ENSAIO • Pesquisa em Educação em Ciências | 2024 | 26:e51515
DOI | http://dx.doi.org/10.1590/1983-21172022240172T

EDITORIAL

OPEN DATA IN SCIENCE EDUCATION RESEARCH: PERSPECTIVES, CHALLENGES, AND POSSIBILITIES

Nathália Helena Azevedo[I]
https://orcid.org/0000-0003-2075-9121

Paula Cristina Cardoso Mendonça[II]
https://orcid.org/0000-0003-1762-4840

ABSTRACT:

In this editorial, we present the open science perspectives adopted by the *Ensaio* journal, emphasizing open data practices and shared data. The availability of research data can improve transparency, collaboration, reproducibility, replication, reuse, and enrichment of scientific production in Science Education. We offer new perspectives on the topic based on our experiences throughout 2023 in implementing a Data Editorship, whose work we detail in a flowchart. Furthermore, we address some of the possible implications of ethical data sharing in the field and report on some of the outcomes of the interactions between our editorial team, authors, and the academic community involved in discussions on open science by summarizing the paths we have taken in this sphere.

Keywords:
Open science;
Open data;
Science education.

DADOS ABERTOS NA PESQUISA EM EDUCAÇÃO EM CIÊNCIAS: PERSPECTIVAS, DESAFIOS E POSSIBILIDADES

RESUMO:

Neste editorial apresentamos as perspectivas de ciência aberta adotada pelo periódico dando destaque para as práticas de dados abertos e dados compartilhados. A disponibilização de dados de pesquisa pode aprimorar a transparência, a colaboração, a reprodutibilidade, a replicação, o reuso e o enriquecimento da produção científica no campo da Educação em Ciências. Por isso, neste editorial, trazemos novas perspectivas à temática com base em nossos aprendizados ao longo de 2023 na implementação de uma editoria de dados, a qual detalhamos o trabalho em um fluxograma. Além disso, abordamos algumas das possíveis implicações do compartilhamento ético de dados na área e relatamos alguns dos frutos das interações entre nossa equipe editorial, pessoas autoras e a comunidade acadêmica envolvida nas discussões sobre ciência aberta e sintetizamos os caminhos que temos trilhado nessa seara.

Palavras-chave:
Ciência aberta;
Dados abertos;
Educação em Ciências.

DATOS ABIERTOS EN LA INVESTIGACIÓN EN EDUCACIÓN EN CIENCIAS: PERSPECTIVAS, DESAFÍOS Y POSIBILIDADES

RESUMEN:

En este editorial presentamos las perspectivas de la ciencia abierta adoptadas por *Ensaio* con énfasis en prácticas de datos abiertos y datos compartidos. La disponibilidad de datos de investigaciones puede mejorar la transparencia, la colaboración, la la reproducibilidad, la

Palabras clave:
Ciencia abierta;
Datos abiertos;
Educación en Ciencias.

I University of Groningen: Groningen, Institute for Science Education and Communication, Groningen, NL.
II Federal University of Ouro Preto, Ouro Preto, Brazil.

replicación, la reutilización y el enriquecimiento de la producción científica en el campo de la Educación en Ciencias. Por lo tanto, presentamos nuevas perspectivas acerca del tema a partir del conocimiento aprendido a lo largo de 2023 en la implantación de una sección de datos, cuyo trabajo detallamos en el flujograma. Además, abordamos algunas de las posibles implicaciones del intercambio ético de datos en el área y relatamos algunos de los resultados de las interacciones que tuvimos con nuestro equipo de edición, autores y comunidad académica involucrada en las discusiones sobre los datos abiertos y también resumimos los caminos recorridos en esta esfera.

The global trend directed to the management and sharing of open data in research, allowing the access and reuse of original data sets (Albagli et al., 2014; Open Knowledge International, 2013) is already present in the guidelines of the agencies that encourage the research nationally (e.g., FAPESP) and internationally (e.g., European Commission). This movement towards open data, which was initiated at the end of the 1990s and is stimulated by international organizations such as the Organization for Economic Cooperation and Development, OECD (2007), and UNESCO (2021), is part of a broader perspective of open science (Ruusalepp, 2008). This vision is promoted by a series of possible benefits, including the promotion of new studies, transparency, reproducibility, replication,[1] economy of costs, diversification of analysis, and reinforcement of open scientific research (Chauvette et al., 2019; Mozersky et al., 2020a; Obels et al., 2020). Additionally, open data can contribute to increased transparency and accountability, resulting in more efficient and economical use of data, mainly in publicly funded studies (Janz, 2016; Mozersky et al., 2020a; Prosser et al., 2022). Open data can also optimize the data value and ease the burden on the participants since it encourages the reuse of data previously collected (Mozersky et al., 2020a).

As an increasing number of researchers, universities, and funding agencies want to promote data opening, it is crucial to acknowledge the ethical, ontological, and epistemological dimensions that can distinguish more qualitative from quantitative data in this debate. This is so because not all the data is equally adequate for openness (Chauvette et al., 2019). Qualitative research, highly present in the studies of Science Education, includes ontological, epistemological, and methodological challenges that are distinct due to the contextual nature and variation in the use of the data when removed from their contexts. It is, therefore, essential to promote a transparent dialog about the complexities inherent in qualitative research, such as preserving data quality and integrity. This is crucial to ensure that the transition to open data does not harm research participants or compromise the validity and proper interpretation of this data in generating knowledge (Prosser et al., 2022).

