literature. Although there are some publications that cover this skill deficiency in the overall process of journalism, from the collection of data for the publication of the articles and therefore are mentioned in a separate category of needs. In the literature, there were mentioned the cases of hackathons (Boyles, 2020) and the involvement of stakeholders (Kassen, 2018a) from other professions to cover this lack of digital skills. On the other hand, there are presented cases of educational curriculums for journalists that are training them in data journalism and the use of open data (Radchenko & Sakoyan, 2016; Splendore et al., 2016). Furthermore, there are publications that cover specific topics in the skills that journalists need. There are especially mentions of the lack of data analysis skills (Baack, 2018; Gray et al., 2018; Ridgway, 2016) but also of low coding tools (Petricek, 2022) that are designed to bypass this problem. Finally, cases where the lack of digital skills is an impediment to the adoption of new technologies by journalists were encountered (Berntzen et al., 2019; Tabary et al., 2016).

Ethics

The ethical risk that was discovered is the possibility of de-anonymisation of the published data (Bozsik et al., 2022; Krotoski, 2012). Although the datasets used by journalists are anonymised as are all the open data datasets, the produced synthesis of them in a new dataset may lead to the possibility that individuals can be identified. Although any incident of de-anonymisation that could expose the identity of any individual has never occurred, it is a great fear for journalists. If an incident like that occurs, it can have severe consequences, especially when the data sets contain sensitive and personal data like health-related information. Therefore, there is a need for stronger anonymisation tools and techniques that journalists can use to avoid that risk.

Table 55: Needs of journalists identified in the literature

Category	Papers	Count
Discoverability	(Böhm et al., 2010; Bozsik et al., 2022; Gottron et al., 2015; Klímek et al., 2018; Martin et al., 2022; Ocaña et al., 2021; Paley et al., 2021)	7
Quality	(Bozsik et al., 2022; Camaj et al., 2022; Porlezza & Splendore, 2019)	3
Analysis	(Andrews & Da Silva, 2013; Handler & Ferrer Conill, 2016; Le Borgne et al., 2016, 2016; Sandoval-Martín & La-Rosa, 2018)	5
Communication	(Araújo, 2019; Brolcháin et al., 2017; Evéquoz & Castanheiro, 2019; Gupta et al., 2016; Lawson, 2022; Rind et al., 2016; Smith, 2016; Windhager et al., 2016)	8
Skills	(Baack, 2018; Berntzen et al., 2019; Boyles, 2020; Gray et al., 2018; Kassen, 2018a; Petricek, 2022; Radchenko & Sakoyan, 2016; Ridgway, 2016a; Splendore et al., 2016; Tabary et al., 2016)	10
Ethics	(Bozsik et al., 2022; Krotoski, 2012)	2

5.4. Conclusions

In this research, there were identified six different user need categories, Discoverability, Quality, Analysis, Communication, Skills, and Ethics. The most prominent category is Skills since it covers the whole spectrum of the data journalism process. The category with the next mentions is Communication, which is interesting since it implies that the focus for the journalists is the presentation of the story and not the data themselves. Another interesting argument on the

Communication category is that journalists must acquire experience in this new type of reporting so they can combine visualisations with compelling storytelling. A special mention must be made for the Analysis since all the publications were focused on the need to be able to analyse big datasets. This is something that could be an interesting research question in the future. Additionally, we found that there was a fear that open data journalism could unwillingly expose the private information of citizens.

Another interesting finding that was discovered during this research, and it is not related to the user needs, is that the lack of skills is an impediment to the adoption of new technologies and, therefore, new data sources by the journalists. This is a discovery that can be evolved into an interesting research topic since it will explain why data journalism is not so popular during a time when there is an abundance of available data sources. Finally, we should mention that this is ongoing research, and although some interesting findings have been detected, we must be cautious when drawing conclusions. The presented results are preliminary, and more analysis of the literature is needed. After that, interviews with experts in the field of journalism must be conducted.

6. Students

6.1. Introduction

Open Data Education is gaining relevance for straightening and training a literate community that can further benefit from Open Data (Cook et al., 2018). Therefore, the insertion of Open Data in educational systems becomes increasingly important in the Open Data movement. For example, the International Open Data Chapter suggests engaging with schools to ensure the inclusiveness of Open Data in society (International Open Data Charter, 2015).

Although Open Data Education might help to stimulate bigger shifts towards Open Data ecosystems (van Loenen et al., 2021a), Open Data is normally created and released without considering its possible use for educational purposes (Coughlan, 2020). Open Data Education should consider the challenges faced by non-data experts such as students and teachers who participate in formal or hybrid (formal/informal) learning environments. For example, several researchers have highlighted that non-data expert users might face barriers associated with the complexity of handling the data and participation in the open-data process (Janssen et al., 2012). Uncovering the student's needs might be the first step in a journey for enhancing the usage of Open Data as an educational resource.

Open Data Education

The novel field of Open Data Education has the potential of empowering students with digital skills and critical thinking through work directly with real facts (Celis Vargas & Magnussen, 2022). Open Data as an educational resource can contribute to create meaningful learning experiences by countering current flaws in educational systems such as the gap between classroom activities and real-life, and the lack of students and teachers' motivation (Cook et al., 2018; Coughlan, 2020; Saddiqa, Rasmussen, et al., 2019; Wolff, Gooch, et al., 2016). Traditional educational models are criticised for their inability to develop essential skills required for civic engagement and the labour market (Elisa Raffaghelli, 2020; Saddiqa et al., 2021a). Literature in education suggests the need for promoting abilities to cope with new socio-cultural challenges and adapt to changes in technology, and a data-driven society (Atenas et al., 2015; Coughlan, 2020; Wolff, Gooch, et al., 2016).

Current experiments in Open Data Education involve students engaging with Open Data in several ways. Firstly, by directly using open datasets as resources for learning subjects such as geography, history or statistics in formal learning activities (Atenas et al., 2015; Coughlan, 2020). Secondly, by engaging in courses or curriculums where Open Data is the central topic of learning (Dermentzi et al., 2022; Palova & Vejacka, 2022). Thirdly, using Open Data to understand local issues and create solutions in informal learning environments associated with local governments (Davis & Shneyer, 2020). Differences between these three approaches are related to learning objectives and educational levels.

6.2. Methods

A literature review was conducted to uncover students' needs in the field of Open Data Education. According to Grant & Booth (2009), the literature review method seeks to identify what has been accomplished, allowing for consolidation and for gaps identification. Uncovering students; needs in current literature allows building on previous work to pursue the research goal of developing tools and methods enhancing the usage of Open Data in elementary schools.

Considering the Open Data and Education domains, keywords such as *Open Data, Open Datasets, School, Classroom*, and *Educational resource* were defined to conduct iterative searches in the SCOPUS database. Searches with various combinations of keywords were followed by the screening of titles and abstracts to identify relevant studies, such as articles, conference papers, and book chapters in the Open Data Education field. According to the research question: *What are the student's needs for using Open Data?* initial searches considered words such as "needs" and "users", however, the

results did not provide insight from a student's perspective. Therefore, an abductive reasoning process was adopted to form hypotheses by discovering new concepts, ideas, or explanations (Rambaree, 2018), allowing for inferring students' needs.

6.3. Results

Students are non-expert data users. Therefore, their needs for successfully using Open Data in meaningful learning activities might be related to other non-specialist user groups, such as citizens. However, specific students' needs consider learning objectives and educational levels. Students are people who engage in a formal learning activity. According to Beetham (2019), a learning activity involves other participants, such as teachers, a learning environment including tools, and learning outcomes. Allocating learners at the centre, the following figure provides an analysis framework for uncovering the students' needs.

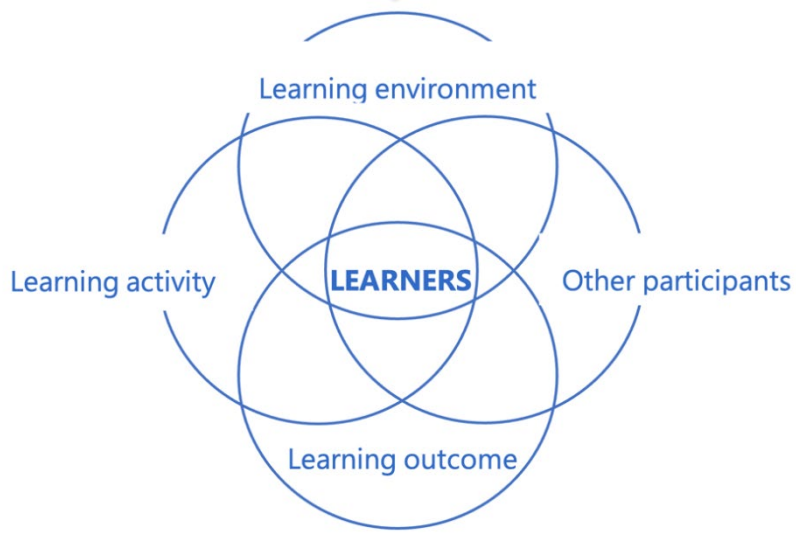

Figure 1: Analysis framework based on Beetham's Learning Activities Design Approach (Beetham, 2019)

Needs related to the learning environment

Learning environments might involve formal or informal settings and certain tools or artifacts which determine different level of interaction.

- *Interception of formal and informal learning environments:*
  On the one hand, formal education is related to primary, elementary, high school, undergraduate programmes and masters and doctorates. Open Data Education is presented as part of the curriculum first, as a specific course, or as part of regular courses such as mathematics and geography, among others. On the other hand, informal education is framed in informal educational or learning environments such as libraries, profit or non-profit, grassroots, and public organisations, and research institutions. For example, Schools of Data initiatives are mainly embodied in courses for a broad public (Dander & Macgilchrist, 2022). There is a need to integrate different levels of formality. Formal/Informal education is related to local government or community-based initiatives aimed at increasing civic participation. Their implementation strategy focuses on students.
- *Concrete tools for students and educators:*
  Students and educators face barriers such as the concept of Open Data being highly abstract and the need for customised hands-on open data collection, interpretation and exploitation activities (Atenas et al., 2015; Coughlan, 2020; Saddiqa et al., 2021a). Educators using Open Data as an educational resource have been facing problems such as finding the right datasets, processing data before giving it to students, and finding sources to ensure the openness of data (Coughlan,

2020). The literature shows the development of experiments on tools for countering barriers and supporting students in using Open Data. Firstly, the authors focused on identifying open datasets for school subjects such as maths, science, and geography through mining techniques, interfaces and online communities (Chicaiza et al., 2017; Saddiqa et al., 2021b; Vallejo-Figueroa et al., 2018). Secondly, the authors presented experiments with tools facilitating the use of Open Data in the classroom for the collection of own local data and data visualisation (Badioze Zaman et al., 2021; Prodromou, 2017; Saddiqa, Kirikova, et al., 2019).

Needs related to other participants

Most learning involves interaction with other participants. These other participants interact with the learners according to different kinds of learning, such as associative, constructive, or situated.

- *Connecting to actors in an Open Data Ecosystem:*
  Students need an Open Data ecosystem which can connect them with data experts, real-world organisations, and other actors. For example, an Open Data Ecosystem might support students to connect to data and problem owners, provide teachers with proper tools to lead learning activities and help school administrators to define guidelines (Radchenko & Sakoyan, 2014; Selwyn et al., 2017). Furthermore, (Saddiqa et al., 2021a) suggest the importance of creating a community of educators and students to share their own datasets, experience and tools.

Needs related to learning outcome

A learning outcome might be seen as an identifiable change in the learner. It might differ according to different kinds of learning.

- *Engaging as active citizens in an Open Data ecosystem:*
  Current literature elaborates on the relevance of Open Data, not just in providing information about reality but in giving input for transforming it. For example, Saddiqa, Rasmussen, et al. (2019) experimented with using local Open Data in the classroom, firstly, to help students understand real facts and, secondly, to come up with ideas to improve their communities. In a competency-based education for active citizenship, students might adopt different roles from an ecosystem perspective, such as Open Data re-users or providers. However, they should be integrated.

Needs related to the learning activity

A learning activity determines the interaction between learners with other people, using certain tools and resources, oriented towards a specific outcome.

- *Developing skills and competencies for understanding and using Open Data:*
  The appropriate skills should allow users not only to use, modify, and share available Open Data (Conradie & Choenni, 2014; Kassen, 2013; Prieto et al., 2012; Shadbolt et al., 2012) but also to understand what kind of perspectives it opens (Van Loenen et al., 2021). For example, Open Data users usually need to be familiar with dataset formats, statistics, text processing software, programming languages or interfaces (Ridgway, 2016b). These technical abilities are often associated with Data Literacy (Van Audenhove et al., 2020; Wolff, Cavero Montaner, et al., 2016). However, the openness of Open Data might need the consideration of skills and competencies, which some authors relate to 21st century skills (Romero et al., 2015). Furthermore, Saddiqa, Kirikova, et al. (2019) and Zapata & Santana (2015) elaborate on the importance of developing skills and competencies, such as the ability to understand local and global issues and critical and scientific thinking.

- *Engaging in meaningful learning experiences:*
  Open Data was used to increase engagement and motivation among students by extending teaching outside the classroom. Activities such as engaging with local settings by collecting data

in real time or working with local data proved to increase students' motivation. Furthermore, games and gamification helped to achieve this goal. For example, open data games helped to collect environmental data and to address local environmental matters in the classroom (Dickinson et al., 2015; Vargianniti & Karpouzis, 2020) (Dickinson et al., 2015; Siriaraya et al., 2018). Open data games motivated students to navigate their city, collect data and discuss their cultural heritage (Chiotaki & Karpouzis, 2020).

6.4. Conclusions

Open Data Education contributes to creating complex learning ecosystems where open data is blended with civic engagement (Elisa Raffaghelli, 2020). Open Data Literacy drives learning activities towards a higher engagement with real-world settings and real-world data. Which contributes to a competence-based education towards community capacity building (Jaskiewicz et al., 2019) and learners engaged in communities that can co-developed capabilities (Bertot et al., 2014).

Future research is needed to gain deeper knowledge about the values, needs and motivations of students in more specific educational contexts. For example, computer science undergraduate students and elementary school students might have different motivations for using Open Data as an educational resource. Therefore, more specific needs might arise regarding the use of Open Data.

In education, the successful integration and adoption of Open Data to solve real-world problems require not just defining the open data skills but also the tools and active learning methodologies (Romero et al., 2015; Wolff, Cavero Montaner, et al., 2016).

Even though current literature in the field of Open Data education is little, the literature is exponentially increasing. Limitations of the current study are related to the novelty of the field. Furthermore, the literature presented the needs from the teacher's perspective. Therefore, new studies and research methods might provide a deeper understanding of latent students' needs.

7. Non-governmental organisations

7.1. Introduction

Non-governmental organisations (NGOs) and non-profit organisations (NPOs) are terms that are often used interchangeably in the literature (Chang, 2005), including open data-related research. The term NGO can include community-based organisations (CBOs) that serve a specific community in a local geographic area, national organisations that operate in developing countries, and international organisations that have headquarters in developed countries and operate in more than one developing country (Chang, 2005). Those organisations are described as non-profit with an elected board of directors and guided by a strategic plan developed in consultation with community stakeholders (Wilson et al., 2010). Hence, NPOs and NGOs are also used interchangeably in this report. Given the NGOs' focus on societal issues and their resolution, they can positively impact society with the use of open data (Mendel, 2013). They are also well positioned to bridge the gap between public policy and its implementation through the involvement in public-private partnerships and the creation of space for the collaboration of other actors in an open data ecosystem (idem).

