

A THEMATIC REVIEW ON RESEARCH DATA MANAGEMENT IN CITIZEN SCIENCE PROJECTS

Cik Ramlah Che Jaafar ^{1,3}
Asmadi Mohammed Ghazali ^{2*}
Mad Khir Johari Abdullah Sani ¹

¹ Faculty of Information Science, Universiti Teknologi MARA (UiTM) Puncak Perdana, MALAYSIA,

² Faculty of Information Science, Universiti Teknologi MARA (UiTM), Kedah Branch, MALAYSIA,

³ Hamzah Sendut Library, Universiti Sains Malaysia, MALAYSIA

*Corresponding author: asmadi615@uitm.edu.my

Article history

Received date : 1-3-2026
Revised date : 2-3-2026
Accepted date : 20-4-2026
Published date : 11-6-2026

To cite this document:

Che Jaafar, C. R., Mohammed Ghazali, A., & Abdullah Sani, M. K. J. (2026). A thematic review on research data management in citizen science projects. *Journal of Islamic, Social, Economics and Development (JISED)*, 11 (83), 272 – 289.

Abstract: *Citizen science (CS) has emerged as a globally recognized participatory approach to scientific inquiry, engaging non-professionals in data collection and analysis. As the volume and complexity of citizen-generated data grow, effective Research Data Management (RDM) becomes essential to ensure data quality, ethical governance, and long-term usability. Despite increasing attention, a gap remains in the consolidated understanding of how RDM is applied, evaluated, and discussed across diverse citizen science contexts. Therefore, this thematic review aims to explore current trends in RDM within CS projects, focusing on peer-reviewed literature published between 2021 and 2025. The study systematically searched the Scopus and Web of Science (WoS) databases and identified 26 relevant journal articles, which were analyzed using ATLAS.ti 24. Four key themes emerged: (i) increasing emphasis on data quality assurance and validation; (ii) movement toward ethical data governance and open science alignment; (iii) integration of digital infrastructures for scalable data management; and (iv) capacity building and empowerment through data literacy and participation. These findings reveal a shift from informal, ad hoc data practices to more structured, ethical, and community-informed approaches in managing CS data. This review contributes a timely synthesis of evolving RDM practices in CS projects and offers practical implications for researchers, practitioners, and policymakers. It also identifies future research opportunities, including context-specific applications, digital ethics, and inclusive frameworks for participatory data stewardship.*

Keywords: *Citizen Science, Research Data Management, Open Science, Thematic Review*

Introduction

Citizen Science (CS) has emerged as a key approach for involving the public in research, offering valuable contributions across diverse fields. Understanding the thematic landscape of Research Data Management (RDM) in CS involves examining data sharing practices, data quality measures, and infrastructure that support collaboration between scientists and the public. Effective data management ensures the quality and interoperability of CS-generated data. Loglisci et al. (2024) emphasize the need for interoperability standards and infrastructure, especially in environmental monitoring, to streamline data sharing and enhance collaboration. Steven et al. (2019) note the growing use of structured survey methods to improve data quality and reliability for broader repositories. In addition, Rowbotham et al. (2023) highlight the increasing use of structured survey methods to improve data quality and reliability for broader repositories. Lotfian et al. (2022) observe that heterogeneous datasets, due to varied participant backgrounds, require multiple validation mechanisms to maintain reliability. Frameworks like the DataONE life cycle (Shwe, 2020) help systematize data collection, quality assurance, and preservation. Engaging local stakeholders is also critical. Mandeville et al. (2022) argue that using knowledge of CS trends improves participation and aligns data with monitoring goals. Moreover, Mason et al. (2024) highlight the importance of open data policies, while Guerrini et al. (2021) emphasize collaboration and shared ownership. Therefore, the thoroughness of CS projects is influenced by various factors, including structured methodologies for data collection, engagement with community stakeholders, and the establishment of effective frameworks for data interoperability. Hence, this paper aims to conduct a thematic review of relevant literature from Scopus and Web of Science (WoS), to address the following research questions (RQ): What are the current trends in research data management related to citizen science found in the literature from 2021 to 2025?

Materials And Methods

The study employed a thematic review (TR) method, which was introduced by Mohd Zairul (2021); Zairul (2020, 2022); Zairul et al. (2023) and has been protected by copyright, under the registration number CRLY2023W02032 (Zairul, 2023). TR is utilized because the methodology of this study involves applying a thematic analysis procedure in conducting a literature review. Clarke and Braun (2013) define thematic analysis as a process of identifying the patterns and constructing themes over a thorough reading of the subject. Furthermore, patterns are categorized to understand trends, particularly regarding RDM in CS. This study applies Thematic Review Flowz, a five-stage methodological framework, in conducting a systematic TR in the context of RDM in CS, as depicted in **Figure 1**. The process begins with formulating an RQ that is appropriately scoped to ensure relevance and focus. In the second stage, article screening is conducted using at least two databases using Boolean operators and carefully selected keywords, to facilitate the identification of potentially relevant literature. The third phase, article filtering, involves applying inclusion and exclusion criteria to ensure that only studies aligning with the RQ are retained. In the fourth stage, the cleaning process involves verifying metadata accuracy and organizing references using Mendeley. The final phase, synthesis and reporting, involves integrating quantitative and qualitative data to comprehensively address the RQ. This stage also includes thematic coding and analysis using ATLAS.ti. Overall, the framework provides a structured, transparent, and replicable approach in conducting TR, ensuring methodological rigour and the production of credible and insightful research outcomes.

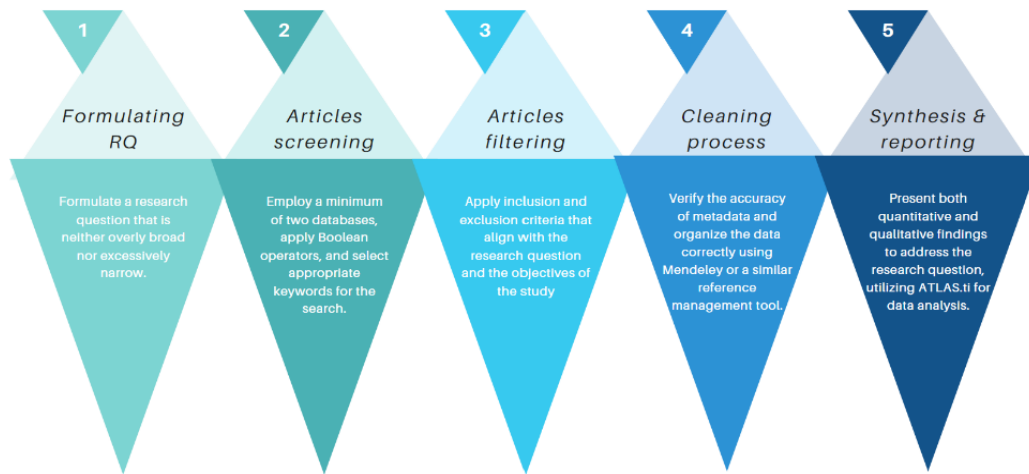


Figure 1: Thematic Review Flowz (Zairul, 2023)