NGOs and NPOs historically played a part in pushing for data openness, developing the open data research field, and resolving the practicalities of open data use. Especially because of the cost benefits of using open data, given their non-profit nature. For example, NPO Open Knowledge Foundation (OKF) was founded in 2004 in the UK, representing civil society and providing the infrastructure for open knowledge projects (van Loenen et al., 2018). OKF created their definition of openness about data and content with a first draft created in 2005, followed by improved versions of the definition, which was used as a standardised definition for different actors to use (Open Knowledge Foundation, n.d.). Given the role in society that NGOs and NPOs have, and their representation of the problems, interests, and values of various user groups, it is beneficial to understand their needs related to open data. Hence, the Research Question: *What are the needs of non-governmental (non-profit) organisations related to the use of open data?*

7.2. Method

We performed a systematic literature review to answer the research question. The review is conducted using the approach suggested by (Xiao & Watson, 2019b). The query string is: ("open data" OR "open government data") AND ("n-profit" OR "n-governmental"). We collected 116 papers from two databases: Scopus and Web of Science. The initial papers were screened for duplicates, filtered to be in English language only, and filtered for the publication type such as book chapter, conference paper, and journal article. The filters used did not limit the year of publication. The selection criterion was the focus on NGOs or NPOs' needs as a user of open data. After the screening and the evaluation using the eligibility criterion, only four articles were found to be relevant. This showed that there was a lack of literature on the topic. Thus, we conducted additional systematic literature reviews to inductively identify the needs of NGOs and NPOs based on the activities they perform and the barriers they face in relation to open data.

NG(P)Os are often not investigated on their own but grouped together with other types of intermediaries, so they should not only be searched separately, even when they are outlined as one of the intermediary types included in the study. Therefore, while conducting a systematic literature review, some relevant literature may be excluded if the search is focused solely on NGOs and NPOs. Following such an approach, we did an additional search in the Scopus and Web of Science databases. A barriers-related query string was: ("open data" OR "open government data") AND (barriers OR risks OR challenges OR impediments). Initially, 713 works were collected. The inclusion criterion for the works was that they focus on the open data barriers specific to the NGOs and NPOs as users. After the duplicate deletion and exclusion of the irrelevant or poor-quality articles, four relevant papers were found.

To investigate the activities performed by the NG(P)Os, a more general query string was used, focusing on intermediaries overall: ("open data" OR "open government data") AND ("infomediaries" OR "intermediaries" OR "citizen engagement" OR "civic engagement"). Initially, 255 works were collected from Scopus and the Web of Science. The inclusion criterion for the works was that they focus on the activities of NG(P)Os as open data intermediaries. After the evaluation, this search provided us with sixteen papers in total. Overall, based on three different searches, twenty-four papers were selected.

7.3. Results

After using thematic analysis to identify patterns and themes in the selected literature, five main categories of NG(P)Os' needs as users were identified. First, **Access and Findability** related needs that the organisations have due to barriers they face when trying to reuse open data. Some of the data that NGOs and NPOs need is unpublished, or access to them is restricted (Chattapadhyay, 2014; Cranefield et al., 2014; Yoon & Copeland, 2020). (Erete et al., 2016) highlight the need for some NPOs to have a centralised data hub that would make it easier to look for the needed open data while also having many reliable data sources in one place. Moreover, the authors point out that to find certain datasets NPOs may rely on external actors' advice (Erete et al., 2016; Yoon et al., 2018a). Another need is to have access to a dataset that would have different levels of data available for the analysis, for example, non-aggregated micro-level data (Erete et al., 2016; Yoon & Copeland, 2020). Often such data is not collected or only available in aggregated form (Yoon & Copeland, 2020). Several authors also point out the need to have existing and inclusive infrastructure for NPOs and NGOs to have better access to open data (Chattapadhyay, 2014; Yoon et al., 2018). Given that many NG(P)Os perform activities related to application development, data science, and information dissemination based on open data, they need to be able to find and access the data they require (Baack, 2015; Enaholo, 2017; Gonzalez-Zapata & Heeks, 2015; Hasselwander et al., 2022; Johnson & Greene, 2017a; Ricker et al., 2020; Sangiambut & Sieber, 2017; P. Thakuriah et al., 2017).

Second, NGOs and NPOs have various **Data and Technology** related needs. The organisations may need a different format of open data available, where less data-literate employees might have issues working only with raw data, but more data-savvy ones might require working API (Erete et al., 2016). The less data-savvy organisations might also seek data visualisations available through open data portals or other intermediaries (Brugger, Fraefel, Riedl, Fehr, Schöneck, et al., 2016; Erete et al., 2016; Saxena & Muhammad, 2018; Yoon et al., 2018a). There is also a need to tackle open data usability issues that NG(P)Os can face (Cranefield et al., 2014). Open data and their metadata should be complete, should not have missing data (Chattapadhyay, 2014; Saxena & Muhammad, 2018), and should be cleaned, reformatted, and checked for sensitive information (Hou & Wang, 2017). As many NG(P)Os perform activities related to application development and data science, they need to have usable open data and access to technology to work with open data (Baack, 2015; Enaholo, 2017a; Gonzalez-Zapata & Heeks, 2015; Hasselwander et al., 2022; Johnson & Greene, 2017a; Ricker et al., 2020; Sangiambut & Sieber, 2017; P. Thakuriah et al., 2017).

Third, we identified the need of NGOs and NPOs to have **Skills** or access to an external actor with the necessary skills to analyse open data. (Erete et al., 2016), (Hou & Wang, 2017), and (Yoon & Copeland, 2020) highlight the lack of resources in NPOs that they can allocate towards gaining skills or hiring an additional skilled employee for open data use and analysis. While technical skills are important for NGOs to perform activities related to data and technology, they are also important for open data training that some organisations provide to other intermediaries (Baack, 2015; Enaholo, 2017; Gasco-Hernandez et al., 2018; Gonzalez-Zapata & Heeks, 2015; Kassen, 2017b, 2017a, 2019; Ricker et al., 2020; Santos-Hermosa et al., 2023; P. Thakuriah et al., 2017; Yoon et al., 2018; Yoon & Copeland, 2020). The ability to analyse the open data is also important for NPOs to be able to evaluate their

programmes and measure the outcomes both for internal decisions and funding accountability (Yoon et al., 2018).

Fourth, from the activities performed by the NG(P)Os we can see that there is a need for **Communication** channels between the organisations and other actors such as open data providers, users, and other intermediaries. (Chattapadhyay, 2014a) writes that to understand and respond to the data needs of the NGOs, it is important to have a line for the direct interactions between them and open data providers. One of the kind of activities NPOs perform is organisational, meaning they organise events like conferences and hackathons, connect users to provide feedback to the open data providers, and start campaigns to open the data (Enaholo, 2017; Gonzalez-Zapata & Heeks, 2015; Heimstädt et al., 2014; Kassen, 2017b, 2017a, 2018b, 2019). To conduct these activities, the organisations need to establish connections with different actors to communicate with them, bring them together, and gather and provide feedback related to open data.

Fifth, there is a need for NG(P)Os to have clear **Regulations** related to open data. Especially when it comes to licensing standards and possible privacy concerns, the NGOs require regulations that can make it easier for them to avoid possible complications (Chattapadhyay, 2014; Cranefield et al., 2014). It is also beneficial for NGOs to have regulations that require open data providers to provide open data according to certain standards and have feedback channels to take the needs of NGOs into account (Chattapadhyay, 2014; Cranefield et al., 2014). Table 6 presents all categories of the NG(P)Os' needs with the respective sources from the systematic literature reviews.

Table 6: Needs categories of the NG(P)Os and their respective sources.

Needs' categories	Sources
Access and Findability	Chattapadhyay, 2014; Cranefield et al., 2014; Yoon & Copeland, 2020; Erete et al., 2016; Yoon et al., 2018; Baack, 2015; Enaholo, 2017; Gonzalez-Zapata & Heeks, 2015; Hasselwander et al., 2022; Johnson & Greene, 2017; Ricker et al., 2020; Sangiambut & Sieber, 2017; P. Thakuriah et al., 2017
Data and Technology	Brugger et al., 2016; Erete et al., 2016; Saxena & Muhammad, 2018; Yoon et al., 2018; Cranefield et al., 2014; Chattapadhyay, 2014; Hou & Wang, 2017; Baack, 2015; Enaholo, 2017; Gonzalez-Zapata & Heeks, 2015; Hasselwander et al., 2022; Johnson & Greene, 2017; Ricker et al., 2020; Sangiambut & Sieber, 2017; P. Thakuriah et al., 2017
Skills	Erete et al., 2016; Hou & Wang, 2017; Yoon & Copeland, 2020; Baack, 2015; Enaholo, 2017; Gasco-Hernandez et al., 2018; Gonzalez-Zapata & Heeks, 2015; Kassen, 2017b, 2017a, 2019; Ricker et al., 2020; Santos-Hermosa et al., 2023; P. Thakuriah et al., 2017; Yoon et al., 2018
Communication	Chattapadhyay, 2014; Enaholo, 2017; Gonzalez-Zapata & Heeks, 2015; Heimstädt et al., 2014; Kassen, 2017b, 2017a, 2018, 2019
Regulations	Cranefield et al., 2014; Chattapadhyay, 2014

7.4. Conclusions

There are five categories of needs that NGOs and NPOs have that we found in the literature from looking into their needs, barriers, and activities: 1) Access and Findability, 2) Data and Technology, 3) Skills, 4) Communication, and 5) Regulations. As NG(P)Os are not a homogenous group, the needs may differ depending on the individual, organisational issues, or goals, as some may have more resources or specific priorities, which make certain needs less prominent. Moreover, it should be considered that there is a lack of literature that focuses solely on NGOs and NPOs and their needs. Thus, these findings should be seen as non-exhaustive, and the context of a specific NG(P)O should be considered when looking into its needs.

8. Central/regional government

8.1. Introduction

The open movement has considered open data a benign phenomenon from the outset (Bates, 2012; Kitchin, 2021; Longo, 2011; Safarov et al., 2017). However, over the last years, a growing body of literature has pointed out that the distribution of the benefits of open data is uneven, with some users more capable than others of exploiting their potential (Bates, 2012; Kitchin, 2021). As the use and provision of data by governments increase, they are suspected of having mixed or negative effects on social equity (Ruijer et al., 2022). Developed primarily in the North American public administration tradition, social equity is considered the fourth core public value informing governmental activities, along with efficiency, effectiveness, and economy (Cepiku & Mastrodascio, 2021; Frederickson, 2021). Amid different definitions, social equity can be defined as a multi-dimensional concept that embeds fair and just treatment, justice, and equal and equitable distribution of benefits to the members of society (Cepiku & Mastrodascio, 2021). Social equity differentiates itself from equality. While equality means that people should be treated equally, social equity puts the emphasis on fairness and promotes deviation from "even" distributions of benefits to achieving a just society (Ruijer et al., 2022). Governments should strive for social equity in policy design and policy implementation through equal access, procedural fairness, quality of service provision and equal outcomes (Ruijer et al., 2022).

While recent investigations connected the topic of social equity with data-driven service provisions (Ruijer et al., 2022), contributions to social equity and open data are not named in the literature. As such, the needs of users from disadvantaged groups are not investigated by the literature on open data in the context of governmental activities. With the aim of addressing the lack of analysis of disadvantaged groups' user needs, this contribution will answer the following research question: *What are the needs of a user of open data from a social equity and governmental perspective?*

8.2. Method

In this contribution, we performed a systematic literature review conducted with the PRISMA (the Preferred Reporting Items for Systematic Reviews and Meta-Analyses) process. As a widely adopted systematic review and meta-analysis protocol, PRISMA is used to improve transparency in systematic reviews through a pre-defined and reproducible methodology for the identification of literature (Liberati et al., 2009). Following the PRISMA methodology, we first searched the Scopus and Web of Science (WoS) databases with a predefined set of keywords. Keywords were developed with the aim of obtaining a broad range of articles, and their definition was based on the exam of those used in other systematic reviews on social equity in data-drive public service provision (Ruijer et al., 2022) and on social equity in public service (Cepiku & Mastrodascio, 2021).

The following keywords were used in the database searches: ("Open Data") AND ("Justice" OR "Social equit*" OR "Equit*" OR "Equal*" OR "Inclusion" OR "Divid*" OR "Disparit*" OR "Bias*" OR "Fairness").

The search process was concluded on the 16th of February 2023 and led to the identification of N=1575 records. The initial number of records was screened for duplicates, language (only articles written in English language were accepted), publication type (only peer-reviewed articles), and relevance (only studies related to social equity and open data). Despite the extensive search, after the screening and the eligibility criteria, only twenty-one articles were considered relevant. The analysis of the articles was performed in two steps. We first deductively coded the articles, with the aim of identifying '"user needs"' from disadvantaged groups. The second step involved the identification of patterns and themes through thematic analysis. The findings presented in this contribution are part of a larger continuing study on social equity and open data. The results of the literature review's conclusions will be supported and corroborated by empirical data.

8.3. Descriptive results of the literature review

The literature on open data and social equity refers to the needs of marginalised users in a variety of ways. We present here recurring themes extracted from the literature that problematize the needs of users (or potential users) of open data from a social equity perspective.

In general, data literacy, presented in some cases as associated with digital equity, is mentioned in different studies (Shibuya et al., 2022; B. Wilson & Chakraborty, 2019; B. Wilson & Cong, 2021a; S. Zhang, 2022) with some authors considering the concept as outdated and to be absorbed by the one of digital inequality (Bezuidenhout et al., 2017). Concerns over data quality (Fernández-Ardèvol & Rosales, 2022; Ossom-Williamson et al., 2021), access to data (Overby et al., 2022), and data availability (Jarke, 2019; Schwoerer, 2022) are also present in different studies. Some authors suggest strategies for data governance (Fusi et al., 2022; Walter et al., 2021), as in the case of the indigenous population, and more specifically, how to address the lack of institutional capacity at the local level (Svärd, 2018; Zuiderwijk et al., 2018). According to Table 7, the literature tends to devote higher attention to the local level, intended as municipalities and smart cities. The literature also presents intermediaries, such as University libraries, and contexts of open data and open access use, as in the case of laboratories.

Table 77: Concepts associated with the user needs of disadvantaged groups

Concepts associated with user needs	Source	Level of government
Data literacy, digital literacy, and digital equity	(Shibuya et al., 2022; Wilson & Chakraborty, 2019; Wilson & Cong, 2021; Zhang, 2022)	Local government (Shibuya et al., 2022; Wilson & Chakraborty, 2019; Wilson & Cong, 2021) N.A. (i.e., University library in Zhang, 2022)
Digital divide	(Ghose & Appel, 2016; Lee et al., 2023; Zhang, 2022) (Bezuidenhout et al., 2017)	Local government (Lee et al., 2023) N.A. (University libraries in Ghose & Appel, 2016; Zhang, 2022) N.A. (Laboratories in Low and Middle Income Countries, LMIC)
Data availability	(Jarke, 2019; Schwoerer, 2022)	Local government
Data divide	(Fusi et al., 2022)	Central government
Data quality	(Fernández-Ardèvol & Rosales, 2022; Ossom-Williamson et al., 2021)	Supra-national governance (Fernández-Ardèvol & Rosales, 2022) Local and Central Government (Ossom-Williamson et al., 2021)
Access to data	(Overby et al., 2022)	State/Province government
Data governance	(Fusi et al., 2022; Walter et al., 2021)	Central government
Uneven access to data across territories	(Svärd, 2018; Zuiderwijk et al., 2018)	Local government

8.4. Main findings

This section aims at articulating how user needs for marginalised user groups are presented in the literature under different themes.

Data literacy, digital literacy, and digital equity

Different authors refer to barriers for users and potential users of open data as being the result of the lack of data literacy (Shibuya et al., 2022; Wilson & Chakraborty, 2019; Wilson & Cong, 2021; Zhang, 2022). Open government data, including spatial data, are used more by literate users, which often include governmental staff, businesses, and journalists (Shibuya et al., 2022; Zhang, 2022), whereas other users are left behind (Wilson & Cong, 2021). Indeed, not all users possess the same data literacy, and governments should invest in building data literacy and digital equity (Wilson & Cong, 2021). Digital equity captures what is needed for actors to participate in and through open government data and is composed of both the acquisition of the necessary skills (defined as digital literacy) and technologies (Wilson & Cong, 2021).

In their investigation of the effects of open data on citizens' behaviour during the Covid-19 outbreak, Shibuya et al. (2022) conclude that information sharing through open data did not have an equal impact on all citizens. Indeed, citizens with high digital literacy were more used to making decisions based on data even before the pandemic" (p. 6). The authors, therefore, suggest that open data research focuses on understanding the needs for information of all citizens.

While open data presents opportunities, a more active role of different actors is invoked to fill the data literacy gap. Zhang (2022) focuses on equitable approaches to teaching spatial literacy, considered to promote the engagement of the public and as a key to people's empowerment. Author also discusses the implication of teaching spatial data literacy in relation to themes focused on equity. By such an approach, it is possible to link geospatial data to relevant issues and reach the public's engagement.

Wilson and Chakraborty (2019) investigate civic technology as a field that deploys open data with the aim of giving visibility to problems not addressed by governments through collaboration. While many factors are described as key to the success of civic technologies initiatives, digital literacy is reckoned as a challenge for untapping the potential of these forms of collaboration. Authors also put an emphasis on transitioning to a new form of "just" planning leveraging on open data. However, this transition requires additional data skills for planners to allow them to consider open data in their activities.

Overall, the different contributions to digital literacy, data literacy, and digital equity call for addressing user needs through training and technology provided to different actors (e.g., citizens or planners) by different providers (i.e., governments or libraries as data intermediaries).

Digital divide

Access to data is uneven and is limited by digital divide (Ghose & Appel, 2016; Lee et al., 2023). Ghose & Appel (2016) research participation in geospatial data and the role of university libraries as viable options for intermediating access to open data for less privileged groups. Indeed, the digital divide prevents users and potential users from accessing the appropriate and relevant geospatial data. The digital divide is further exacerbated by imbalances in power that lead to the development of infrastructures, actors' collaboration, and policies in a way that does not satisfy the needs of disadvantaged and marginalised groups (Ghose & Appel, 2016). The same groups cannot reap the benefits of open data as access to relevant datasets in the geospatial domain implies, in some cases, negotiating power and skills.

Lee et al. (2023) focus on smart cities and present, among others, the case of the City of Portland, in which the local government took actions to bridge the digital divide that resulted from decades of marginalisation through actions targeted primarily at marginalised areas keeping as priority community engagement to solve the most pressing problems that characterise underserved areas The

"open and people-first approach to digital transformation" (p. 88) was channelled through an Open Data Resolution specifically aimed at participation and reuse from data by citizens.

According to Bezuidenhout et al. (2017), the digital divide is an obsolete concept to which we should prefer "digital inequality". The focus of digital inequality is access intended as both a social and technological issue. With the adoption of a broader definition of the digital divide, it is possible to understand and solve the variety of reasons that prevent the reuse of data, and that includes a "complex mixture of social, psychological, economic and pragmatic reasons" (Selwyn 2004: 348, as cited by Bezuidenhout et al.).

Data availability
Schwoerer (2022) investigates citizens' user needs in local governments and finds out that data availability is a major issue. The information needs of citizens are different from those of other actors, and, as already pointed out by previous studies (Ojo et al. 2018), open datasets concerning specific policy areas such as health and environment are considered more relevant. Local governments are not only providers of data but also "catalysts of "hidden" demand for open data" of different user groups. Issues regarding data availability are identified also by Jarke (2019). Approaches based on co-creation through open data should deal with different citizens' information needs. Indeed, open data that citizens consider relevant cannot be available, and therefore, data collection and data creation should come first as priorities. To this end, collaborations with different data owners should be envisaged to meet the open data availability gap.

Data divide
Fusi et al. (2022) investigates open government data for environmental justice and argues that governments should be responsible for filling the data divide that prevents the participation of disadvantaged groups (in this case, vulnerable population) in policymaking. The unprivileged socio-economic background of some groups of citizens limits the opportunities for accessing data and creates phenomena of data divide. Governments need to consider differences in access and treat it as an important equity issue. Also, the focus on "technical" features of data, such as "machine readability, quantity and granularity", widens these gaps by giving a competitive advantage in the use of open data to actors who are already empowered and skilful.

Data quality
According to Fernández-Ardèvol & Rosales (2022), data use and reuse need to consider the potential biases of secondary datasets, intended as datasets that were created for different purposes. Data quality assessment, indeed, does not ensure that data are bias-free and, more specifically, that data adequately reflect the existence and the characteristics of different user groups, such as minorities or social groups on which data are not collected. The definition of data quality is "socially constructed" and, therefore, it is important to acknowledge and consider that open data that meet existing standards might not satisfy user needs. Also, the issue of data quality should be seen in connection with the lack of data skills. Users might lack critical approaches towards data and, therefore, they might not be able to recognize that data are not objective and that they can embed limitations.

Legal, political, and organisational barriers for data access
Overby et al. (2022) investigate legal, political, and organisational barriers to data accessibility. Analysing the case of a conservation easement and land records in the United States, the Authors find out that concerns over privacy, as well as political and organisational barriers, prevent wide access and participation of citizens in environmental governance through open data. Barriers to access prevent a "just" environmental governance, and it is suggested that governments strive for equitable data access to facilitate the participation of diverse groups and satisfy societal needs.

Data governance
Governance through new data principles might emphasise collective ownership, control, and self-determination, as well as the needs of marginalised groups. One leading example is the one of the CARE (Collective benefit, Authority to control, Responsibility, Ethics) principles, proposed as a remedy for the flaws of the FAIR (Findable, Accessible, Interoperable, Reusable) principles (Walter et al., 2021). Indeed, the FAIR principles, developed in the Western tradition, do not reflect, or consider the needs of disadvantaged groups. For this reason, the RDA International Indigenous Data Sovereignty Interest Group developed the CARE principles (Walter et al., 2021). These new data principles can be a source of inspiration for developing more inclusive data governance in other contexts, such as in environmental justice and open data governance, as suggested by Fusi et al. (2022).

Uneven access across territories
Access to data for users might be uneven across different territories. In practice, some local governments might share more data than others. Focusing on governmental user needs at the local level in relation to the implementation of the PSI Directive, Svärd (2018) contends that financial constraints impede the adoption of open data-sharing practices. As a result, the availability of data changes across municipalities, and this might result in inequities of access.

For other authors (Zuiderwijk et al., 2018), while budget constraints might play a role, differences in open data uptake at the local level can also result from uneven motivations among municipalities for adopting them. Therefore, at the central/national level, policies need to consider differences that might arise in relation to the adoption of open data policies.

8.5. Conclusions

In the review of the literature on open data and social equity, we identified a variety of user needs that are connected to the disadvantaged groups that encompass: 1) Data literacy, digital literacy, and digital equity, 2) Digital divide and Digital inequality, 3) Data availability, 4) Data Divide, 5) Data Quality, 6) Legal, political, and organisational barriers for data access, 7) Data Governance, 8) Uneven access across territories. The identified user needs do not always involve governments as users of open data but rather as mediators of the demand for open data (data availability), providers of training or funding for training (data literacy, digital literacy, and digital equity) or "guarantors" of data quality and absence of biases in data. Also, the focus of the literature is on local governments, with fewer contributions about regional and central governments. Overall, the literature on open data and social equity, while not systematically discussing implications for marginalised groups in relation to the use of open data, suggests different avenues for further research.

9. Companies

9.1. Introduction

Open Data (OD) refers to data that can be accessed and shared without restriction. OD is revolutionizing how businesses operate in the modern digital age. Many companies are adopting OD initiatives to stimulate innovation, enhance decision-making, and uncover latent market potential (Hammell et al., 2012). As organizations become increasingly data-driven, it is essential that they comprehend and effectively align their requirements. This section examines companies' indispensable requirements for open data. We will consider both the benefits and obstacles associated with its implementation.

To make informed decisions, companies require access to high-quality, reliable open data sources (Zuiderwijk et al., 2015). The significance of data quality cannot be overstated, as inaccurate information can result in wrong strategies and wasted resources. For instance, energy companies can use open data on solar radiation and wind patterns to identify optimal locations for renewable energy projects, ensuring reliable energy generation. On the other hand, waste management companies can use open data on population density, waste generation rates, and traffic patterns to optimize waste collection routes, thereby increasing efficiency and decreasing environmental impact.

Accessibility and the seamless integration of diverse datasets are essential user requirements for businesses (Väyrynen et al., 2017). Companies require user-friendly (accessible) interfaces on open data platforms that enable them to search, filter, and download data with minimal effort. In addition, interoperability between diverse datasets is essential, as organizations frequently need to combine and analyze multiple data sources to derive insightful conclusions. OD on public transit routes, schedules, and ridership can help transportation agencies and companies optimize public transportation systems, improve service coverage, and enhance the commuter experience.

Businesses have diverse and specific data requirements depending on their industry, market, and goals. Therefore, companies require OD platforms that offer customization and adaptability regarding data formats, frequency, and granularity. This customization allows companies to tailor the data to their specific requirements and gain insights related to their business area. For example, businesses in the energy sector can leverage open data on energy consumption patterns to identify high-demand areas, enabling them to develop targeted energy efficiency programs and optimize grid operations. The utilization of open data about traffic flow, road conditions, and accidents can be advantageous for transportation companies in implementing efficient traffic management tactics, optimizing signal timings, and mitigating congestion.

OD can help companies in their decision-making processes. They can make educated judgments with the availability of comprehensive data, which lowers uncertainty and minimizes risks (Zuiderwijk et al., 2015). For example, transportation authorities can use open data on traffic flow, road conditions, and accidents to implement effective traffic management strategies, optimizing signal timings and reducing congestion. In addition, OD fosters innovation by providing businesses with insightful information that can disclose new market opportunities, trends, and growth areas. Energy companies can use open data on energy consumption patterns to identify high-demand areas, enabling them to develop targeted energy efficiency programs and optimize grid operations.

Challenges associated with the adoption of OD should also be mentioned. The massive quantity of accessible data can be overwhelming for businesses, making it difficult to identify and prioritize the most pertinent information for their purposes. Companies can overcome this obstacle, however, by focusing on specific open data sources relevant to their initiatives. There may also be data quality concerns, as open data sources may contain errors or inconsistencies. Companies can mitigate this

difficulty and ensure the veracity of the information they use by thoroughly evaluating and validating their data sources.

9.2. Method

The present study employed a two-fold approach comprising a comprehensive literature review and a thorough analysis of practical applications. A comprehensive literature review was conducted to investigate the impact of OD on business operations. The task necessitated the examination of academic literature, technical documents, and analytical studies pertaining to open data endeavours. The literature was meticulously selected to guarantee its pertinence and reliability, furnishing a comprehensive outlook on the ramifications of open data in various sectors.

After the literature review, we examined various scenarios that showcase the pragmatic application of OD in the process of making business decisions. The instances were carefully chosen from diverse sectors, such as energy, waste management, and transportation, to demonstrate a variety of situations. The purpose of these examples is to demonstrate how enterprises can effectively employ open data to fulfil their unique needs while highlighting potential challenges that may arise. Through the integration of these two stages, a comprehensive understanding of the open data business user requirements was achieved. The methodology employed in this study incorporated theoretical and practical perspectives derived from relevant literature and real-world applications.

9.3. Results

Following the literature review and the applications analysis, the following needs were identified.

1. Access to high-quality and reliable OD sources: Companies need accurate and reliable data to make good choices and avoid wrong strategies. They need trustworthy sources of OD relevant to their business area.
2. Integration and accessibility of diverse datasets: Companies must have access to user-friendly interfaces, and datasets must be interoperable between them. They require access to open data platforms that facilitate data discovery, refinement, and download. By combining and analysing data from different sources, companies can gain important insights that would not otherwise be possible.
3. Customization and adaptability of data platforms: Different industries, markets, and business goals have different and unique data needs. They need OD tools that let them change the style, frequency, and level of detail of the data. This availability lets them change the info to fit their needs and learn more about the situation.
4. Availability of data: Companies can take better decisions when they have all the needed information. OD platforms should give companies access to a wide range of data sources to get the needed information.
5. Data quality assurance: Companies need to know that their OD sources are of good quality. They need ways to check and confirm the data sources to ensure they are correct, reliable, and consistent.
6. Overwhelming amount of data: There are vast amounts of data companies can access. They need ways to find and rank the most important information for their goals. This problem can be solved by focusing on specific OD sources that are useful for their projects.
7. Data quality: OD sources may have errors or be inconsistent. Concerns about data quality need to be dealt with by companies by carefully examining and confirming their data sources. Truthful data helps ensure the decision-making process.
8. Support: For companies to use OD effectively, they need help and direction. They need access to tools, expertise, and help, to maximize the benefits OD brings to the business.

9.4. Conclusions

In conclusion, open data has the potential to significantly revolutionise how organisations conduct their daily operations whilst providing a range of benefits, including improved decision-making

abilities, lower costs, and the potential to stimulate innovation. The effective implementation of open data initiatives requires careful consideration of various factors such as data quality, accessibility, customisation, security, and support requirements for enterprise users. It would, in turn, promote informed decision-making, enhance operational processes, and promote a sustainable and effective future for organisations. The utilisation of open data poses certain obstacles; nevertheless, businesses that effectively overcome these challenges can attain a competitive edge.

10. Artificial Users

10.1. Introduction

The open data ecosystem holds significant importance in the current digital age and serves as a crucial resource for a diverse range of users (Runeson et al., 2021). In recent times, there has been a notable surge in the significance of artificial users. In the context of open data, "artificial users" refers to entities that utilize open data for consumption, analysis, learning, and informed decision-making processes (Helm et al., 2020). The category of artificial users encompasses entities such as artificial intelligence (AI), machine learning, and bots (Nikitas et al., 2020).

Artificial users are integral components of various contemporary systems, including but not limited to recommendation engines, search algorithms, and data analytics tools (Kosala & Blockeel, 2000). In our technology-centric society, they play a crucial role as the foundation of numerous routine functions, ranging from tailored suggestions on digital interfaces to anticipatory analysis in domains such as finance, healthcare, and transportation, among others.

Despite the extensive research conducted to understand the requirements of human users in open data ecosystems, the specific needs of artificial users have not been given equal importance. The development of synthetic users necessitates a distinctive methodology owing to their diverse needs. Accurate identification and comprehensive understanding of particular requirements are essential for enhancing the efficiency of artificial users and maximizing their potential in open data ecosystems (Janssen et al., 2020).

The subsequent segment will examine the needs of artificial users in the domain of open data. The objective of this research is to offer significant perspectives and direction to experts who are involved in the management of open data and artificial intelligence. A comprehensive evaluation of the prerequisites and their probable consequences on open data systems will be executed to attain this objective. The main aim is to facilitate the establishment of open data eco-systems customized to meet synthetic users' needs, thereby enhancing their effectiveness and influence on our data-driven society (Welle Donker & van Loenen, 2017).

10.2. Method

We employed a rigorous research methodology to gain a comprehensive understanding of the requisites of synthetic users in the domain of open data. The initial methodology employed encompassed a thorough review of relevant literature and an examination of established practices within the domains of artificial intelligence and open data ecosystems. This enabled us to acquire a comprehensive perspective and identify shortcomings in the current understanding of the requirements of artificial users.

During the second phase of our methodology, our primary objective was to develop and implement an analytical framework to assess the distinct requirements of synthetic users from diverse viewpoints. A thorough examination was carried out on diverse categories of synthetic users, encompassing both artificial intelligence (AI) systems and software bots. Furthermore, we assessed various scenarios in which these users engage with open data.

Concurrently, we conducted a study which evaluated various state of the art algorithms such as BERT, RAKE, YAKE, TEXTRANK, and CHATGPT, with a focus on their performance comparison (Campos et al., 2020; Hu et al., 2018; Qian et al., 2021; Song et al., 2023). Furthermore, a new hybrid methodology referred to as BRYT was introduced and integrated into the comparative analysis. The methodology employed in this study focuses on the extraction of pertinent metadata from datasets to assist artificial users, as illustrated in Figure 2. The algorithms' effectiveness was assessed based on their ability to pull out important keywords from the dataset's descriptions. This is a key factor in improving

data access and discoverability for automated users. The efficiency of the algorithms was evaluated based on their ability to extract important keywords from the dataset's descriptions. This is a key factor in improving accessibility and discoverability for automated users.