A structured literature search was conducted using two major academic databases: SCOPUS and WoS. As shown in **Table 1**, the keywords used in SCOPUS are: (TITLE-ABS-KEY ("Data Management") AND TITLE-ABS-KEY ("Citizen Science")). To ensure recency and relevance, the search was limited to articles published between 2021 and 2025 by applying the filter PUBYEAR > 2021 AND PUBYEAR < 2025. Further filters included document type as journal articles (LIMIT-TO (DOCTYPE, "ar")), source type as journals (LIMIT-TO (SRCTYPE, "j")), language as English, publication stage as final (LIMIT-TO (PUBSTAGE, "final")), and open access availability (LIMIT-TO (OA, "all")). This refined query yielded 40 relevant articles. In WoS, a broader search strategy was applied using the query: "Data Management" (All Fields) AND "Citizen Science"(All Fields). This query was similarly restricted to open-access, English-language journal articles, ensuring consistency with the SCOPUS search parameters. The WoS search returned 43 articles. Both search strategies were designed to capture high-quality, peer-reviewed, and openly accessible studies that discuss how data generated from CS is organized, stored, validated, shared, and reused across disciplines.

Table 1: Search Strings from Scopus and WoS

Academic Database	Keywords	Results
SCOPUS	(TITLE-ABS-KEY ("Data Management") AND TITLE-ABS-KEY ("Citizen Science")) AND PUBYEAR > 2021 AND PUBYEAR < 2025 AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (OA , "all")) AND (LIMIT-TO (PUBSTAGE , "final")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English"))	40 articles
WoS	" Data Management " (All Fields) and " Citizen Science" (All Fields) and Article (Document Types) and English (Languages) and All Open Access (Open Access)	43 articles

Figure 2 illustrates the article selection process based on records retrieved from SCOPUS and WoS databases. Initially, 40 articles were identified from SCOPUS and 43 from WoS, but only 15 and 11 articles were available for full-text review. After merging and screening, 10 duplicates were found, and 47 were excluded through eligibility assessment using predefined inclusion and exclusion criteria. Several studies were excluded at this stage due to misalignment

with the research scope, inaccessibility of full articles, broken links, and incomplete metadata. Ultimately, 26 studies met all criteria and were included in the final analysis. Accordingly, the final articles were uploaded into ATLAS.ti 24 for further analysis. This flow diagram demonstrates a transparent, rigorous approach to article selection, ensuring the relevance and quality of the sources used for the thematic synthesis.

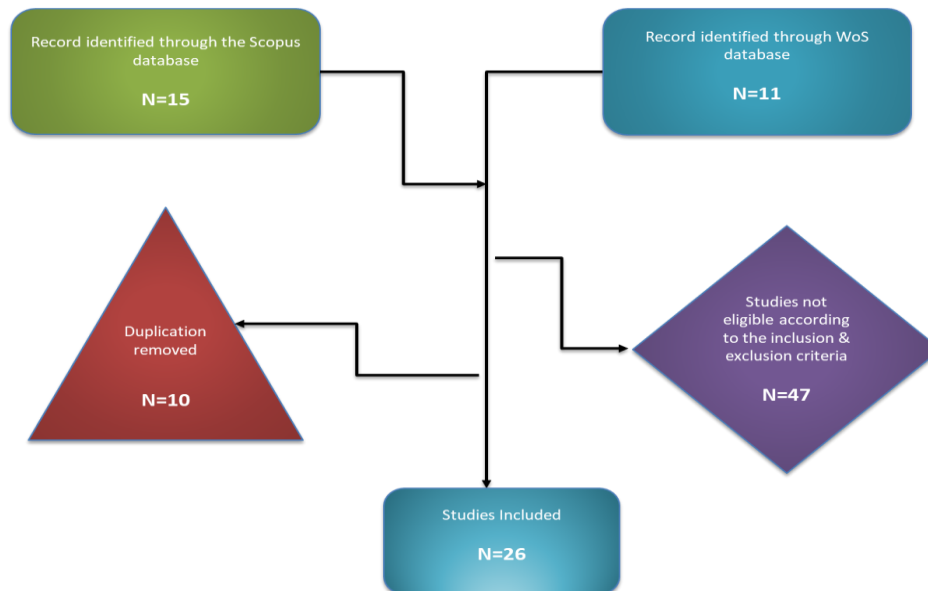


Figure 2: Inclusion and exclusion criteria in the thematic review

To strengthen the methodological rigor of the thematic review, the included studies were further appraised through article quality ranking and bias assessment. Source quality was first identified based on each article's journal status (Scopus and/or WoS) and the journal quality category or quartile available at the time of article extraction. This ranking was used to contextualize the strength and visibility of the evidence base rather than to automatically exclude papers, because thematically relevant studies may still contribute valuable insights across interdisciplinary citizen science contexts. In the next step, the selected articles were uploaded to Mendeley for data processing, which included removing duplicates, standardizing author names, and ensuring metadata accuracy.

Coding was conducted using ATLAS.ti 24 through a transparent multi-stage process. First, each article was assigned a study identifier and classified according to year of publication and country to support trend analysis, as shown in **Figure 3**. The selected articles were then read in full to enable familiarisation with the data. Second, descriptive open coding was applied line-by-line or paragraph-by-paragraph to relevant text segments discussing quality assurance and validation, data governance, data infrastructure, and training. Third, the initial codes were compiled into a working codebook, which included the code name, operational definition, inclusion and exclusion criteria, and brief example excerpts. Next, similar codes were reviewed iteratively and merged into broader categories through constant comparison across the 26 selected articles. Tentative themes were then developed to group articles with similar findings. These themes were repeatedly reapplied to the dataset and refined until closely related themes were combined, and no new substantial themes emerged. Finally, the category-level patterns were examined to generate the final themes, which were checked against the full dataset to

ensure internal coherence, clear distinction between themes, and adequate representation of the reviewed literature. This process resulted in four final themes.