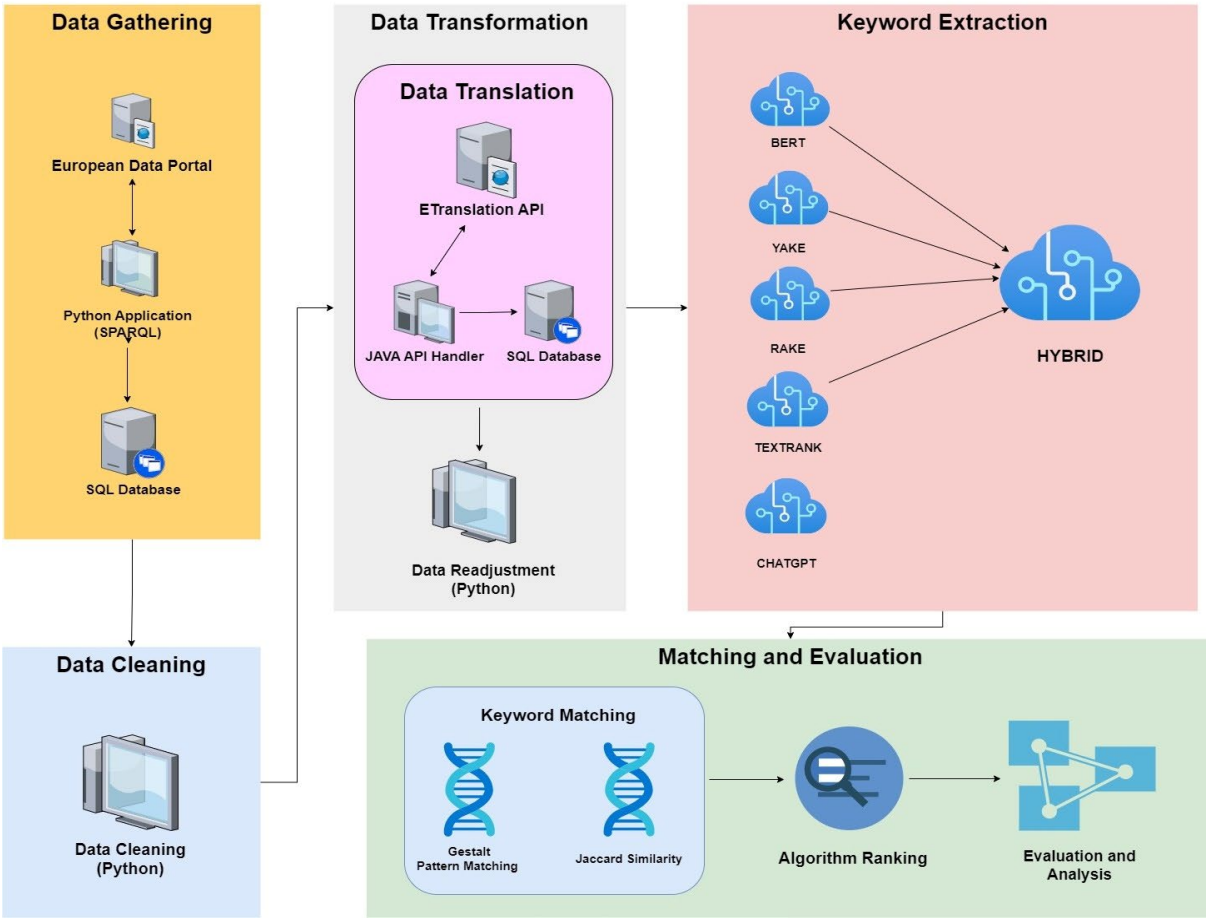

Figure 2: BRYT Keyword Extraction

10.3. Results

Our findings highlighted several important requirements for artificial users pertaining to open data.

Firstly, our research stressed the need for reliable and consistent data for artificial users. This is because these organizations depend on precise, thorough, pertinent, and timely data to function. Inaccurate results may be produced if the data sources are faulty, or the data formats are inconsistent. Secondly, our study demonstrated how important it is for data to be accessible and discoverable in the open data ecosystem. Artificial users should be able to explore data systems with ease and find pertinent facts in a sea of information.

Thirdly, we noted the need for standardization and compatibility. This calls for data standardization since it is necessary for data from many sources to be readily combined and used together. Furthermore, it is essential for artificial users to have access to real-time or frequently updated material given the quickly changing nature of information nowadays. We also emphasized the need of strong privacy and security safeguards and the significance of upholding ethical principles while using data.

Lastly, the outcomes of our suggested technique, BRYT, in terms of our keyword extraction algorithms were promising. When it came to collecting representative keywords from the dataset's descriptions, it

consistently outperformed other examined approaches. This is a crucial aspect of improving data findability.

10.4. Conclusions

In conclusion, this research sheds vital light on the unique needs of artificial users in the context of open data, underlining the necessity to cater to these requirements for maximizing the effectiveness of these entities. The findings reveal that factors such as data quality, accessibility, interoperability, real-time updates, and ethical considerations are of paramount importance. Furthermore, our proposed methodology, BRYT, shows promising potential in enhancing data findability, a critical need for artificial users.

This research, however, is just the beginning. As the field of AI continues to evolve, it's crucial to continue exploring the needs of artificial users and refine our strategies accordingly. The focus should be on creating open data ecosystems that are tailored to the needs of artificial users, thereby fostering a digital environment where artificial and human users can coexist and thrive.

11. Open data intermediaries

11.1. Introduction

There are socio-technical barriers to the meaningful use of open data, such as a lack of knowledge about the data, ambiguity surrounding data licenses, and a lack of the necessary software to process data. Open data intermediaries play an important role in addressing these challenges (Davies & Edwards, 2012). They are defined as "third-party actors who provide specialized resources and capabilities to (i) enhance the supply, flow, and/or use of open data and/or (ii) strengthen the relationships among various open data stakeholders" (Shaharudin et al., 2023). Understanding the needs of open data intermediaries is necessary to ensure that they can play their role more effectively and contribute to a sustainable open data ecosystem.

11.2. Method

A systematic literature review was conducted to better understand who indeed open data intermediaries are by looking into their types of actors, tasks, and objectives (Shaharudin et al., 2023)[2]. From the types of actors, tasks, and objectives of open data intermediaries gathered, their needs were inductively identified. Additionally, the same literature pool in (Shaharudin et al., 2023) was utilised to capture challenges faced by open data intermediaries, which are also useful in the identification of the needs. To complement the findings from the literature, an interview with a representative of an open data intermediary, Esri Nederland, which is one of the partners of ODECO, was also conducted.

11.3. Results

From the literature, various types of actors of open data intermediaries were identified (Table 8). While most of them are users of open data, some of them advocate for or facilitate access to open data. They are not necessarily organizations – some of them are individuals such as entrepreneurs, individual developers, and researchers.

Table 8: Types of actors of open data intermediaries (Shaharudin et al., 2023)

Type of actor	Sources
Civil society organizations (CSOs)	(Mayer-Schönberger & Zappia, 2011), (Cañares, 2014), (González-Zapata & Heeks, 2015), (Brugger, Fraefel, Riedl, Fehr, Schöeneck, et al., 2016), (Germano et al., 2016)
Entrepreneurs/ businesses	(Cañares, 2014), (Janssen & Zuiderwijk, 2014), (Germano et al., 2016), (Andrason & van Schalkwyk, 2017), (Glassey, 2017)
Media	(Cañares, 2014), (Baack, 2015b), (Brugger, Fraefel, Riedl, Fehr, Schöeneck, et al., 2016), (Meng, 2016), (Johnson & Greene, 2017)
Public organizations	(Janssen & Zuiderwijk, 2014), (Chan et al., 2016), (Johnson & Greene, 2017), (Robinson & Mather, 2017), (Kim, 2018)
Researchers	(Meng, 2016), (Johnson & Greene, 2017b), (Park & Gil-Garcia, 2017), (Corbett et al., 2018), (Kim, 2018)
Multi-partner	(Hielkema & Hongisto, 2013), (Meijer & Potjer, 2018)

Open data intermediaries do a wide range of tasks (Table 9) at various stages of the open data lifecycle, deploying various types (Shaharudin et al., 2023). Typically, multiple tasks are needed for them to serve their functions. Most of the tasks entail active processing of open data, such as collecting, augmenting, contextualizing, visualizing data, and developing products and services with open data. However, some tasks do not necessarily require them to actively process open data, for example, building data capacity, facilitating stakeholders' interactions, and channelling feedback.

[2]This systematic literature review has been published as a research article in a peer-reviewed journal whose objective is to propose a common definition of open data intermediaries.

Table 9: Tasks of open data intermediaries (Shaharudin et al., 2023)

Task	Source
Compile data	(Dumpawar, 2015), (González-Zapata & Heeks, 2015), (van Schalkwyk et al., 2015), (Chan et al., 2016), (Meng, 2016)
Build data capacity	(Davies & Edwards, 2012), (da Silva Craveiro & Albano, 2017), (Enaholo, 2017), (Maail, 2017), (Robinson & Mather, 2017)
Augment data	(Davies & Edwards, 2012c), (Dumpawar, 2015), (Andrason & van Schalkwyk, 2017), (Young & Verhulst, 2017), (Corbett et al., 2018)
Contextualize data	(Dumpawar, 2015), (González-Zapata & Heeks, 2015), (Brugger, Fraefel, Riedl, Fehr, Schöeneck, et al., 2016), (Meng, 2016), (da Silva Craveiro & Albano, 2017)
Curate data	(Davies & Edwards, 2012c), (Dumpawar, 2015), (Chan et al., 2016), (Andrason & van Schalkwyk, 2017), (den Haan, 2018)
Develop products and services	(González-Zapata & Heeks, 2015), (Chan et al., 2016), (Meng, 2016), (Andrason & van Schalkwyk, 2017), (Corbett et al., 2018)
Interpret data	(Dumpawar, 2015), (van Schalkwyk et al., 2015), (Meng, 2016), (Enaholo, 2017b), (Corbett et al., 2018)
Validate data	(Dumpawar, 2015), (González-Zapata & Heeks, 2015), (Corbett et al., 2018), (den Haan, 2018), (Kim, 2018)
Demand open data	(González-Zapata & Heeks, 2015), (Enaholo, 2017b), (Corbett et al., 2018), (Kim, 2018), (Meng et al., 2019)
Visualize data	(Dumpawar, 2015), (Brugger, Fraefel, Riedl, Fehr, Schöeneck, et al., 2016), (Meng, 2016), (Enaholo, 2017b), (den Haan, 2018)
Facilitate stakeholders' interactions	(Juell-Skielse et al., 2014), (Dumpawar, 2015), (Chan et al., 2016), (Meng, 2016), (den Haan, 2018)
Channel feedback	(Chan et al., 2016), (Enaholo, 2017b), (den Haan, 2018), (Hablé, 2019), (Navalkha, 2021)
Improve technical openness of data	(Meng, 2016), (Maail, 2017), (den Haan, 2018), (Meng et al., 2019), (Navalkha, 2021)
Identify risks of opening data	(Davies & Edwards, 2012)

There are several objectives that open data intermediaries set to achieve (Table 10). These objectives can generally be grouped into two: (i) to enhance the supply, flow, and/or use of open data; and (ii) to strengthen the relationships among various open data stakeholders (Shaharudin et al., 2023).

Table 1010: Objectives of open data intermediaries (Shaharudin et al., 2023)

Objective	Source (non-exhaustive)
Facilitate use	(Chattapadhyay, 2014), (Maail, 2017), (Robinson & Mather, 2017), (P. (Vonu) Thakuriah et al., 2017), (Yoon et al., 2018)
Increase the accessibility to open data	(Chattapadhyay, 2014), (Baack, 2015), (Meng, 2016), (van Schalkwyk et al., 2016), (Enaholo, 2017b)
Close the feedback loop	(Hielkema & Hongisto, 2013), (Frank & Waddell, 2014), (Meng, 2016), (Enaholo, 2017), (Maail, 2017)
Provide services to citizens	(Davies & Edwards, 2012), (Frank & Waddell, 2014), (Andrason & van Schalkwyk, 2017), (Glassey, 2017), (Sangiambut & Sieber, 2017b)
Bring stakeholders together	(Mayer-Schönberger & Zappia, 2011), (Hielkema & Hongisto, 2013), (Juell-Skielse et al., 2014), (Andrason & van Schalkwyk, 2017), (Maail, 2017)

Objective	Source (non-exhaustive)
Enhance trust between stakeholders	(Andrason & van Schalkwyk, 2017), (Johnson & Greene, 2017b) (Maail, 2017), (Kim, 2018)
Improve open data practices	(Maail, 2017), (Park & Gil-Garcia, 2017), (Meng et al., 2019)

Studies on the challenges faced by open data intermediaries are limited (Table 11). The challenges are based on the types of actors, tasks, and objectives of the open data intermediaries. For example, (Enaholo & Dina, 2020) found the lack of data competencies as the main challenge faced by journalists in Nigeria as they are from traditional media of which data competencies were not necessarily expected. Meanwhile, for businesses that act as open data intermediaries since their inception, data competencies are less likely a challenge, but designing products for the less digitally savvy end-users is (Andrason & van Schalkwyk, 2017).

Table 1111: Challenges faced by open data intermediaries

Challenge	Source
Cost of product deployment	(Andrason & van Schalkwyk, 2017)
Data competencies	(Enaholo & Dina, 2020)
Data quality	(Dumpawar, 2015)
Designing products for less digitally savvy users	(Andrason & van Schalkwyk, 2017)

From the types of actors, tasks, objectives, and challenges of open data intermediaries gathered, their needs were inductively identified (Table 12).

Table 1212: Needs of open data intermediaries

Needs	Note
Adequate funding and sustainable business model	Open data intermediaries include several non-profit groups, namely, civil society organizations, public organizations, and researchers. These groups mostly rely on sponsorship to conduct their work. Meanwhile, businesses need sustainable business models to continue working with open data. Cost of product deployment, financial sustainability, and competition to hire skilled staff are indeed some of the challenges found in the literature (Andrason & van Schalkwyk, 2017; Flores, 2020).
Timely, consistent, and high-quality data	Since most of the tasks of open data intermediaries involve the processing of open data (see Table 8), they rely on timely and high-quality data. Data quality is identified as one of the challenges faced by open data intermediaries (Dumpawar, 2015). Based on an interview with Esri Nederland, timeliness and consistency are also an issue as some data is irregularly updated and the format is not consistent over time.
Data skills	As not all open data intermediaries are data-oriented from the start (e.g., some CSOs, media, and researchers), data skills training for them is necessary. This is consistent with the finding by (Enaholo & Dina, 2020) citing data competencies as one of the challenges faced by journalists acting as open data intermediaries.
Domain-specific knowledge support	Some objectives of open data intermediaries, such as providing services to citizens and closing the feedback loop, require domain-specific knowledge. Challenges of designing products for less digitally savvy users and scepticism of end-users (Andrason & van Schalkwyk, 2017) are related to the lack of domain-specific knowledge. While some open data intermediaries may already have domain-specific knowledge in-house, some may need to hire people with this knowledge or form collaborations with external parties that have this knowledge.

Needs	Note
Domain-specific data standards	Related to data quality and domain-specific knowledge is domain-specific data standards. The interview with Esri Nederland revealed that in some domains/industries (e.g., gas industry), the data standards are not mature yet resulting in different data standards by different data providers within the same industry/domain. This makes it difficult for open data intermediaries to compile the data.
Knowledge of ethical data practices and impacts	Ethical concerns are one of the challenges faced by open data intermediaries. For example, (Dumpawar, 2015) found out that certain products by open data intermediaries unfairly benefit certain communities over others. Thus, the ethical considerations that need to be considered by open data intermediaries go beyond how to process data ethically (e.g., protecting privacy), to include how to ensure the data-based products do not negatively impact certain communities. Besides, certain tasks by open data intermediaries such as contextualizing data and curating data entail them putting meaning to the data of which (implicit) bias may be imposed if not careful.
Reliable infrastructure	In the case of developing countries, such as Ghana (Andrason & van Schalkwyk, 2017), open data intermediaries encountered basic infrastructure issues such as reliable access to the Internet. Nevertheless, infrastructure also includes services such as continuous and uninterrupted access to data providers' application programming interface (API), which is also a concern in developed countries.

11.4. Conclusions

By gathering the types of actors, tasks, objectives, and challenges of open data intermediaries from the literature and conducting an interview with Esri Nederland, seven needs of open data intermediaries were identified: (i) adequate funding and sustainable business model, (ii) timely, consistent, and high-quality data, (iii) data skills training, (iv) domain-specific knowledge support, (v) domain-specific data standards, (vi) knowledge of ethical data practices and impacts, and (vii) reliable infrastructure. In general, the needs are related to finance, data, skills & knowledge, and infrastructure. Since open data intermediaries are a diverse group, some needs are more relevant to certain open data intermediaries than others.