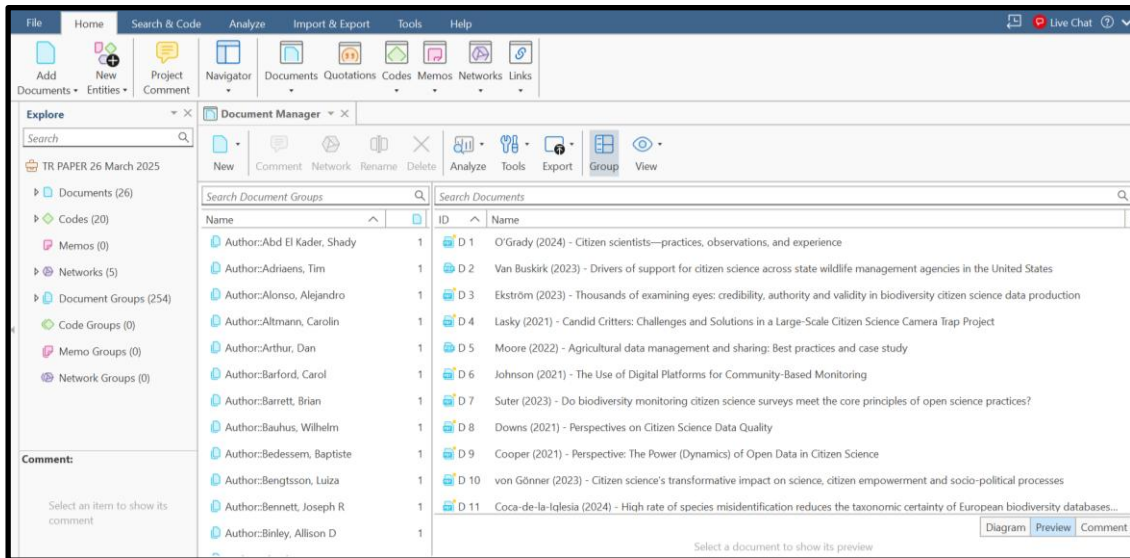


Figure 3: The code group established from ATLAS.ti 24

Results And Discussions

The study's findings are reported in two sections: quantitative and qualitative. The quantitative analysis presents numerical data through graphs, maps, and tables, while the qualitative analysis identifies and develops themes from the selected articles. These analyses offer both numerical insights and thematic depth, facilitating an integrated discussion on the evolution of RDM in CS.

Quantitative Findings

The quantitative part begins with an analysis of the word cloud from 26 final documents. This was followed by analyses of publications by year, geographical dispersion, thematic distribution, and distribution by journal and year.

The word cloud in **Figure 4** visually depicts the most frequently occurring terms in the literature on RDM and CS, offering insight into thematic priorities and conceptual trends. Dominant terms such as “projects” (10,042 times), “data” (4,312), “science” (3,872), “citizen” (2,106), “research” (1,065), and “community” (785) emphasize the participatory and data-centric foundation of CS. The high frequency of “management” (552), “monitoring” (488), “participants” (455), and “quality” (397) shows a clear concern with the integrity and organization of citizen-generated data. Additionally, terms such as “open” (377), “platforms” (264), “sharing” (262), and “digital” (210) reflect alignment with broader movements in open science and technological integration, pointing to the increasing reliance on digital infrastructure for scalable and transparent data practices. Other important and related terms, such as “indigenous” (249), “access” (253), “policy” (208), and “engagement” (172), indicate a growing discourse on inclusivity, equity, and responsible governance in CS. The frequent appearance of “training” (239), “support” (220), and “volunteers” (276) underscores the importance of capacity building, ensuring that citizen scientists are empowered not just as data contributors but as active partners in scientific processes. Domain-specific terms such as “biodiversity” (384), “conservation” (369), and “environmental” (348) further emphasize the

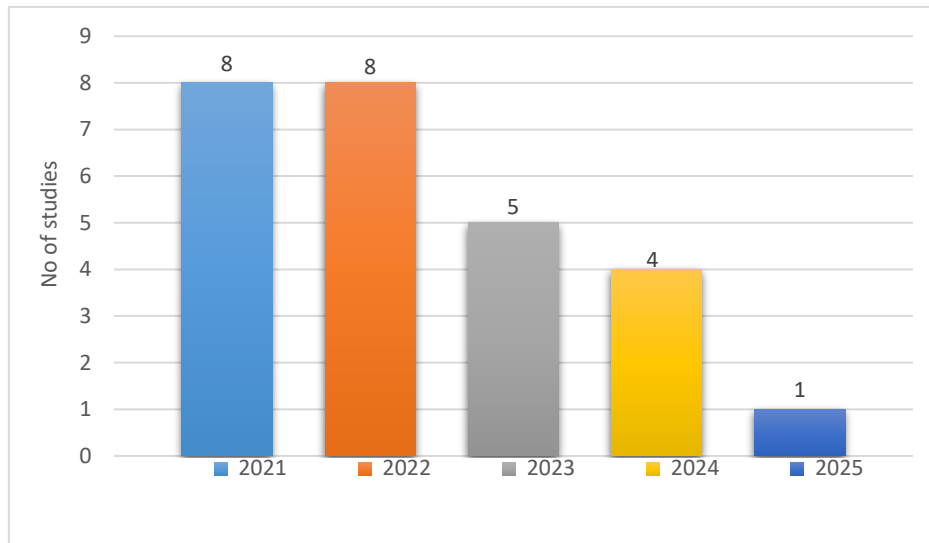


Figure 5: Number of Studies by Year of Publication

The geographical distribution of scholarly contributions in the domain of RDM within CS between 2021 and 2025 demonstrates a clear Western-centric trend, with notable emerging contributions from other global regions. According to **Figure 6**, the United States is the most prolific contributor, with a total of 11 publications, reflecting its continued leadership in CS initiatives for community-driven data infrastructure. Within Europe, the United Kingdom (3), Austria (2), and Spain (2) show active engagement, supported by strong institutional frameworks promoting open science, biodiversity monitoring, and participatory environmental research. Other European countries, such as Belgium, Germany, Ireland, Italy, Norway, Sweden, and Switzerland, each contributed 1 publication, highlighting a widespread but varied regional engagement. Beyond Europe and North America, Canada and Korea each contributed 1 study, consistent with national interests in environmental stewardship and digital innovation. Significantly, Ethiopia appears as a single contributor from the Global South, pointing to growing interest in adopting CS methods to address local development and environmental monitoring challenges. While contributions from Africa and parts of Asia remain limited, this distribution suggests a gradual diffusion of CS research beyond traditionally dominant regions, potentially driven by international policy agendas such as the Sustainable Development Goals (SDGs) and increased access to digital platforms.

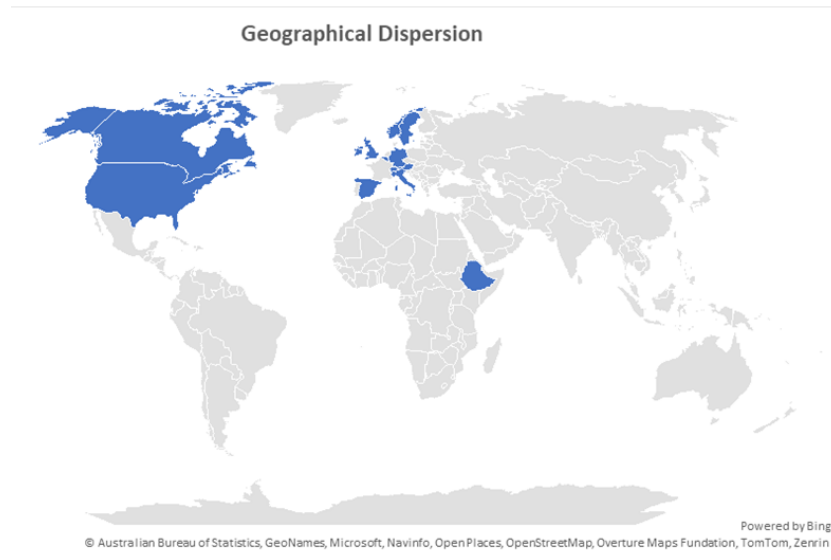


Figure 6: Geographical Dispersion of RDM in CS (2021–2025)

Table 2 shows the thematic analysis of 26 peer-reviewed articles published between 2021 and 2025, revealing four dominant trends in the discourse surrounding RDM in CS. Each article was classified according to its alignment with one or more of the following themes: (T01) data quality assurance and validation, (T02) ethical data governance and open science, (T03) digital infrastructure for scalable data management, and (T04) data literacy and community empowerment. Theme 1, which focuses on the increasing emphasis on data quality assurance and validation, emerged as the most represented, with 8 articles addressing issues related to the credibility, accuracy, and reliability of citizen-generated data. Key contributions in this theme include works by Ekstrom (2023), Downs et al. (2021), Binley and Bennett (2023), and Kebede Mengistie et al. (2024), all of which explore mechanisms for improving data validity, ranging from peer monitoring and statistical controls to quality benchmarking and multi-site data comparisons. This trend reflects a growing scholarly interest in strengthening the scientific rigor of CS datasets. Theme 2, concerning the movement toward ethical data governance and alignment with open science, featured in 6 articles. These works, such as Cooper et al. (2021), Suter et al. (2023), and Thuermer et al. (2023) delve into issues related to data ownership, stewardship, transparency, and the application of FAIR and CARE principles. The literature on this theme emphasizes the need for inclusive governance frameworks that recognize the contributions and rights of citizen scientists while promoting responsible, open data practices.

Theme 3, which examines the integration of digital infrastructure for scalable data management, appeared in 6 articles. Studies by Johnson et al. (2021), Moore et al. (2022), and Soriano-González et al. (2024) highlight the increasing use of digital platforms, data repositories, and sensor-based technologies to manage large-scale, heterogeneous datasets. These digital solutions are enabling better data storage, interoperability, and cross-institutional collaborations, marking a shift toward more technologically mature CS projects. Theme 4, focusing on capacity building and empowerment through data literacy and participation, was observed in 6 articles, including O’Grady and Mangina (2024), Ramirez-Andreotta et al. (2021), and Sorensen and Jordan (2025). These works emphasize the value of training, feedback mechanisms, and participatory design in enhancing the engagement and agency of citizen scientists. The growing discourse in this theme reflects a broader recognition of citizen empowerment as both a goal and a driver of effective data practices in CS. In summary, the

thematic distribution suggests that while data quality assurance remains a foundational concern, there is increasing scholarly attention to ethical governance, digital infrastructure, and the social dimensions of participation. These four themes collectively illustrate a maturing research landscape in which CS is being redefined not only as a data collection method but also as a holistic, ethical, and inclusive model of scientific inquiry and collaboration.

Table 2: Thematic Distribution of Trends in RDM for CS (2021–2025)

Authors	Theme Code			
	T01	T02	T03	T04
O’grady and Mangina (2024)	0	0	0	1
Van Buskirk et al. (2023)	0	0	0	1
Ekstrom (2023)	1	0	0	0
Lasky et al. (2021)	0	0	0	1
Moore et al. (2022)	0	0	1	0
Johnson et al. (2021)	0	0	1	0
Suter et al. (2023)	0	1	0	0
Downs et al. (2021)	1	0	0	0
Cooper et al. (2021)	0	1	0	0
Von Gonner et al. (2023)	0	0	0	1
Coca-De-La-Iglesia et al. (2024)	0	0	1	0
Roman et al. (2021)	0	1	0	0
Ramírez-Andreotta et al. (2021a)	0	0	0	1
Fraisl et al. (2022)	1	0	0	0
Eicken et al. (2021)	0	0	1	0
Binley and Bennett (2023)	1	0	0	0
Koo et al. (2022)	1	0	0	0
De Sherbinin et al. (2021)	0	0	1	0
Sorensen and Jordan (2025)	0	0	0	1
Thuermer et al. (2023)	0	1	0	0
Finch et al. (2022)	1	0	0	0
Kebede Mengistie et al. (2024)	1	0	0	0
Pappers et al. (2022)	1	0	0	0
Soriano-González et al. (2024)	0	0	1	0
Sterner & Elliott (2024)	0	1	0	0
Price-Jones Et Al. (2022)	0	1	0	0

The distribution of publications from 2021 to 2025 shows consistent academic interest in CS and RDM, with variation in annual output and journal representation, as shown in **Table 3**. A total of 26 articles were published across 22 different journals, with 2022 recording the highest number of publications (n = 8), followed by 2021 (n = 8), 2023 (n = 5), 2024 (n = 4), and 2025 (n = 1). The post-pandemic wave may reflect a sharp focus on open science, public engagement, and environmental data. The *Frontiers in Climate* journal led with three articles in 2021, highlighting its climate-focused CS engagement. A similar level of recurring contribution was observed in *BioScience* and *Citizen Science: Theory and Practice*, each with two publications across different years, affirming their relevance in the field. Several journals hosted single publications, yet each contributes valuable disciplinary perspectives. For instance, *Agronomy Journal*, *Journal of Hydrology*, and *Environmental Monitoring and Assessment* reflect contributions from applied environmental and agricultural sciences, underscoring the cross-

sectoral application of CS in data-intensive disciplines. Likewise, publications in Science as Culture, Humanities and Social Sciences Communications, and Socio-Ecological Practice Research point to the socio-cultural dimensions of data use and participation, indicating a multidimensional engagement with citizen-generated data. Notably, 2022 featured the widest journal spread, with eight journals publishing a single article each, indicating expanding interest and publication diversity. In contrast, only one article appeared in (Frontiers in Environmental Science), possibly due to ongoing review cycles. Overall, the temporal and disciplinary patterns indicate that CS and RDM are gaining momentum across both STEM and social sciences, reflecting a multidisciplinary convergence around data stewardship and public participation.

Table 3: Distribution of Publications by Journal and Year (2021–2025)

Publications	2021	2022	2023	2024	2025	Totals
Agronomy Journal	0	1	0	0	0	1
Aslib Journal of Information Management	0	1	0	0	0	1
BioScience	2	0	0	0	0	2
Citizen Science: Theory and Practice	1	0	1	0	0	2
Conservation Biology	0	1	0	0	0	1
Data Technologies and Applications	1	0	0	0	0	1
Diversity	0	1	0	0	0	1
Environmental Monitoring and Assessment	0	0	1	0	0	1
Frontiers in Climate	3	0	0	0	0	3
Frontiers in Conservation Science	0	1	0	0	0	1
Frontiers in Environmental Science	0	0	0	0	1	1
Frontiers in Sustainable Food Systems	1	0	0	0	0	1
Humanities and Social Sciences Communications	0	0	0	1	0	1
Journal of Hydrology	0	0	0	1	0	1
Methods in Ecology and Evolution	0	0	1	0	0	1
Nature Reviews Methods Primers	0	1	0	0	0	1
NeoBiota	0	1	0	1	0	1
Ocean & Coastal Management	0	0	0	0	0	1
Science as Culture	0	0	1	0	0	1
Scientific Reports	0	0	0	1	0	1
Socio-Ecological Practice Research	0	0	1	0	0	1
Transportation Research Interdisciplinary Perspectives	0	1	0	0	0	1
Totals	8	8	5	4	1	26

Beyond publication counts, the article quality ranking showed that the evidence base was drawn from established peer-reviewed journals in citizen science, biodiversity, environmental science, data management, and interdisciplinary research. This indicates that the review is anchored in recognized scholarly outlets and not in informal or non-indexed sources. The ranking exercise, therefore, strengthens confidence in the corpus's credibility while acknowledging disciplinary diversity across the represented journals. The bias assessment further indicated that most studies presented clear aims and relevant empirical or conceptual contributions, placing them generally within low to moderate risk categories. Nevertheless, several recurring sources of bias were observed across the corpus, including uneven reporting of data validation procedures, limited discussion of sampling or participant selection constraints, insufficient reflexive treatment of

platform or technological limitations, and context-specific case designs that limit generalisability. These issues were considered during theme development so that stronger, more transparently reported studies carried greater analytical weight in the final interpretation.