12. Discussion and Conclusions

In this section, we present conclusions about the research question: *What are the user needs of open data users?* considering the open data user groups which were explored in previous chapters. First, we present general conclusions, and a categorization of user needs which summarizes the results and reflect on the work undertaken. Then we present the main limitations of the study, and finally, in the last section, we set out a research agenda.

Through identifying user needs, we could also identify some interconnections between the different user groups, which might contribute to increased circularity and inclusivity in an open data ecosystem. Figure 3 provides an example of these interconnections and possible shared user needs. On one side, students might immerse in a bigger group of non-specialist users. Meanwhile, in non-homogeneous user groups such as Government, Intermediaries, and Companies, and NG(P)Os, some of the actors might be non-specialists or *Content Experts,* as quoted in the second chapter of this report. Finally, Journalists might position in interceptions between non-specialist users, Intermediaries and Companies, and NG(P)Os. Nevertheless, by considering these interconnections, it is possible to recognise different degrees of literacy that could be addressed by the interaction of several user groups between non-specialised and specialised actors in the OD ecosystem. Furthermore, bridges between user groups within an ecosystem perspective might help to address shared needs. For example, the strengths of a group can help to solve the needs of another group, such as the case of artificial users supporting the non-specialised user's need for findability or discoverability of data. Although these seven user groups were analysed for this report and ongoing research is aimed at improving our understanding of them, the results helped to recognise the complexity and variety of user groups and types, which might vary according to task, roles, and the specific context in an open data ecosystem.

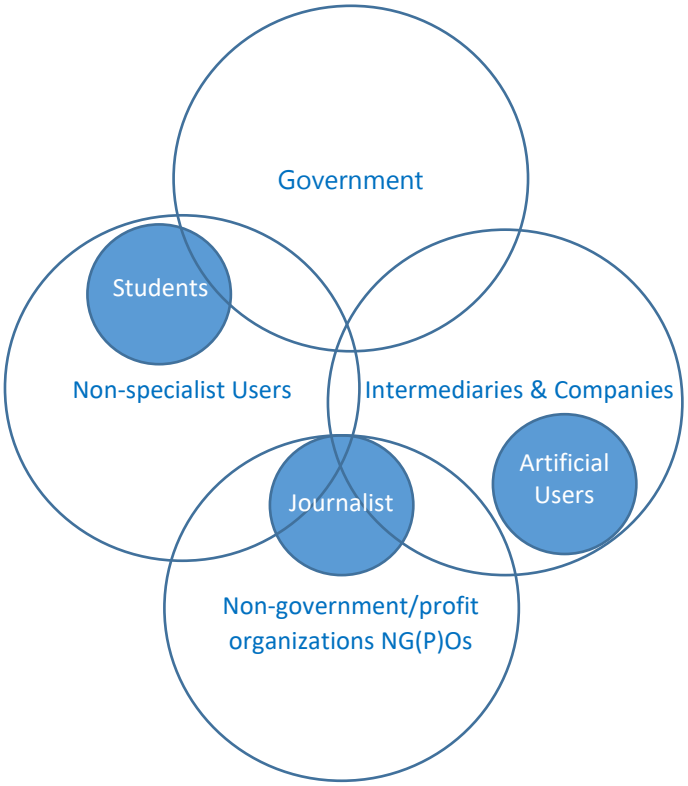

Figure 3: Interconnection between user groups according to shared needs.

Overall, the literature highlights those roles, needs as well as desires change over time. In some cases, as per the *local governments*, the roles and needs or desires that they can acquire within the lifecycle of opening data can vary according to the goals to be achieved and the context of data use. The identified user needs do not always involve *central and regional governments* as users of open data but rather as mediators of the demand for open data (data availability), providers of training or funding for training (data literacy, digital literacy, and digital equity) or "guarantors" of data quality and absence of biases in data. As *NG(P)Os* are not a homogeneous group, their needs may differ depending on the individual organisational issues or goals, as some may have more resources or specific priorities, which make certain needs less prominent. The same considerations apply to *open data intermediaries* that constitute a diverse group, and, therefore, some needs are more relevant to certain open data intermediaries than others. Nevertheless, we attempt to group together general user needs.

12.1. Categorization of user needs

In this report, we conducted a literature review to identify user needs for a variety of user contexts that included: non-specialist data users, local government, journalists, students, NGOs, central/regional government, companies, artificial users, and open data intermediaries. Although the list of nine user types is not exhaustive, with this approach, we attempted to investigate the needs of an open data ecosystem in a comprehensive way by including actors who have been neglected by previous research, such as disadvantaged and marginalized groups. Drawing on the different strands of literature presented in the previous chapters, we can identify differences and similarities among the needs of the users of open data. The commonalities in the multidisciplinary analysis allow for a broad categorization of user needs under the following themes.

Literacy

The literature highlights the importance for open data users of acquiring a broad range of skills. *Non-specialist users* participating in hackathons need formation methods for mixing and interacting with specialist users, as well as freedom and support to learn new data skills. *NGOs* need to have skills or access to external actors with the necessary skills to analyse open data. Skills acquisition, together with ethics, is needed by *Journalists* and amounts to the most prominent category of their needs, since it covers the whole spectrum of the data journalism process. *Students* need to develop skills and competencies for understanding and using open data. The skills and competencies needed by *students* are usually associated with data literacy and might also go beyond to include 21st century skills or the ability to understand local and global issues, and critical and scientific thinking. Data literacy and digital literacy are also needed to allow the participation of all users, including disadvantaged groups, as found out by the literature on *Government* (Central/Regional). Indeed, not all users possess the same data literacy, and governments should invest in building data literacy and digital equity. Digital equity captures what is needed for actors to participate in and through open government data and is composed of both the acquisition of the necessary skills (defined as digital literacy) and technologies. Finally, data skills are also needed by *Open data intermediaries,* who might not be data-oriented from the beginning (e.g., some CSOs, media, and researchers). Data skills training is, therefore, necessary for them.

Access to data, availability, and findability

The needed open data are not always accessible, available, and findable. Access to data might be uneven across different territories. Therefore, citizens might be impacted by the levels of resources or willingness to adopt and make available open data of their *governments*. The need for data access has two dimensions with reference to local governments. First, reliable data is required for which metadata is crucial for establishing and operating open government data. Second, some groups of citizens can be excluded, with a lack of representation of data and participation from specific groups. Availability of data is also a major need as the information needs of citizens are different from those of other actors. *NGOs* face barriers to data access, with some data unpublished or not fully accessible due to

restrictions. In addition, NGOs need to rely on external actors' advice for finding data. *Artificial users* can help extract metadata from datasets, making them more discoverable and searchable. As data continues to grow and become more complex, the development of efficient metadata extraction techniques is considered paramount in improving the accessibility and usability of this data.

Quality

Data quality can be considered more generally as the need for better quality data, as in the case of *Journalists*, or more specifically, as the need for timely, consistent, and reliable data, as per the literature on *open data intermediaries*. Needs for data quality at the *governmental level* put the emphasis also on representativeness, data protection and validity. Data quality standards do not always ensure that data are bias-free and, more specifically, that data adequately reflect the existence and the characteristics of different user groups, such as minorities or social groups on which data are not collected. The definition of data quality is "socially constructed" and, therefore, it is important to acknowledge and consider that open data that meet existing standards might not satisfy user needs. Also, the issue of data quality should be seen in connection with the lack of data skills of certain user groups. Users might lack critical skills towards data and, therefore, they might not be able to recognise that data are not objective and that they can embed some limitations.

Data infrastructure

Open data users need a reliable data infrastructure. Research on local governments highlighted various issues related to IT Technical Infrastructure and operational data infrastructure, such as the need for technical support, specialised software, and training for utilizing open data, automating the publication and maintenance of data, the need for evolutionary development, and the socio-political barriers to integrating diverse. The data infrastructure also affects *Open data intermediaries* who encounter basic infrastructure issues such as reliable access to the Internet, as well as continuous and uninterrupted access to data providers' application programming interface (API), which is also a concern in developed countries and, more generally, for even users' participation to the open data ecosystem.

Funding

Open data users need adequate funding. *Open data intermediaries* include several non-profit groups, namely, civil society organisations, public organisations, and researchers. These groups mostly rely on sponsorship to conduct their work. Meanwhile, businesses need sustainable business models to continue working with open data. Cost of product deployment, financial sustainability, and competition to hire skilled staff are indeed some of the challenges found in the literature. *Local governments also* need to have the economic resources and the appropriate business models to drive the commitment of the different participants.

Data ethics

Open data raise ethics concerns and the need for knowledge of ethical data practices and impacts, as highlighted by the literature on *journalists*. In the case of open data intermediaries, ethical considerations that need to be considered by open data intermediaries go beyond how to process data ethically (e.g., protecting privacy) to include how to ensure the data-based products do not negatively impact certain communities.

Regulation

NGOs need clear regulations related to open data, with specific reference to licensing standards and privacy concerns. At the *Governmental* level, concerns over privacy, as well as political and organisational barriers, prevent wide access and participation of citizens in environmental governance through open data. Barriers to access prevent a "just" environmental governance, and it is suggested that governments strive for equitable data access to facilitate the participation of diverse groups and satisfy societal needs.

Governance and coordination
The literature on *Governments* suggests that governance through new data principles might emphasise collective ownership, control, and self-determination, as well as the needs of marginalised groups. One leading example is one of the CARE (Collective benefit, Authority to control, Responsibility, Ethics) principles, proposed as a remedy for the flaws of the FAIR (Findable, Accessible, Interoperable, Reusable) principles. New data principles can be a source of inspiration for developing more inclusive data governance in other contexts, such as in environmental justice and open data governance. Furthermore, there is a need for coordination. *Local governments,* as implementers of open data initiatives, need trust and coordination mechanisms so that they can prevent distrust and competition between public and private actors.

Communication
Communication is referenced as a need with different meanings. *NGOs* need communication channels to connect with other actors, such as open data providers, users, and other intermediaries who can help satisfy their needs. For *Journalists*, communication is directed towards their audience (the public), and, as such, they need to be able to present complicated data in a form that can be easily understandable through visualization and, most importantly, storytelling.

When *local governments* act as the initiators of open data initiatives, they need proper communication and coordination so that open data initiatives have the support and engagement of local businesses and communities and align standards, business rules, and architecture. Communication is paramount to enable data flow and drive innovation and value-creation.

Even though categories of needs were identified as transversal to user groups, a deeper knowledge of different contexts of use can lead to more specific needs. For example, "the need for concrete tools for students and educators", "the need for meaningful learning experiences" (*students*), "the need for connecting to Open Data Ecosystems (*students*)", "domain-specific knowledge support" and "domain-specific data standards", as in the case of *open data intermediaries*.

12.2. Limitations
It is important to review some limitations of the report. The most important limitation of the study lies in the fact that literature on open data do not always directly engage with the concept of user needs. Therefore, challenges faced by users are often derived from the tasks, impediments, barriers, and struggles faced by different user types in approaching open data. The report is also limited by the novelty of the field. While the number of studies on open data is exponentially growing, the literature is still fragmented, with research gaps concerning the different user types analysed in this report. Another limitation of the study is the lack of conceptual clarity in the literature. The roles of open data users are not always interpreted in the same way by the literature, and, therefore, what constitutes, for instance, an open data intermediary is not always clear due to a lack of common definitions. Despite its limitations, the report adds to our understanding of the needs of different types of users who have been neglected by previous research and contributes to setting out a research agenda to fill existing gaps in the literature.

12.3. Towards a research agenda
Based on our analysis, we can delineate a research agenda. First, empirical research is needed to both corroborate the results of the literature review and provide new insights into the needs of open data users for which previous research is fragmented, scarce or unfocused, such as NGOs, non-specialist users, and journalists. Second, further research is also needed regarding different stages of the open data lifecycle, considering how the user's needs emerge in a different way based on the multiple roles that different users experience (e.g., governments both sharing and using data). Finally, and most importantly, additional research is needed to understand how to address the user needs from both a technological and a governance perspective. As seen in the report, while some user needs are

expected to be solved by technological solutions, such as improving findability, others are deeply intertwined with governance issues, as in the case of improving literacy, channelling funding, and setting new regulations. To this end, future ODECO studies (deliverable D2.2 and deliverable D2.3) will build on this report and focus on technological and governance measures to satisfy user needs.

References

Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2019). The process of open data publication and reuse: The Process of Open Data Publication and Reuse. *Journal of the Association for Information Science and Technology*, *70*(3), 296–300. https://doi.org/10.1002/asi.24116

Andrason, A., & van Schalkwyk, F. (2017). *Opportune Niches in Data Ecosystems: Open Data Intermediaries in the Agriculture Sector in Ghana* (SSRN Scholarly Paper No. 2949722). Social Science Research Network. https://doi.org/10.2139/ssrn.2949722

Andrews, P., & Da Silva, F. S. C. (2013). Using parliamentary open data to improve participation. In *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance (ICEGOV '13)*. Association for Computing Machinery, New York, NY, USA, 242–249. https://doi.org/10.1145/2591888.2591933.

Araújo, A. C. (2019). Gender and Numbers: Using data from International Women's Day coverage on the sites of three major Brazilian newspapers. In *Brazilian Journalism Research* (Vol. 15, Issue 1, pp. 74–101). Associacao Brasileira de Pesquisadores de Jornalismo. https://doi.org/10.25200/BJR.v15n1.2019.1061

Atenas, J., Havemann, L., & Priego, E. (2015). Open Data as Open Educational Resources: Towards transversal skills and global citizenship. *Open Praxis*, *7*(4), 377. https://doi.org/10.5944/openpraxis.7.4.233

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

Baack, S. (2018). Practically Engaged: The entanglements between data journalism and civic tech. *Digital Journalism*, *6*(6), 673–692. https://doi.org/10.1080/21670811.2017.1375382

Badioze Zaman, H., Baharin, H., & Ahmad, A. (2021). Fusion Technology and Visualisation to Share STEM Data Using PETS Robots (i-COMEL) for Open Data Readiness Amongst Primary School Children. In H. Badioze Zaman, A. F. Smeaton, T. K. Shih, S. Velastin, T. Terutoshi, B. N. Jørgensen, H. Aris, & N. Ibrahim (Eds.), *Advances in Visual Informatics* (Vol. 13051, pp. 184–194). Springer International Publishing. https://doi.org/10.1007/978-3-030-90235-3_16

Bates, J. (2012). "This is what modern deregulation looks like": Co-optation and contestation in the shaping of the UK's Open Government Data Initiative. *The Journal of Community Informatics*, *8*(2). https://doi.org/10.15353/joci.v8i2.3038

Beetham, H. (2019). Learning Activities and Activity Systems. In H. Beetham & R. Sharpe (Eds.), *Rethinking Pedagogy for a Digital Age* (3rd ed., pp. 32–48). Routledge. https://doi.org/10.4324/9781351252805-3

Berntzen, L., Johannessen, M. R., Andersen, K. N., & Crusoe, J. (2019). Parliamentary open data in Scandinavia. *Computers, 8 (*3). https://doi.org/10.3390/computers8030065

Bertot, J. C., Butler, B. S., & Travis, D. M. (2014). Local big data: The role of libraries in building community data infrastructures. *Proceedings of the 15th Annual International Conference on Digital Government Research*, 17–23. https://doi.org/10.1145/2612733.2612762

Bezuidenhout, L. M., Leonelli, S., Kelly, A. H., & Rappert, B. (2017). Beyond the digital divide: Towards a situated approach to open data. *Science and Public Policy*, *44*(4), 464–475. https://doi.org/10.1093/scipol/scw036

Böhm, C., Naumann, F., Freitag, M., George, S., Höfler, N., Köppelmann, M., Lehmann, C., Mascher, A., & Schmidt, T. (2010). Linking open government data: What journalists wish they had known. In *I-SEMANTICS '10: Proceedings of the 6th International Conference on Semantic Systems*. Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/1839707.1839751

boyd, danah, & Crawford, K. (2012). CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Boyles, J. L. (2020). Laboratories for news? Experimenting with journalism hackathons. In *Journalism* (Vol. 21, Issue 9, pp. 1338–1354). SAGE Publications Ltd. https://doi.org/10.1177/1464884917737213