Qualitative Findings

This qualitative analysis includes trends and patterns generated from 20 initial codes after reviewing the 26 articles. After successive rounds of recording and code merging in ATLAS.ti 24, four code clusters were finalized and identified as important themes.

Figure 7 shows the overall themes that address the central RQ: *What are the current trends in RDM related to CS found in the literature from 2021 to 2025?* Through rigorous analysis, four interconnected themes were identified, each representing a critical area of development in CS data practices. Theme 1, which focuses on data quality assurance and validation, reveals a collective concern across multiple studies (e.g., Downs et al., 2021; Ekstrom, 2023) regarding the reliability of citizen-generated data and the need for robust validation mechanisms. This theme closely intersects with Theme 2, ethical data governance and open science alignment, as studies such as Cooper et al. (2021) and Thuermer et al. (2023) argue that transparency, data rights, and ethical stewardship are essential for maintaining public trust in the quality and integrity of data. Furthermore, Theme 3, which emphasizes the integration of digital infrastructure for scalable data management, demonstrates how technological tools are being leveraged to address both quality and governance challenges, providing platforms for secure data storage, interoperability, and real-time validation (Johnson et al., 2021; Moore et al., 2022). This technological advancement is not isolated but is foundational to the success of Theme 4, capacity building and empowerment through data literacy, where participation is enhanced by tools that facilitate not only data submission but also understanding and use of data (Ramirez-Andreotta et al., 2021; Sorensen & Jordan, 2025). These themes are not isolated but represent a multi-dimensional shift in how CS data is conceptualized and managed.

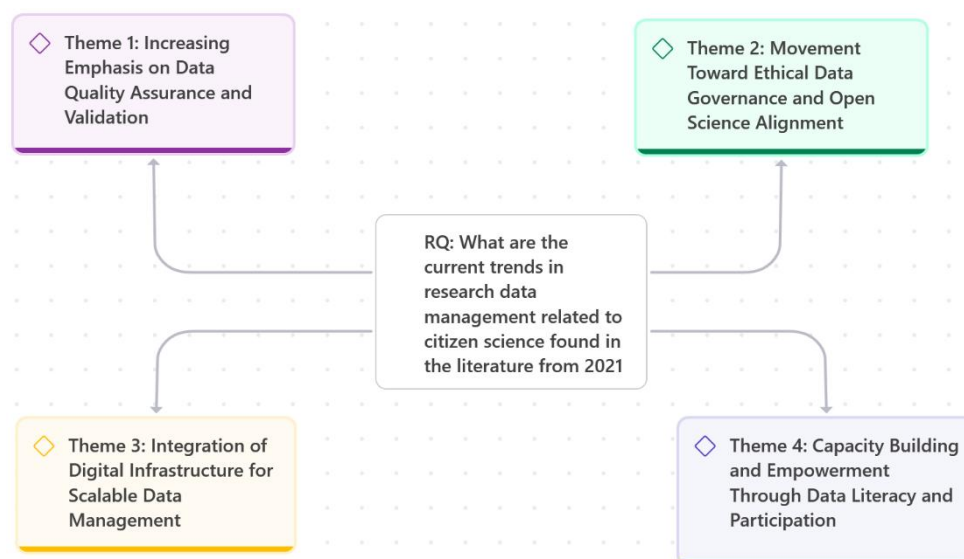


Figure 7: Overall Theme

Theme 1: Increasing Emphasis on Data Quality Assurance and Validation

The focus on data quality assurance and validation in CS, as shown in **Figure 8**, reflects a shared concern for ensuring rigor and credibility in volunteer-contributed data. Ekstrom (2023) and Downs et al. (2021) highlight tensions between expert and amateur knowledge and the challenges of maintaining integrity in decentralized data collection. Additionally, several studies like Koo et al. (2022) and Pappers et al. (2022) propose platform-specific and localized methods for improving accuracy. Similarly, Kebede Mengistie et al. (2024) and Fraisl et al. (2022) stress the importance of standardized protocols for environmental and ecological data. The comparative lens adopted by Binley and Bennett (2023) questions the unequal quality standards between institutional and citizen data, while Finch et al. (2022) advocate for outcome-based validation in conservation efforts. Collectively, these eight studies emphasize that robust validation, clear quality benchmarks, and context-sensitive data assessment are essential for enhancing the credibility and utility of CS.

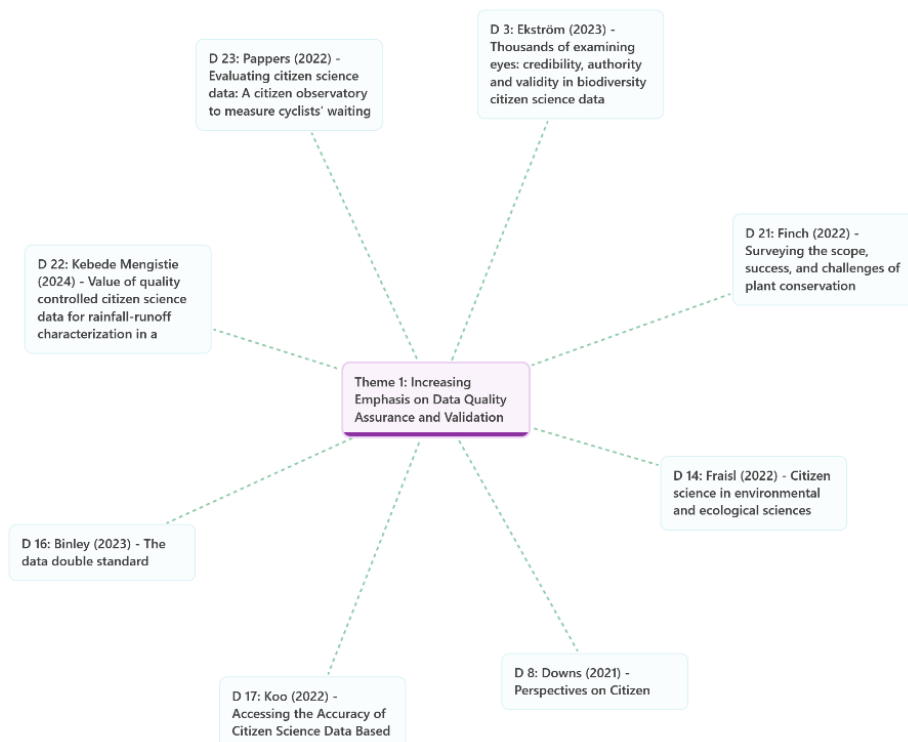


Figure 8: Theme 1

Theme 2: Movement Toward Ethical Data Governance and Open Science Alignment

Figure 9 shows that the movement toward ethical data governance and alignment with open science principles represents a growing focus in CS literature, reflecting a shift from data accessibility to accountability, inclusiveness, and responsible stewardship. Cooper et al. (2021) reveal how power imbalances in open data ecosystems can marginalize citizen contributors despite their essential role in data production. Sterner and Elliott (2024) further argue that ethical data governance is crucial in biodiversity science to promote trust and equitable participation. Additionally, Suter et al. (2023) assess biodiversity monitoring projects and find inconsistencies between open science ideals, such as transparency and participation, and actual data practices, particularly in the handling of citizen-generated environmental data. Similarly, Roman et al. (2021) analyze pollution-focused CS projects and highlight gaps in ethical