Bozsik, S., Cheng, X., Kuncham, M., & Mitchell, E. (2022). Democratizing Housing Affordability Data: Open Data and Data Journalism in Charlottesville, VA. *2022 Systems and Information Engineering Design Symposium (SIEDS)*, 178–183. https://doi.org/10.1109/SIEDS55548.2022.9799410

Briscoe, G., & Mulligan, C. (2014). *Digital innovation: The hackathon phenomenon*. https://qmro.qmul.ac.uk/xmlui/handle/123456789/11418

Brolcháin, N. Ó., Porwol, L., Ojo, A., Wagner, T., Lopez, E. T., & Karstens, E. (2017). Extending open data platforms with storytelling features. In H. C.C & O. A (Eds.), *ACM International Conference Proceeding Series: Vol. Part F128275* (pp. 48–53). Association for Computing Machinery. https://doi.org/10.1145/3085228.3085283

Brugger, J., Fraefel, M., Riedl, R., Fehr, H., Schöneck, D., & Weissbrod, C. S. (2016). Current barriers to open government data use and visualization by political intermediaries. *Proceedings of the 6th International Conference for E-Democracy and Open Government, CeDEM 2016*, 219–229. https://doi.org/10.1109/CeDEM.2016.18

Cabitza, F., Locoro, A., & Batini, C. (2020). Making Open Data more Personal Through a Social Value Perspective: A Methodological Approach. *Information Systems Frontiers*, *22*(1), 131–148. https://doi.org/10.1007/s10796-018-9854-7

Camaj, L., Martin, J., & Lanosga, G. (2022). The Impact of Public Transparency Infrastructure on Data Journalism: A Comparative Analysis between Information-Rich and Information-Poor Countries. In *Digital Journalism*. Routledge. https://doi.org/10.1080/21670811.2022.2077786

Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., & Jatowt, A. (2020). YAKE! Keyword extraction from single documents using multiple local features. *Information Sciences*, *509*, 257–289. https://doi.org/10.1016/j.ins.2019.09.013

Cañares, M. P. (2014). Opening the local: Full disclosure policy and its impact on local governments in the Philippines. *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, 89–98. https://doi.org/10.1145/2691195.2691214

Cantador, I., Cortés-Cediel, M. E., & Fernández, M. (2020). Exploiting Open Data to analyze discussion and controversy in online citizen participation. *Information Processing & Management*, *57*(5), Business Source Premier. https://doi.org/10.1016/j.ipm.2020.102301

Celis Vargas, A., & Magnussen, R. (2022). A Game-based Approach for Open Data in Education: A Systematic Mapping Review. *European Conference on Games Based Learning*, *16*(1), 139–146. https://doi.org/10.34190/ecgbl.16.1.572

Cepiku, D., & Mastrodascio, M. (2021). Equity in Public Services: A Systematic Literature Review. *Public Administration Review*, *81*(6), 1019–1032. https://doi.org/10.1111/puar.13402

Chan, M., Johnson, P., & Shookner, M. (2016). Assessing the Use of Government Open Data and the Role of Data Infomediaries The Case of Nova Scotia's Community Counts Program. *JeDEM - EJournal of EDemocracy and Open Government*, *8*(1), Article 1. https://doi.org/10.29379/jedem.v8i1.370

Chang, W.-W. (2005). Expatriate Training in International Nongovernmental Organizations: A Model for Research. *Human Resource Development Review*, *4*(4), 440–461. https://doi.org/10.1177/1534484305281035

Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., & Ferro, E. (2018a). The Multiple Life Cycles of Open Data Creation and Use. In Y. Charalabidis, A. Zuiderwijk, C. Alexopoulos, M. Janssen, T. Lampoltshammer, & E. Ferro, *The World of Open Data* (Vol. 28, pp. 11–31). Springer International Publishing. https://doi.org/10.1007/978-3-319-90850-2_2

Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., & Ferro, E. (2018b). *The World of Open Data: Concepts, Methods, Tools and Experiences* (Vol. 28). Springer International Publishing. https://doi.org/10.1007/978-3-319-90850-2

Chattapadhyay, S. (2014). Access and use of government data by research and advocacy organisations in India: A survey of (potential) open data ecosystem. *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, 361–364. https://doi.org/10.1145/2691195.2691262

Chicaiza, J., Piedra, N., Lopez-Vargas, J., & Tovar-Caro, E. (2017). Recommendation of open educational resources. An approach based on linked open data. *2017 IEEE Global Engineering Education Conference (EDUCON)*, 1316–1321. https://doi.org/10.1109/EDUCON.2017.7943018

Chiotaki, D., & Karpouzis, K. (2020). Open and Cultural Data Games for Learning. *International Conference on the Foundations of Digital Games*, 1–7. https://doi.org/10.1145/3402942.3409621

Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, *31*, S10–S17. https://doi.org/10.1016/j.giq.2014.01.003

Cook, K., Çakirlar, C., Goddard, T., DeMuth, R. C., & Wells, J. (2018). Teaching Open Science: Published Data and Digital Literacy in Archaeology Classrooms. *Advances in Archaeological Practice*, *6*(2), 144–156. https://doi.org/10.1017/aap.2018.5

Corbett, J., Templier, M., & Takeda, H. (2018, January 3). The Making of a "Top" Open Data City: A Case Study of Edmonton's Open Data Initiative. *Open Data, Information Processing, and Datification in Government*. Hawaii International Conference on System Sciences 2018. https://doi.org/10.24251/HICSS.2018.308

Coughlan, T. (2020). The use of open data as a material for learning. *Educational Technology Research and Development*, *68*(1), 383–411. https://doi.org/10.1007/s11423-019-09706-y

Cranefield, J., Robertson, O., & Oliver, G. (2014). Value in the mash: Exploring the benefits, barriers and enablers of open data apps. *ECIS 2014 Proceedings - 22nd European Conference on Information Systems*.

da Silva Craveiro, G., & Albano, C. (2017). Open data intermediaries: Coproduction in budget transparency. *Transforming Government: People, Process and Policy*, *11*(1), 119–131. https://doi.org/10.1108/TG-12-2015-0057

Dander, V., & Macgilchrist, F. (2022). School of Data and Shifting Forms of Political Subjectivity. In P. Bettinger (Ed.), *Educational Perspectives on Mediality and Subjectivation* (pp. 45–67). Springer International Publishing. https://doi.org/10.1007/978-3-030-84343-4_3

Davies, T. (2011). *Open Data: Infrastructures and ecosystems*.

Davies, T., & Edwards, D. (2012). Emerging Implications of Open and Linked Data for Knowledge Sharing in Development. *IDS Bulletin*, *43*(5), 117–127. https://doi.org/10.1111/j.1759-5436.2012.00372.x

Davis, K., & Shneyer, E. (2020). Teaching Computational Thinking through Civics: NYC's Hack League. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 1344–1344. https://doi.org/10.1145/3328778.3372636

Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, *33*(1), 15–27. https://doi.org/10.1016/j.giq.2016.01.003

den Haan, D. (2018). *Intermediation as a dissolver of barriers: How intermediaries can help overcome barriers in open data use.* [Master's thesis, Wageningen University]. https://studenttheses.uu.nl/handle/20.500.12932/31148

Dermentzi, E., Zotou, M., Tambouris, E., & Tarabanis, K. (2022). Using the problem based learning method and educational technologies to teach open data: A design-based research approach. *Education and Information Technologies*, *27*(6), 8859–8882. https://doi.org/10.1007/s10639-022-10995-9

Dickinson, A., Lochrie, M., & Egglestone, P. (2015). UKKO: Enriching Persuasive Location based Games with Environmental Sensor Data. *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 493–498. https://doi.org/10.1145/2793107.2810324

Dumpawar, S. (2015). *Open government data intermediaries: Mediating data to drive changes in the built environment* [Master's thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/97994

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

Enaholo, P. (2017). Beyond mere advocacy: CSOs and the role of intermediaries in Nigeria's open data ecosystem. In F. van Schalkwyk, S. Verhulst, G. Magalhaes, J. Pane, & J. Walker (Eds.), *The Social Dynamics of Open Data* (pp. 89–108). African Minds. http://www.africanminds.co.za/the-social-dynamics-of-open-data/

Enaholo, P., & Dina, D. (2020). Journalists and the intermediation of open data: A Nigerian perspective. In *Situating Open Data: Global Trends in Local Contexts* (pp. 31–49). African Minds. https://www.africanminds.co.za/situation-open-data-global-trends-in-local-contexts/#:~:text=Open%20data%20and%20its%20effects,focus%20on%20its%20situated%20use.

Erete, S., Ryou, E., Smith, G., Fassett, K. M., & Duda, S. (2016). Storytelling with Data: Examining the Use of Data by Non-Profit Organizations. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 1273–1283. https://doi.org/10.1145/2818048.2820068

European Commission. (2011, December 12). *Digital Agenda: Commission's Open Data Strategy, Questions & answers* [Text]. European Commission. https://ec.europa.eu/commission/presscorner/detail/en/MEMO_11_891

Evéquoz, F., & Castanheiro, H. (2019). Which Lobby Won the Vote? Visualizing Influence of Interest Groups in Swiss Parliament. In L. I, J. M, L. H, P. A, R. B. M.P, S. H.J, & T. E (Eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 11685 LNCS* (pp. 155–167). Springer. https://doi.org/10.1007/978-3-030-27325-5_12

Faini, F., & Palmirani, M. (2016). The right to know through the freedom of information and open data. *Proceedings of the European Conference on E-Government, ECEG*, *2016-January*, 54–62.

Fernández-Ardèvol, M., & Rosales, A. (2022). Quality Assessment and Biases in Reused Data. *American Behavioral Scientist*. https://doi.org/10.1177/00027642221144855

Flores, M. A. A. (2020). *Creating public value with open government data in Latin America* [Master's thesis]. Westfälische Wilhelms-Universität, Münster.

Foulonneau, M., Martin, S., & Turki, S. (2014). How Open Data Are Turned into Services? In M. Snene & M. Leonard (Eds.), *Exploring Services Science* (Vol. 169, pp. 31–39). Springer International Publishing. https://doi.org/10.1007/978-3-319-04810-9_3

Frank, M., & Waddell, P. (2014). *Solving the Democratic Deficit: The Role of Open Data and Intermediaries*. Conference for E-Democracy and Open Governement, Krems an der Donau, Austria.

Frederickson, H. G. (2021). Thick Social Equity. *Perspectives on Public Management and Governance*, *4*(3), 229–236. https://doi.org/10.1093/ppmgov/gvab012

Fusi, F., Zhang, F., & Liang, J. (2022). Unveiling environmental justice through open government data: Work in progress for most US states. *Public Administration*. https://doi.org/10.1111/padm.12847

Gao, Y., & Janssen, M. (2020). Generating Value from Government Data Using AI: An Exploratory Study. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12219 LNCS*, 319–331. https://doi.org/10.1007/978-3-030-57599-1_24

Gao, Y., & Janssen, M. (2022). The Open Data Canvas–Analyzing Value Creation from Open Data. *Digit. Gov.: Res. Pract.*, *3*(1). https://doi.org/10.1145/3511102

Gasco-Hernandez, M., Martin, E., Reggi, L., Pyo, S., & Luna-Reyes, L. (2018). Promoting the use of open government data: Cases of training and engagement. *GOVERNMENT INFORMATION QUARTERLY*, *35*(2), 233–242. https://doi.org/10.1016/j.giq.2018.01.003

Germano, E. C., Souza, C. A. de, & Sun, V. (2016). Business models adopted by intermediaries in the use of open government data. *REBRAE*, *9*(1), Article 1. https://doi.org/10.7213/rebrae.09.001.AO05

Glassey, O. (2017). Proactive Transparency and Open Data: A Tentative Analysis. In M. Janssen, K. Axelsson, O. Glassey, B. Klievink, R. Krimmer, I. Lindgren, P. Parycek, H. J. Scholl, & D. Trutnev

(Eds.), *Electronic Government* (pp. 314–323). Springer International Publishing. https://doi.org/10.1007/978-3-319-64677-0_26

Golub, K., & Lund, A. (2021). Why open government data? The case of a Swedish municipality. *Journal of Data and Information Science*, *6*(1), 120–138. https://doi.org/10.2478/jdis-2021-0012

González-Zapata, F., & Heeks, R. (2015, May 20). *Understanding Multiple Roles of Intermediaries in Open Government Data*. 13th International Conference on Social Implications of Computers in Developing Countrie, Negombo, Sri Lanka.

Gottron, T., Schmitz, J., & Middleton, S. E. (2014). Focused exploration of geospatial context on Linked Open Data. In: 3rd International Conference on Intelligent Exploration of Semantic Data (ICIESD).

Government of Scotland. (n.d.). *User needs*. Mygov.Scot. https://resources.mygov.scot/content-standards/content-delivery/content-discovery/user-needs/

Government of UK. (2017, March 23). *Learning about users and their needs*. GOV.UK Service Manual. https://www.gov.uk/service-manual/user-research/start-by-learning-user-needs

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies: A typology of reviews, *Maria J. Grant & Andrew Booth*. *Health Information & Libraries Journal*, *26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Gray, J., Bounegru, L., Chambers, L., European Journalism Centre, & Open Knowledge Foundation (Eds.). (2012). *The data journalism handbook* (1st ed). O'Reilly Media.

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data and Society*, *5*(2). https://doi.org/10.1177/2053951718786316

Gupta, K., Sampat, S., Sharma, M., & Rajamanickam, V. (2016). Visualization of election data: Using interaction design and visual discovery for communicating complex insights. *EJournal of EDemocracy and Open Government*, *8*(2), 59–86. https://doi.org/10.29379/jedem.v8i2.422

Hablé, J. (2019). *Value Drivers and Inhibitors in Municipal Open Government Data Ecosystems: An in-depth analysis of stakeholder perceptions on values, barriers and success factors in open government data initiatives in Dutch municipalities* [Master's thesis, Delft University of Technology]. https://repository.tudelft.nl/islandora/object/uuid%3A1347feeb-dce7-44c7-b8b7-3f58f2dd254b

Hammell, R., Bates, C., Lewis, H., Perricos, C., Brett, L., & Branch, D. (2012). *Open data: Driving growth, ingenuity and innovation*. https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-data-driving-growth-ingenuity-and-innovation.pdf

Handler, R. A., & Ferrer Conill, R. (2016). Open Data, Crowdsourcing and Game Mechanics. A case study on civic participation in the digital age. In *Computer Supported Cooperative Work: CSCW: An International Journal* (Vol. 25, Issues 2–3, pp. 153–166). Springer Netherlands. https://doi.org/10.1007/s10606-016-9250-0

Harrison, T. M., Pardo, T. A., & Cook, M. (2012). Creating Open Government Ecosystems: A Research and Development Agenda. *Future Internet*, *4*, 900–928. https://doi.org/10.3390/fi4040900

Hasselwander, M., Kiko, M., & Johnson, T. (2022). Digital civic engagement, open data, and the informal sector: A think piece. *Transportation Research Interdisciplinary Perspectives*, *16*. Scopus. https://doi.org/10.1016/j.trip.2022.100700

Heijs, W. (2022). *User needs by Systematic Elaboration (USE): A theory-based method for user needs analysis, programming and evaluation*. Springer Nature.

Heimstädt, M., Saunderson, F., & Heath, T. (2014). From toddler to teen: Growth of an open data ecosystem: A Longitudinal Analysis of Open Data Developments in the UK. *EJournal of EDemocracy and Open Government*, *6*(2), 123–135. https://doi.org/10.29379/jedem.v6i2.330

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N. (2020). Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. *Current Reviews in Musculoskeletal Medicine*, *13*(1), 69–76. https://doi.org/10.1007/s12178-020-09600-8

Hielkema, H., & Hongisto, P. (2013). Developing the Helsinki Smart City: The Role of Competitions for Open Data Applications. *Journal of the Knowledge Economy*, *4*(2), 190–204. https://doi.org/10.1007/s13132-012-0087-6

Hlabano, M., & Van Belle, J.-P. (2019). Tracing the impact of the city of cape town's open data initiative on communities and development. *IFIP Advances in Information and Communication Technology*, *551*, 284–294. https://doi.org/10.1007/978-3-030-18400-1_23

Hossain, M. A., Dwivedi, Y. K., & Rana, N. P. (2016). State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 14–40. https://doi.org/10.1080/10919392.2015.1124007

Hou, Y., & Wang, D. (2017). Hacking with NPOs: Collaborative Analytics and Broker Roles in Civic Data Hackathons. *Proceedings of the ACM on Human-Computer Interaction*, *1*(CSCW), 1–16. https://doi.org/10.1145/3134688

Hu, J., Li, S., Yao, Y., Yu, L., Yang, G., & Hu, J. (2018). Patent Keyword Extraction Algorithm Based on Distributed Representation for Patent Classification. *Entropy*, *20*(2), Article 2. https://doi.org/10.3390/e20020104

International Open Data Charter. (2015). Principles. *International Open Data Charter*. http://opendatacharter.net/principles/

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, *37*(3), 101493. https://doi.org/10.1016/j.giq.2020.101493

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. https://doi.org/10.1080/10580530.2012.716740

Janssen, M., & Zuiderwijk, A. (2014). Infomediary Business Models for Connecting Open Data Providers and Users. *Social Science Computer Review*, *32*(5), 694–711. https://doi.org/10.1177/0894439314525902

Jarke, J. (2019). Open government for all? Co-creating digital public services for older adults through data walks. *Online Information Review*, *43*(6), 1003–1020. https://doi.org/10.1108/OIR-02-2018-0059

Jarke, J. (2021). Mobile age: Co-creating digital public services with and for older citizens. *Public Administration and Information Technology*, *6*, 53–70. https://doi.org/10.1007/978-3-030-52873-7_4

Jaskiewicz, T., Mulder, I., Morelli, N., & Pedersen, J. (2019). Hacking the hackathon format to empower citizens in outsmarting "smart" cities. *Interaction Design and Architecture(s)*, *43*, 8–29. https://doi.org/10.55612/s-5002-043-001

Johnson, P. A. (2016). Reflecting on the Success of Open Data: How Municipal Government Evaluates their Open Data Programs. *International Journal of E-Planning Research*, *5*(3), 1–12. https://doi.org/10.4018/IJEPR.2016070101

Johnson, P. A., & Greene, S. (2017a). Who are government opendata infomediaries? A preliminary scan and classification of open data users and products. *URISA Journal*, *28*(1), 9–18.

Juell-Skielse, G., Hjalmarsson, A., Juell-Skielse, E., Johannesson, P., & Rudmark, D. (2014). Contests as innovation intermediaries in open data markets. *Information Polity*, *19*(3–4), 247–262. https://doi.org/10.3233/IP-140346

Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*, *30*(4), 508–513. https://doi.org/10.1016/j.giq.2013.05.012

Kassen, M. (2017a). Open data in Kazakhstan: Incentives, implementation and challenges. *Information Technology and People*, *30*(2), 301–323. Scopus. https://doi.org/10.1108/ITP-10-2015-0243

Kassen, M. (2017b). Understanding transparency of government from a Nordic perspective: Open government and open data movement as a multidimensional collaborative phenomenon in Sweden. *Journal of Global Information Technology Management*, *20*(4), 236–275. Scopus. https://doi.org/10.1080/1097198X.2017.1388696

Kassen, M. (2018a). Adopting and managing open data: Stakeholder perspectives, challenges and policy recommendations. In *Aslib Journal of Information Management* (Vol. 70, Issue 5, pp. 518–537). Emerald Group Holdings Ltd. https://doi.org/10.1108/AJIM-11-2017-0250

Kassen, M. (2018b). Open data and its intermediaries: A cross-country perspective on participatory movement among independent developers. *Knowledge Management Research and Practice*, *16*(3), 327–342. https://doi.org/10.1080/14778238.2018.1471377

Kassen, M. (2019). Open data and e-government–related or competing ecosystems: A paradox of open government and promise of civic engagement in Estonia*. *Information Technology for Development*, *25*(3), 552–578. https://doi.org/10.1080/02681102.2017.1412289

Kim, D. (2018). *Data + disasters: Rethinking the role of Metro Boston's data intermediaries in disaster management* [Master's thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/117828

Kitchin, R. (2021). *The data revolution: Big data, open data, data infrastructures and their consequences* (2nd edition). Sage Publications Ltd.

Kleiman, F., Jansen, S. J. T., Meijer, S., & Janssen, M. (2023). Understanding civil servants' intentions to open data: Factors influencing behavior to disclose data. *Information Technology & People*. https://doi.org/10.1108/ITP-12-2020-0893

Klímek, J., Kučera, J., Nečaský, M., & Chlapek, D. (2018). Publication and usage of official Czech pension statistics Linked Open Data. In *Journal of Web Semantics* (Vol. 48, pp. 1–21). Elsevier B.V. https://doi.org/10.1016/j.websem.2017.09.002

Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. *ACM SIGKDD Explorations Newsletter*, *2*(1), 1–15. https://doi.org/10.1145/360402.360406

Krotoski, A. K. (2012). Data-driven research: Open data opportunities for growing knowledge, and ethical issues that arise. In *Insights: The UKSG Journal* (Vol. 25, Issue 1, pp. 28–32). United Kingdom Serials Group. https://doi.org/10.1629/2048-7754.25.1.28

Kujala, S. (2002). *User Studies: A Practical Approach to User Involvement for Gathering User Needs and Requirements* [PhD Thesis, Helsinki University of Technology]. https://aaltodoc.aalto.fi/bitstream/handle/123456789/2181/isbn9512259001.pdf?sequence=1&isAllowed=y

Kujala, S., Kauppinen, M., & Rekola, S. (2001). *Bridging the gap between user needs and user requirements*. 45–50.

Kumpula-Natri, M. (2021, April 13). Building a European data strategy. *The Parliament Magazine*. https://www.theparliamentmagazine.eu/news/article/building-a-european-data-strategy

Lawson, B. T. (2022). Realizing the benefits of open government data: Journalists' coverage of the NHS winter crisis, 2016-17. In *INFORMATION SOCIETY* (Vol. 38, Issue 1, pp. 25–35). TAYLOR & FRANCIS INC. https://doi.org/10.1080/01972243.2021.1998274

Le Borgne, Y.-A., Homolova, A., & Bontempi, G. (2016). OpenTED browser: Insights into European Public Spendings. In Gama J., Zliobaite I., Zliobaite I., & Gavalda R. (Eds.), *CEUR Workshop Proc.* (Vol. 1831). CEUR-WS; Scopus. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85019195878&partnerID=40&md5=fa04428595f7474afce7eac77d25e760

Lee, S. C., Nadri, C., Sanghavi, H., & Jeon, M. (2022). Eliciting User Needs and Design Requirements for User Experience in Fully Automated Vehicles. *International Journal of Human–Computer Interaction*, *38*(3), 227–239. https://doi.org/10.1080/10447318.2021.1937875

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, *6*(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100

Lodato, T. J., & DiSalvo, C. (2016). Issue-oriented hackathons as material participation. *New Media & Society*, *18*(4), 539–557. https://doi.org/10.1177/1461444816629467

Longo, J. (2011). Open Data: Digital-Era Governance Thoroughbred or New Public Management Trojan Horse? *Public Policy and Governance Review*, *2*.

Maail, A. G. (2017). The relational impact of open data intermediation: Experience from Indonesia and the Philippines. In F. van Schalkwyk, S. Verhulst, G. Magalhaes, J. Pane, & J. Walker (Eds.), *The Social Dynamics of Open Data* (pp. 153–166). African Minds. http://www.africanminds.co.za/the-social-dynamics-of-open-data/

Martin, J. A., Camaj, L., & Lanosga, G. (2022). Scrape, Request, Collect, Repeat: How Data Journalists Around the World Transcend Obstacles to Public Data. In *Journalism Practice*. Routledge. https://doi.org/10.1080/17512786.2022.2142837

Mayer-Schönberger, V., & Zappia, Z. (2011). *Participation and power: Intermediaries of open data*. 1st Berlin Symposium on Internet and Society, Berlin, Germany. https://www.hiig.de/wp-content/uploads/2017/08/participation_and_power.pdf

Meijer, A., & Potjer, S. (2018). Citizen-generated open data: An explorative analysis of 25 cases. *Government Information Quarterly*, *35*(4), 613–621. https://doi.org/10.1016/j.giq.2018.10.004

Mendel, S. C. (2013). *How Nonprofit Organizations Create Public Value*. Urban Publications. https://engagedscholarship.csuohio.edu/urban_facpub/686

Meng, A. (2016). *The social impact of open government data* [PhD thesis, Georgia Institute of Technology]. https://smartech.gatech.edu/handle/1853/56292

Meng, A., & DiSalvo, C. (2018). Grassroots resource mobilization through counter-data action. *Big Data and Society*, *5*(2). https://doi.org/10.1177/2053951718796862

Meng, A., DiSalvo, C., Tsui, L., & Best, M. (2019). The social impact of open government data in Hong Kong: Umbrella Movement protests and adversarial politics. *The Information Society*, *35*(4), 216–228. https://doi.org/10.1080/01972243.2019.1613464

Milojevic-Dupont, N., Hans, N., Kaack, L. H., Zumwald, M., Andrieux, F., de Barros Soares, D., Lohrey, S., Pichler, P.-P., & Creutzig, F. (2020). Learning from urban form to predict building heights. *PLoS ONE*, *15*, 1–22. Academic Search Premier. https://doi.org/10.1371/journal.pone.0242010

Najafabadi, M. M., & Cronemberger, F. A. (2022). Systemic effects of an open government program on data quality: The case of the New York State's Food Protection program area. *Transforming Government: People, Process and Policy*. https://doi.org/10.1108/TG-11-2021-0194

Navalkha, C. (2021). *Data for Housing Justice: Examining Activists' Use of Open Government Data for Housing Justice in Boston, MA and New York, NY* [Master's thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/138973

Nikitas, A., Michalakopoulou, K., Njoya, E. T., & Karampatzakis, D. (2020). Artificial Intelligence, Transport and the Smart City: Definitions and Dimensions of a New Mobility Era. *Sustainability*, *12*(7), Article 7. https://doi.org/10.3390/su12072789

Norman, D. (2013). *The Design of Everyday Things: Revised and Expanded Edition*. Hachette UK.

Ocaña, M. G., Opdahl, A. L., & Dang-Nguyen, D.-T. (2021). Emerging News task: Detecting emerging events from social media and news feeds. In Working Notes Proceedings of the MediaEval 2021 Workshop, Online, 13-15 December 2021. CEUR Workshop Proceedings (Vol. 3181). https://ceur-ws.org/Vol-3181/paper7.pdf

Okoli, C. (2015). A Guide to Conducting a Standalone Systematic Literature Review. *Communications of the Association for Information Systems*, *37*. https://doi.org/10.17705/1CAIS.03743

Olausson, K. R. a. A. (2016). *A step towards aligning supply and demand? - User involvement in supply of open data among ten Dutch public sector bodies* [Master Thesis]. https://studenttheses.uu.nl/handle/20.500.12932/24941

On the Full and Open Exchange of Scientific Data. (1995). National Academies Press. https://doi.org/10.17226/18769

Open Knowledge Foundation. (n.d.). The Open Definition. *Open Definition*. http://opendefinition.org/

Ossom-Williamson, P., Williams, I. M., Kim, K., & Kindratt, T. B. (2021). Reporting and availability of COVID-19 demographic data by US health departments (April to October 2020): Observational study. *JMIR Public Health and Surveillance*, *7*(4). https://doi.org/10.2196/24288

Overby, A. T., Dyckman, C. S., Ogletree, S., Fouch, N., Lauria, M., White, D. L., Baldwin, R. F., Amidon, K., & Crum, D. (2022). Evaluating Conservation Easement and Land Records Data in Rapidly Growing Counties in the United States and the Implications for Environmental Governance.

Society and Natural Resources, *35*(2), 167–184. https://doi.org/10.1080/08941920.2021.2008076

Paley, A., Zhao, A. L. L., Pack, H., Servantez, S., Adler, R. F., Sterbentz, M., Pah, A., Schwartz, D., Barrie, C., Einarsson, A., & Hammond, K. (2021). From data to information: Automating data science to explore the U.S. court system. In *Proceedings of the 18th International Conference on Artificial Intelligence and Law, ICAIL 2021* (pp. 119–128). Association for Computing Machinery, Inc. https://doi.org/10.1145/3462757.3466100

Palova, D., & Vejacka, M. (2022). Experience with Open Data in Project Based Learning. *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, 689–694. https://doi.org/10.23919/MIPRO55190.2022.9803595

Park, S., & Gil-Garcia, J. R. (2017). Understanding Transparency and Accountability in Open Government Ecosystems: The Case of Health Data Visualizations in a State Government. *Proceedings of the 18th Annual International Conference on Digital Government Research*, 39–47. https://doi.org/10.1145/3085228.3085318

Petricek, T. (2022). The Gamma: Programmatic Data Exploration for Non-programmers. *2022 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 1–7. https://doi.org/10.1109/VL/HCC53370.2022.9833134

Poikola, A., Kola, P., & Hintikka. (2011). *Public data: An introduction to opening information resources*. Ministry of Transport and Communications. https://core.ac.uk/download/pdf/198191501.pdf

Pollock, R. (2011, March 31). *Building the (Open) Data Ecosystem*. Open Knowledge Foundation Blog. https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/

Porlezza, C., & Splendore, S. (2019). From Open Journalism to Closed Data: Data Journalism in Italy. *Digital Journalism*, *7*(9), 1230–1252. https://doi.org/10.1080/21670811.2019.1657778

Prieto, L. M., Rodríguez, A. C., & Pimiento, J. (2012). Implementation framework for open data in Colombia. *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance*, 14–17. https://doi.org/10.1145/2463728.2463732

Prodromou, T. (Ed.). (2017). *Data Visualization and Statistical Literacy for Open and Big Data:* IGI Global. https://doi.org/10.4018/978-1-5225-2512-7

Publications Office of the European Union. (2022). *Open data maturity report 2021.* Publications Office. https://data.europa.eu/doi/10.2830/394148

Qian, Y., Jia, C., & Liu, Y. (2021). Bert-Based Text Keyword Extraction. *Journal of Physics: Conference Series*, *1992*(4), 042077. https://doi.org/10.1088/1742-6596/1992/4/042077

Radchenko, I., & Sakoyan, A. (2014). The View on Open Data and Data Journalism: Cases, Educational Resources and Current Trends. In D. I. Ignatov, M. Yu. Khachay, A. Panchenko, N. Konstantinova, & R. E. Yavorsky (Eds.), *Analysis of Images, Social Networks and Texts* (Vol. 436, pp. 47–54). Springer International Publishing. https://doi.org/10.1007/978-3-319-12580-0_4

Radchenko, I., & Sakoyan, A. (2016). On some russian educational projects in open data and data journalism. In Mouromtsev D. & d'Aquin M. (Eds.), *Lect. Notes Comput. Sci.* (Vol. 9500, p. 165). Springer Verlag. https://doi.org/10.1007/978-3-319-30493-9_8

Rambaree, K. (2018). Abductive Thematic Network Analysis (ATNA) Using ATLAS-ti. In L. Moutinho & M. Sokele (Eds.), *Innovative Research Methodologies in Management* (pp. 61–86). Springer International Publishing. https://doi.org/10.1007/978-3-319-64394-6_4

Ricker, B., Cinnamon, J., & Dierwechter, Y. (2020). When open data and data activism meet: An analysis of civic participation in Cape Town, South Africa. *Canadian Geographer*, *64*(3), 359–373. https://doi.org/10.1111/cag.12608

Ridgway, J. (2016). Implications of the Data Revolution for Statistics Education: The Data Revolution and Statistics Education. *International Statistical Review*, *84*(3), 528–549. https://doi.org/10.1111/insr.12110

Rind, A., Pfahler, D., Niederer, C., & Aigner, W. (2016). Exploring media transparency with multiple views. In: W. Aigner, G. Schmiedl, K. Blumenstein, M. Zeppelzauer (eds.): Proceedings of the 9th Forum Media Technology 2016, St. P ̈olten, Austria, 24-11-2016, CEUR Workshop Proceedings (Vol. 1734). Available online: https://ceur-ws.org/Vol-1734/fmt-proceedings-2016-paper8.pdf

Robinson, P., & Mather, L. W. (2017). Open Data Community Maturity: Libraries as Civic Infomediaries. *URISA Journal*, *28*(1), 31–39.