engagement and community participation, emphasizing that project designs often overlook local agency in favor of technocratic data collection models. Thuermer et al. (2023) emphasize the need for formalized documentation and data policies in CS, cautioning that without clear protocols, open data risks becoming performative. Similarly, Price-Jones et al. (2022) highlight how ethical frameworks are essential in integrating community knowledge into policy and research, especially in managing biological invasions. Collectively, these studies call for a recalibration of CS toward ethically sound, participatory, and equitable data governance.

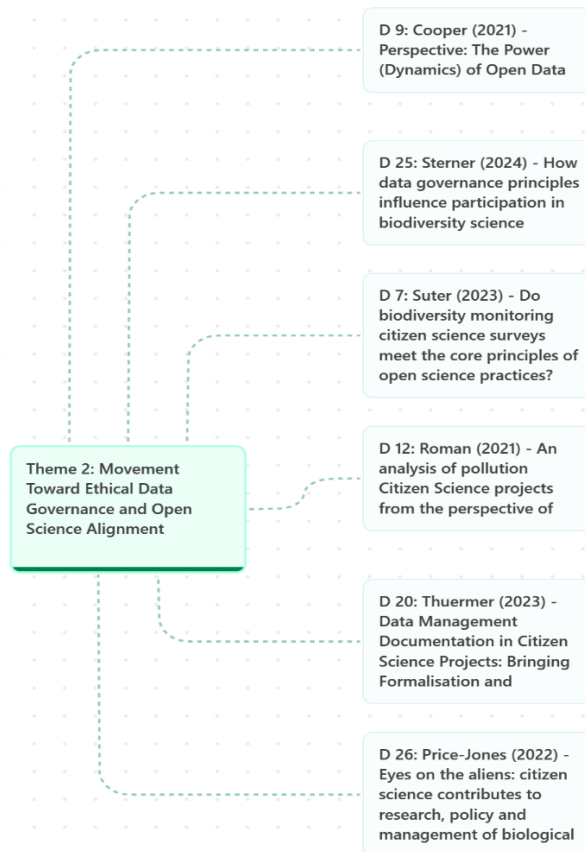


Figure 9: Theme 2

Theme 3: Integration of Digital Infrastructure for Scalable Data Management

The integration of digital infrastructure for scalable data management, as shown in **Figure 10**, has become a key trend in the evolution of CS. Johnson et al. (2021) highlight how community-based platforms enhance data collection and communication, especially in rural areas. Complementing this, Moore et al. (2022) stress the need for interoperable systems that integrate expert and community data in agriculture. de Sherbinin et al. (2021) elevate the discussion by emphasizing the role of robust infrastructure in tackling global sustainability challenges. Moreover, Eicken et al. (2021) reveal the importance of linking top-down and bottom-up environmental observation systems, where digital platforms serve as mediators between institutional science and community knowledge. This approach is exemplified in Soriano-Gonzalez et al. (2024), who document the transformation of a CS initiative into a coastline monitoring system via the Spanish CoastSnap project, showcasing how digital infrastructure facilitates operational continuity and cross-institutional learning. Coca-de-la-Iglesia et al. (2024) underscore the limitations of current digital databases by identifying species

misidentification issues within European biodiversity repositories, reinforcing the need for quality-assurance mechanisms built into digital systems. Thus, these studies show that digital infrastructure is essential for the scalability, reliability, and integration of CS data.

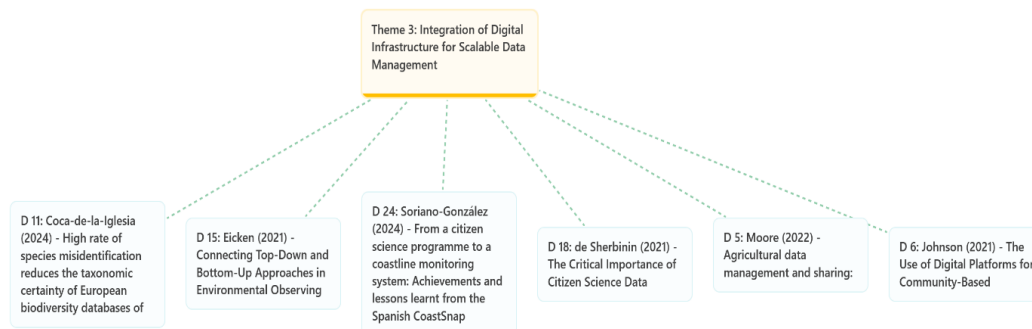


Figure 10: Theme 3

Theme 4: Capacity Building and Empowerment Through Data Literacy and Participation

The theme of capacity building and empowerment through data literacy and participation highlights the growing recognition of citizens as active contributors to science and decision-making, as illustrated in **Figure 11**. O'Grady and Mangina (2024) show how citizen engagement shapes understanding and participation, while Lasky et al. (2021), point to the need for structured training in large-scale CS initiatives. Von Gonner et al. (2023) emphasize the potential of CS to foster scientific literacy and community leadership. Additionally, Sorensen and Jordan (2025) demonstrate how communication strategies influence engagement and ownership. Ramírez-Andreotta et al. (2021) advocate for community-based data use to address health disparities, and Van Buskirk et al. (2023) highlight how institutional support affects CS effectiveness. Collectively, these studies show that empowerment in CS is driven by inclusive design, training, and communication, which makes data literacy both a tool and an outcome for more democratic, responsive science.