Rogers, S. (2008, December 18). Turning official figures into understandable graphics, at the press of a button. *The Guardian*. https://www.theguardian.com/help/insideguardian/2008/dec/18/unemploymentdata

Romero, M., Usart, M., & Ott, M. (2015). Can Serious Games Contribute to Developing and Sustaining 21st Century Skills? *Games and Culture*, *10*(2), 148–177. https://doi.org/10.1177/1555412014548919

Ruijer, E., Grimmelikhuijsen, S., & Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. *Government Information Quarterly*, *34*(1), 45–52. https://doi.org/10.1016/j.giq.2017.01.001

Ruijer, E., Porumbescu, G., Porter, R., & Piotrowski, S. (2022). Social equity in the data era: A systematic literature review of data-driven public service research. *Public Administration Review*, puar.13585. https://doi.org/10.1111/puar.13585

Runeson, P., Olsson, T., & Linåker, J. (2021). Open Data Ecosystems—An empirical investigation into an emerging industry collaboration concept. *Journal of Systems & Software*, *182*, N.PAG-N.PAG. Business Source Premier. https://doi.org/10.1016/j.jss.2021.111088

Saddiqa, M., Kirikova, M., & Pedersen, J. M. (2019). Enterprise Architecture Oriented Requirements Engineering for Open Data Usage in Schools. In M. Pańkowska & K. Sandkuhl (Eds.), *Perspectives in Business Informatics Research* (Vol. 365, pp. 135–147). Springer International Publishing. https://doi.org/10.1007/978-3-030-31143-8_10

Saddiqa, M., Magnussen, R., Larsen, B., & Pedersen, J. M. (2021a). Digital innovation in education: Perspectives, opportunities and challenges of educational open data and sensor data. *CEUR Workshop Proceedings*, *2991*, 74–83.

Saddiqa, M., Magnussen, R., Larsen, B., & Pedersen, J. M. (2021b). Open Data Interface (ODI) for secondary school education. *Computers & Education*, *174*, 104294. https://doi.org/10.1016/j.compedu.2021.104294

Saddiqa, M., Rasmussen, L., Magnussen, R., Larsen, B., & Pedersen, J. M. (2019). Bringing open data into danish schools and its potential impact on school pupils. *Proceedings of the 15th International Symposium on Open Collaboration*, 1–10. https://doi.org/10.1145/3306446.3340821

Safarov, I., Meijer, A., & Grimmelikhuijsen, S. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity*, *22*, 1–24. https://doi.org/10.3233/IP-160012

Sandoval-Martín, T., & La-Rosa, L. (2018). Big data as a differentiating sociocultural element of data journalism: The perception of data journalists and experts. *Communication and Society*, *31*(4), 193–209. Scopus. https://doi.org/10.15581/003.31.4.193-209

Sangiambut, S., & Sieber, R. (2017). The civic open data and crowdsourcing app ecosystem: Actors, materials, and interventions. *URISA Journal*, *28*(1), 49–62.

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* (Vol. 59, pp. 145–177). Springer International Publishing. https://doi.org/10.1007/978-3-031-24193-2_6

Saxena, S., & Muhammad, I. (2018). Barriers to use open government data in private sector and NGOs in Pakistan. *Information Discovery and Delivery*, *46*(1), 67–75. https://doi.org/10.1108/IDD-05-2017-0049

Schrotter, G., & Hürzeler, C. (2020). The Digital Twin of the City of Zurich for Urban Planning. *PFG - Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, *88*(1), 99–112. https://doi.org/10.1007/s41064-020-00092-2

Schwoerer, K. (2022). Whose open data is it anyway? An exploratory study of open government data relevance and implications for democratic inclusion. *Information Polity*, *1*(1). Scopus. https://doi.org/10.3233/IP-220008

Selwyn, N., Henderson, M., & Chao, S.-H. (2017). The possibilities and limitations of applying 'open data' principles in schools. *Cambridge Journal of Education*, *47*(2), 167–187. https://doi.org/10.1080/0305764X.2016.1143449

Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W., & schraefel, m. c. (2012). Linked Open Government Data: Lessons from Data.gov.uk. *IEEE Intelligent Systems*, *27*(3), 16–24. https://doi.org/10.1109/MIS.2012.23

Shaharudin, A., Van Loenen, B., & Janssen, M. (2023). Towards a Common Definition of Open Data Intermediaries. *Digital Government: Research and Practice*. https://doi.org/10.1145/3585537

Shehu, V., Mijushkovic, A., & Besimi, A. (2016). Empowering data driven journalism in Macedonia. In *ACM International Conference Proceeding Series*. Association for Computing Machinery. https://doi.org/10.1145/2955129.2955187

Shibuya, Y., Lai, C.-M., Hamm, A., Takagi, S., & Sekimoto, Y. (2022). Do open data impact citizens' behavior? Assessing face mask panic buying behaviors during the Covid-19 pandemic. *Scientific Reports*, *12*(1). Scopus. https://doi.org/10.1038/s41598-022-22471-y

Siriaraya, P., Kiriu, T., Kawai, Y., & Nakajima, S. (2018). Using Open Data to Create Smart Auditory based Pervasive Game Environments. *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, 611–617. https://doi.org/10.1145/3270316.3271524

Smith, D. A. (2016). Online interactive thematic mapping: Applications and techniques for socio-economic research. *Computers, Environment and Urban Systems*, *57*, 106–117. Scopus. https://doi.org/10.1016/j.compenvurbsys.2016.01.002

Song, M., Jiang, H., Shi, S., Yao, S., Lu, S., Feng, Y., Liu, H., & Jing, L. (2023). *Is ChatGPT A Good Keyphrase Generator? A Preliminary Study* (arXiv:2303.13001). arXiv. https://doi.org/10.48550/arXiv.2303.13001

Splendore, S., Di Salvo, P., Eberwein, T., Groenhart, H., Kus, M., & Porlezza, C. (2016). Educational strategies in data journalism: A comparative study of six European countries. *Journalism*, *17*(1), 138–152. https://doi.org/10.1177/1464884915612683

Susha, I., Grönlund, Å., & Janssen, M. (2015). Organizational measures to stimulate user engagement with open data. *Transforming Government: People, Process and Policy*, *9*(2), 181–206. https://doi.org/10.1108/TG-05-2014-0016

Svärd, P. (2018). Public Information Directive (PSI) implementation in two Swedish municipalities. *Records Management Journal*, *28*(1), 2–17. Scopus. https://doi.org/10.1108/RMJ-04-2016-0012

Tabary, C., Provost, A.-M., & Trottier, A. (2016). Data journalism's actors, practices and skills: A case study from Quebec. *Journalism*, *17*(1), 66–84. https://doi.org/10.1177/1464884915593245

Thakuriah, P. (Vonu), Dirks, L., & Keita, Y. M. (2017). Digital Infomediaries and Civic Hacking in Emerging Urban Data Initiatives. In P. (Vonu) Thakuriah, N. Tilahun, & M. Zellner (Eds.), *Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics* (pp. 189–207). Springer International Publishing. https://doi.org/10.1007/978-3-319-40902-3_11

Tuunanen, T. (2021). *Tool Support for Open Source Software License Compliance: The First Two Decades of the Millennium* [University of Jyväskylä]. https://jyx.jyu.fi/handle/123456789/75272

Vallejo-Figueroa, S., Rodriguez-Artacho, M., Castro-Gil, M., & Cristobal, E. S. (2018). Using text mining and linked open data to assist the mashup of educational resources. *2018 IEEE Global Engineering Education Conference (EDUCON)*, 1606–1611. https://doi.org/10.1109/EDUCON.2018.8363427

Van Audenhove, L., Van den Broeck, W., & Mariën, I. (2020). Data literacy and education: Introduction and the challenges for our field. *Journal of Media Literacy Education*, *12*(3), 1–5. https://doi.org/10.23860/JMLE-2020-12-3-1

Van Der Bijl Brouwer, M., & Dorst, K. (2014, January 1). How deep is deep? A four-layer model of insights into human needs for design innovation. In: *Colors of Care*. Proceedings of the 9th International Conference on Design & Emotion, Bogota. https://opus.lib.uts.edu.au/handle/10453/33538

van Loenen, B. (2018). Towards a user-oriented open data strategy. *Open Data Exposed*, 33–53.

van Loenen, B., Vancauwenberghe, G., & Crompvoets, J. (Eds.). (2018). *Open Data Exposed*. T.M.C. Asser Press. https://link.springer.com/book/10.1007/978-94-6265-261-3

Van Loenen, B., Vancauwenberghe, G., Crompvoets, J., & Dalla Corte, L. (2018). Open Data Exposed. In *Open Data Exposed* (pp. 1–10).

Van Loenen, B., Zuiderwijk, A., Vancauwenberghe, G., Lopez-Pellicer, F. J., Mulder, I., Alexopoulos, C., Magnussen, R., Saddiqa, M., Dulong de Rosnay, M., Crompvoets, J., Polini, A., Re, B., & Casiano Flores, C. (2021). Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda. *JeDEM - EJournal of EDemocracy and Open Government*, *13*(2), 1–27. https://doi.org/10.29379/jedem.v13i2.644

van Schalkwyk, F., Chattapadhyay, S., Cañares, M., & Andrason, A. (2015). *Open Data Intermediaries in Developing Countries* [Working Paper]. World Wide Web Foundation, Washington, D.C. https://idl-bnc-idrc.dspacedirect.org/handle/10625/56288

van Schalkwyk, F., Willmers, M., & McNaughton, M. (2016). Viscous Open Data: The Roles of Intermediaries in an Open Data Ecosystem. *Information Technology for Development*, *22*(sup1), 68–83. https://doi.org/10.1080/02681102.2015.1081868

Vargianniti, I., & Karpouzis, K. (2020). Using Big and Open Data to Generate Content for an Educational Game to Increase Student Performance and Interest. *Big Data and Cognitive Computing*, *4*(4), 30. https://doi.org/10.3390/bdcc4040030

Väyrynen, H., Helander, N., & Vasell, T. (2017). KNOWLEDGE MANAGEMENT FOR OPEN INNOVATION: COMPARING RESEARCH RESULTS BETWEEN SMEs AND LARGE COMPANIES. *International Journal of Innovation Management*, *21*(05), 1740004. https://doi.org/10.1142/S1363919617400047

Veglis, A., & Bratsas, C. (2017). Reporters in the age of data journalism. *JOURNAL OF APPLIED JOURNALISM & MEDIA STUDIES*, *6*(2), 225–244. https://doi.org/10.1386/ajms.6.2.225_1

Walravens, N., Ballon, P., Van Compernolle, M., & Borghys, K. (2021). Data Ownership and Open Data: The Potential for Data-Driven Policy Making. In G. Concilio, P. Pucci, L. Raes, & G. Mareels (Eds.), *The Data Shake* (pp. 19–34). Springer International Publishing. https://doi.org/10.1007/978-3-030-63693-7_2

Walter, M., Lovett, R., Maher, B., Williamson, B., Prehn, J., Bodkin-Andrews, G., & Lee, V. (2021). Indigenous Data Sovereignty in the Era of Big Data and Open Data. *Australian Journal of Social Issues*, *56*(2), 143–156. Scopus. https://doi.org/10.1002/ajs4.141

Wang, T. (2016). *Why Big Data needs Thick Data. Medium*. https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7

Welle Donker, F., & van Loenen, B. (2017). How to assess the success of the open data ecosystem? *International Journal of Digital Earth*, *10*(3), 284–306. https://doi.org/10.1080/17538947.2016.1224938

Wilson, B., & Chakraborty, A. (2019). Planning Smart(er) Cities: The Promise of Civic Technology. *Journal of Urban Technology*, *26*(4), 29–51. Scopus. https://doi.org/10.1080/10630732.2019.1631097

Wilson, B., & Cong, C. (2021). Beyond the supply side: Use and impact of municipal open data in the U.S. *Telematics and Informatics*, *58*, 101526. https://doi.org/10.1016/j.tele.2020.101526

Wilson, M. G., Lavis, J. N., Travers, R., & Rourke, S. B. (2010). Community-based knowledge transfer and exchange: Helping community-based organizations link research to action. *Implementation Science*, *5*(1), 33. https://doi.org/10.1186/1748-5908-5-33

Windhager, F., Mayr, E., Schreder, G., & Smuc, M. (2016). Linked information visualization for linked open government data: A visual synthetics approach to governmental data and knowledge collections. In *EJournal of eDemocracy and Open Government* (Vol. 8, Issue 2, pp. 87–116). Department for E-Governance and Administration. https://doi.org/10.29379/jedem.v8i2.436

Wolff, A., Cavero Montaner, J. J., & Kortuem, G. (2016). Urban Data in the primary classroom: Bringing data literacy to the UK curriculum. *The Journal of Community Informatics*, *12*(3). https://doi.org/10.15353/joci.v12i3.3278

Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an Understanding of Data Literacy for a Data-driven Society. *The Journal of Community Informatics*, *12*(3). https://doi.org/10.15353/joci.v12i3.3275

Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education and Research*, *39*(1), 93–112. https://doi.org/10.1177/0739456X17723971

Yoon, A., & Copeland, A. (2020). Toward community-inclusive data ecosystems: Challenges and opportunities of open data for community-based organizations. *Journal of the Association for Information Science and Technology*, *71*(12), 1439–1454. https://doi.org/10.1002/asi.24346

Yoon, A., Copeland, A., & McNally, P. J. (2018). Empowering communities with data: Role of data intermediaries for communities' data utilization. *Proceedings of the Association for Information Science and Technology*, *55*(1), 583–592. https://doi.org/10.1002/pra2.2018.14505501063

Young, A., & Verhulst, S. (2017). *Aclímate Colombia: Open data to improve agricultural resiliency*. GovLab - Open Data's Impact. https://doi.org/10.1007/978-3-319-63743-3_6

Zapata, C. B., & Santana, R. C. G. (2015). Transparency and open data in the classroom: A pedagogical exercise to construct civic awareness about access to public digital data in Brazil. *International Journal of Electronic Governance*, *7*(4), 313. https://doi.org/10.1504/IJEG.2015.074330

Zhang, J., Li, S., & Wang, Y. (2021). Shaping a Smart Transportation System for Sustainable Value Co-Creation. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-021-10139-3

Zhang, S. (2022). Teaching Spatial Literacy in an Equity-Informed Public Engagement: An Academic Library's Case Study. *Journal of Map and Geography Libraries*. Scopus. https://doi.org/10.1080/15420353.2022.2147626

Zhu, Z., Wulder, M. A., Roy, D. P., Woodcock, C. E., Hansen, M. C., Radeloff, V. C., Healey, S. P., Schaaf, C., Hostert, P., Strobl, P., Pekel, J.-F., Lymburner, L., Pahlevan, N., & Scambos, T. A. (2019). Benefits of the free and open Landsat data policy. *Remote Sensing of Environment*, *224*, 382–385. https://doi.org/10.1016/j.rse.2019.02.016

Zickler, C., Di Donna, V., Kaiser, V., Al-Khodairy, A., Kleih, S., Kübler, A., Malavasi, M., Mattia, D., Mongardi, S., & Neuper, C. (2009). BCI applications for people with disabilities: Defining user needs and user requirements. *Assistive Technology from Adapted Equipment to Inclusive Environments, AAATE*, *25*, 185–189.

Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, *19*(1–2), 17–33. https://doi.org/10.3233/IP-140329

Zuiderwijk, A., Janssen, M., Poulis, K., & van de Kaa, G. (2015). Open data for competitive advantage: Insights from open data use by companies. *Proceedings of the 16th Annual International Conference on Digital Government Research*, 79–88. https://doi.org/10.1145/2757401.2757411

Zuiderwijk, A., Volten, C., Kroesen, M., & Gill, M. (2018). Motivation perspectives on opening up municipality data: Does municipality size matter? *Information (Switzerland)*, *9*(11). Scopus. https://doi.org/10.3390/info9110267