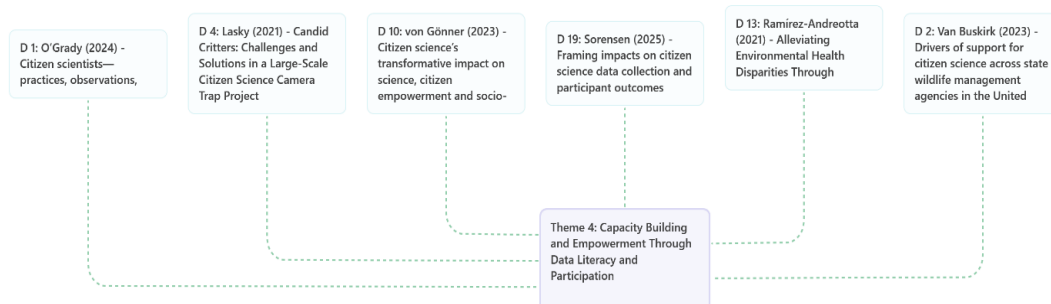


Figure 11: Theme 4

Conclusions

By utilizing a thematic review, this study examined current trends in RDM within CS projects, focusing on literature published between 2021 and 2025. Through a systematic analysis of 26 peer-reviewed articles, four interrelated themes were identified: (1) increasing emphasis on data quality assurance and validation, (2) movement toward ethical data governance and open science alignment, (3) integration of digital infrastructure for scalable data management, and (4) capacity building and empowerment through data literacy and participation. These themes

collectively reflect a field in transition from informal, fragmented practices to more structured, transparent, and inclusive approaches to managing citizen-generated data. The key contribution of this review lies in synthesizing multidisciplinary insights to highlight how RDM is evolving in CS, both as a technical process and a socially embedded practice. It advances the discourse by showing that high-quality, ethically governed, and technologically supported data infrastructures are essential for realizing the full scientific and societal potential of CS. Furthermore, the review underscores that empowering citizens through data literacy and inclusive participation is not only ethically imperative but also critical to the sustainability and credibility of CS initiatives. In summary, future studies might build on these findings by exploring case-specific applications of RDM frameworks across different domains such as public health, climate science, and biodiversity monitoring. Additionally, emerging topics such as AI-assisted validation, indigenous data governance, and participatory policymaking in data ethics demand deeper investigation. By aligning technological advancement with ethical and community-centered values, future studies might help shape CS as a model for responsible, open, and impactful data practices.

References

- Binley, A. D., & Bennett, J. R. (2023). The data double standard. *Methods in Ecology and Evolution*, 14(6), 1389–1397. <https://doi.org/10.1111/2041-210X.14110>
- Clarke, V., & Braun, V. (2013). *Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning*. 1–14. [https://doi.org/http://eprints.uwe.ac.uk/21155/3/Teaching thematic analysis Research Repository version.pdf](https://doi.org/http://eprints.uwe.ac.uk/21155/3/Teaching%20thematic%20analysis%20Research%20Repository%20version.pdf)
- Coca-de-la-Iglesia, M., Gallego-Narbón, A., Alonso, A., & Valcárcel, V. (2024). High rate of species misidentification reduces the taxonomic certainty of European biodiversity databases of ivies (*Hedera L.*). *Scientific Reports*, 14(1), 1–13. <https://doi.org/10.1038/s41598-024-54735-0>
- Cooper, C. B., Rasmussen, L. M., & Jones, E. D. (2021). Perspective: The Power (Dynamics) of Open Data in Citizen Science. *Frontiers in Climate*, 3(July), 1–6. <https://doi.org/10.3389/fclim.2021.637037>
- de Sherbinin, A., Bowser, A., Chuang, T. R., Cooper, C., Danielsen, F., Edmunds, R., Elias, P., Faustman, E., Hultquist, C., Mondardini, R., Popescu, I., Shonowo, A., & Sivakumar, K. (2021). The Critical Importance of Citizen Science Data. *Frontiers in Climate*, 3(March), 1–7. <https://doi.org/10.3389/fclim.2021.650760>
- Downs, R. R., Ramapriyan, H. K., Peng, G., & Wei, Y. (2021). Perspectives on Citizen Science Data Quality. *Frontiers in Climate*, 3(April), 1–7. <https://doi.org/10.3389/fclim.2021.615032>
- Eicken, H., Danielsen, F., Sam, J.-M., Fidel, M., Johnson, N., Poulsen, M. K., Lee, O. A., Spellman, K. V., Iversen, L., Pulsifer, P., & Enghoff, M. (2021). Connecting Top-Down and Bottom-Up Approaches in Environmental Observing. *BioScience*, 71(5), 467–483. <https://doi.org/10.1093/biosci/biab018>
- Ekström, B. (2023). Thousands of examining eyes: credibility, authority and validity in biodiversity citizen science data production. *Aslib Journal of Information Management*, 75(1). <https://doi.org/10.1108/AJIM-10-2021-0292>
- Finch, J., Jasny, M., Kucera, K. F., & Kiefer, G. (2022). Surveying the scope, success, and challenges of plant conservation community science. *Frontiers in Conservation Science*, 3(November), 1–17. <https://doi.org/10.3389/fcosc.2022.933292>
- Fraisl, D., Hager, G., Bedessem, B., Gold, M., Hsing, P.-Y., Danielsen, F., Hitchcock, C. B., Hulbert, J. M., Piera, J., Spiers, H., Thiel, M., & Haklay, M. (2022). Citizen science in environmental and ecological sciences. *Nature Reviews Methods Primers*, 2(1), 64. <https://doi.org/10.1038/s43586-022-00144-4>
- Guerrini, C. J., Crossnohere, N. L., Rasmussen, L., & Bridges, J. F. P. (2021). A best–worst scaling experiment to prioritize concern about ethical issues in citizen science reveals heterogeneity on people-level v. data-level issues. *Scientific Reports*, 11(1), 1–9. <https://doi.org/10.1038/s41598-021-96743-4>
- Johnson, N., Druckenmiller, M. L., Danielsen, F., & Pulsifer, P. L. (2021). The Use of Digital Platforms for Community-Based Monitoring. *BioScience*, 71(5), 452–466. <https://doi.org/10.1093/biosci/biaa162>
- Kebede Mengistie, G., Demissie Wondimagednehu, K., Walker, D. W., & Tamiru Haile, A. (2024). Value of quality-controlled citizen science data for rainfall-runoff characterization in a rapidly urbanizing catchment. *Journal of Hydrology*, 629(January), 130639. <https://doi.org/10.1016/j.jhydrol.2024.130639>
- Koo, K.-S., Oh, J.-M., Park, S.-J., & Im, J.-Y. (2022). Accessing the Accuracy of Citizen Science Data Based on iNaturalist Data. *Diversity*, 14(5), 316. <https://doi.org/10.3390/d14050316>

- Lasky, M., Parsons, A., Schuttler, S., Mash, A., Larson, L., Norton, B., Pease, B., Boone, H., Gatens, L., & Kays, R. (2021). Candid Critters: Challenges and Solutions in a Large-Scale Citizen Science Camera Trap Project. *Citizen Science: Theory and Practice*, 6(1). <https://doi.org/10.5334/cstp.343>
- Loglisci, N., Boni, G., Cauteruccio, A., Faccini, F., Milelli, M., Paliaga, G., & Parodi, A. (2024). The role of citizen science in assessing the spatiotemporal pattern of rainfall events in urban areas: a case study in the city of Genoa, Italy. *Natural Hazards and Earth System Sciences*, 24(7), 2495–2510. <https://doi.org/10.5194/nhess-24-2495-2024>
- Lotfian, M., Ingensand, J., & Brovelli, M. A. (2022). An Approach for Real-Time Validation Of The Location Of Biodiversity Observations Contributed In A Citizen Science Project. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-4/W1-2022(4/W1-2022), 271–278. <https://doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-271-2022>
- Mandeville, C. P., Nilsen, E. B., & Finstad, A. G. (2022). Spatial distribution of biodiversity citizen science in a natural area depends on area accessibility and differs from other recreational area use. *Ecological Solutions and Evidence*, 3(4), 1–14. <https://doi.org/10.1002/2688-8319.12185>
- Mason, E., Gascuel-Oudou, C., Aldrian, U., Sun, H., Miloczki, J., Götzinger, S., Burton, V. J., Rienks, F., Di Lonardo, S., & Sandén, T. (2024). Participatory soil citizen science: An unexploited resource for European soil research. *European Journal of Soil Science*, 75(2), 1–17. <https://doi.org/10.1111/ejss.13470>
- Mohd Zairul. (2021). A thematic review on industrialised building system (IBS) publications from 2015-2019: analysis of patterns and trends for future studies of IBS in Malaysia. *Pertanika Journal of Social Sciences and Humanities*, 29(1), 635–652. <https://doi.org/10.47836/pjssh.29.1.35>
- Moore, E. K., Kriesberg, A., Schroeder, S., Geil, K., Haugen, I., Barford, C., Johns, E. M., Arthur, D., Sheffield, M., Ritchie, S. M., Jackson, C., & Parr, C. (2022). Agricultural data management and sharing: Best practices and case study. *Agronomy Journal*, 114(5), 2624–2634. <https://doi.org/10.1002/agj2.20639>
- O’Grady, M., & Mangina, E. (2024). Citizen scientists—practices, observations, and experience. *Humanities and Social Sciences Communications*, 11(1), 1–9. <https://doi.org/10.1057/s41599-024-02966-x>
- Pappers, J., Keserü, L., & De Wilde, L. (2022). Evaluating citizen science data: A citizen observatory to measure cyclists’ waiting times. *Transportation Research Interdisciplinary Perspectives*, 14(December 2021), 100624. <https://doi.org/10.1016/j.trip.2022.100624>
- Price-Jones, V., Brown, P. M. J., Adriaens, T., Tricarico, E., Farrow, R. A., Cardoso, A. C., Gervasini, E., Groom, Q., Reyserhove, L., Schade, S., Tsinaraki, C., & Marchante, E. (2022). Eyes on the aliens: citizen science contributes to research, policy and management of biological invasions in Europe. *NeoBiota*, 78, 1–24. <https://doi.org/10.3897/neobiota.78.81476>
- Ramírez-Andreotta, M. D., Walls, R., Youens-Clark, K., Blumberg, K., Isaacs, K. E., Kaufmann, D., & Maier, R. M. (2021a). Alleviating Environmental Health Disparities Through Community Science and Data Integration. *Frontiers in Sustainable Food Systems*, 5. <https://doi.org/10.3389/fsufs.2021.620470>
- Roman, D., Reeves, N., Gonzalez, E., Celino, I., Abd El Kader, S., Turk, P., Soylu, A., Corcho, O., Cedazo, R., Re Calegari, G., Scandolari, D., & Simperl, E. (2021). An analysis of pollution Citizen Science projects from the perspective of data science and open science. *Data Technologies and Applications*, 55(5), 622–642. <https://doi.org/10.1108/DTA-10-2020-0253>

- Rowbotham, S., Walker, P., Marks, L., Irving, M., Smith, B. J., & Laird, Y. (2023). Building capacity for citizen science in health promotion: a collaborative knowledge mobilisation approach. *Research Involvement and Engagement*, 9(1), 36. <https://doi.org/10.1186/s40900-023-00451-4>
- Shwe, K. M. (2020). Study on the Data Management of Citizen Science: From the Data Life Cycle Perspective. *Data and Information Management*, 4(4), 279–296. <https://doi.org/10.2478/dim-2020-0019>
- Sorensen, A. E., & Jordan, R. C. (2025). Framing impacts on citizen science data collection and participant outcomes. *Frontiers in Environmental Science*, 12. <https://doi.org/10.3389/fenvs.2024.1496203>
- Soriano-González, J., Sánchez-García, E., & González-Villanueva, R. (2024). From a citizen science programme to a coastline monitoring system: Achievements and lessons learnt from the Spanish CoastSnap network. *Ocean and Coastal Management*, 256(July). <https://doi.org/10.1016/j.ocecoaman.2024.107280>
- Sterner, B., & Elliott, S. (2024a). How data governance principles influence participation in biodiversity science. *Science as Culture*, 33(3), 366–391. <https://doi.org/10.1080/09505431.2023.2214155>
- Steven, R., Barnes, M., Garnett, S. T., Garrard, G., O’connor, J., Oliver, J. L., Robinson, C., Tulloch, A., & Fuller, R. A. (2019). Aligning citizen science with best practice: threatened species conservation in Australia. *Conservation Science and Practice*, 1(10), 1–12. <https://doi.org/10.1111/csp2.100>
- Suter, S., Barrett, B., & Welden, N. (2023a). Do biodiversity monitoring citizen science surveys meet the core principles of open science practices? *Environmental Monitoring and Assessment*, 195(2), 295. <https://doi.org/10.1007/s10661-022-10887-y>
- Thuermer, G., Guardia, E. G., Reeves, N., Corcho, O., & Simperl, E. (2023). Data Management Documentation in Citizen Science Projects: Bringing Formalisation and Transparency Together. *Citizen Science: Theory and Practice*, 8(1), 25. <https://doi.org/10.5334/cstp.538>
- Van Buskirk, A. N., Boley, B. B., Killmaster, C. H., Johannsen, K. L., & D’Angelo, G. J. (2023). Drivers of support for citizen science across state wildlife management agencies in the United States. *Conservation Biology*, 37(3), 1–15. <https://doi.org/10.1111/cobi.14028>
- von Gonner, J., Herrmann, T. M., Bruckermann, T., Eichinger, M., Hecker, S., Klan, F., Lorke, J., Richter, A., Sturm, U., Voigt-Heucke, S., Brink, W., Liedtke, C., Premke-Kraus, M., Altmann, C., Bauhus, W., Bengtsson, L., Büermann, A., Dietrich, P., Dörler, D., ... Bonn, A. (2023). Citizen science’s transformative impact on science, citizen empowerment and socio-political processes. *Socio-Ecological Practice Research*, 5(1), 11–33. <https://doi.org/10.1007/s42532-022-00136-4>
- Zairul, M. (2020). A thematic review on student-centred learning in the studio education. *Journal of Critical Reviews*, 7(2), 504–511. <https://doi.org/10.31838/jcr.07.02.95>
- Zairul, M. (2022). Opening the pandora’s box of issues in the industrialised building system (ibs) in malaysia: A thematic review. *Journal of Applied Science and Engineering*, 25(2), 297–310. [https://doi.org/10.6180/jase.202204_25\(2\).0006](https://doi.org/10.6180/jase.202204_25(2).0006)
- Zairul, M. (2023). *Thematic Review template* (Patent Patent No. 003009261–M).
- Zairul, M., Azli, M., & Azlan, A. (2023). Defying tradition or maintaining the status quo? Moving towards a new hybrid architecture studio education to support blended learning post-COVID-19. *Archnet-IJAR: International Journal of Architectural Research*, 17(3), 554–573. <https://doi.org/10.1108/ARCH-11-2022-0251>