Data is fundamental information either collected or generated systematically to approach issues or solve research problems. This data is the raw material for developing reasoning, discussions, and calculations inherent to the research (Jaunsen, 2018). Data can take different forms, depending on the processing stage. Raw data, for example, represents information obtained directly from a source, such as completed questionnaires, interview recordings, ethnographic observations, artifacts, or textual and visual documents. On the other hand, processed data refers to organized and coded information not yet submitted to an in-depth analysis, such as typed and coded questionnaires or transcribed and categorized data. As treated data, they represent a more refined and analyzed form of data, including, for example, content analysis to identify themes and subthemes, quantitative synthesis, statistical analyses, templates, algorithms, and visualizations. This distinction between raw, processed, and treated data is crucial, for it underlines the transformation along the research cycle, from the initial collection to the production of results and substantial conclusions.

When exploring the sharing of data and supplementary material in the research, it is fundamental to understand that the availability of this data can occur in different processing phases, showing distinct access levels (Foster, n.d.). The notion of open data refers to information available to all, with no restrictions, which can be licensed for different purposes. On the other hand, shared data allows broad access, but it is subject to conditions such as restriction to commercial reuse or the need to grant a usage license. In certain circumstances, access to these data can be granted only to specific groups, like colleagues from other academic institutions. Closed data, instead, implies confidentiality restrictions, mainly when dealing with sensitive data. However, even in these cases, it is essential to disclose research metadata, and the access can be controlled through specific repositories, thus ensuring that only authorized people have access. Hence, when we talk about data and research material sharing, we also consider the distinct levels of openness to ensure transparency and security according to the requirements of each situation.

We finished the editorial of 2023 — *Experiences of the journal Ensaio Research in Science Education with open peer review* (Mendonça et al., 2023) — with the section *From OPR (open peer review) to open data: our perspectives*, in which we describe the initial actions taken by our editorial team to implement a policy for open sharing of data and supplementary materials related to the research. The availability of research data is a reality, and we consider it can improve transparency, collaboration, replicability, and scientific production in science education. Therefore, in this editorial, we bring new perspectives to the theme based on our learnings throughout the year 2023 with the implementation of a Data Editorship, reporting some results of the interactions between our team, the authors, and the academic community involved in the discussions on open science, and then summarizing the many ways we have traveled in this area.

WHAT ARE THE POSSIBLE POSITIVE IMPLICATIONS OF SHARING OPEN DATA?

Considering the recent and prolific literature concerning research open data, there is no reason not to sustain, at first, that the practice of open sharing of data and supplementary materials in the research of science education cannot be supported by a series of fundamental reasons that strengthen the credibility and contribute to the construction of new knowledge. In general, researchers from different areas of knowledge (including humanities, but with scarce information about the education area) have adopted a positive perspective towards the open sharing of research data, recognizing, for the most part, that the practice brings benefits to science (CWTS & Elsevier, 2017).

The availability of open data allows researchers and the public in general to examine, evaluate, and check the results of the studies, thus reducing the possibilities of scientific misconduct (Corti et al., 2016) and honest mistakes. Under this context, the debates around open data sharing persuade researchers and institutions to adopt ethical standards, mainly to avoid manipulating results, which contributes to the integrity of the research and the debates related to it. In this way, credibility and transparency are frequently mentioned as the main reasons for openly sharing research data since this level of transparency is essential to establish trust in the research (Uhlir & Schröder, 2007).

In addition, open data allows for replicability, reproducibility, and the search to validate research. Facilitating the replication/reproduction of studies can confirm the validity of the results, thus increasing the trust and strength of scientific studies (Uhlir & Schröder, 2007; Wilbanks et al., 2006). Despite this, several areas of knowledge, including the humanities and the biological sciences, have reported challenges in the face of the so-called reproducibility crisis (Freese & Peterson, 2017; Munafò et al., 2017). About this subject, when preparing a consensus report about reproducibility and replicability in science, the National Academy of Sciences, Engineering, and Medicine of the United States highlighted that

> The scientific enterprise depends on the ability of the scientific community to scrutinize scientific claims and to gain confidence over time in results and inferences that have stood up to repeated testing. Important throughout this process is the sharing of data and methods and the estimation, characterization, and reporting of uncertainty. Reporting uncertainty in scientific results is a central tenet of the scientific process, and it is incumbent on scientists to convey the appropriate degree of uncertainty to accompany original claims (NAS, 2019, p.06).

Among the significant benefits of sharing open data is promoting collaboration and innovation. When data are available, experts can take advantage of several perspectives and alternative approaches for analysis, which often result in new knowledge and innovation (Bispo & Kuula-Luumi, 2017). This practice also contributes to more efficiency of resources, as the research teams can avoid duplication of activities in the collection of similar data and, instead of that, build upon already existing works (Chavette et al., 2019). This allows more efficient use of the available financial and human results and maximizes the impact of the research (Mozersky et al., 2020a).

Open data sharing also enables the implementation of meta-analytic studies, in which several data sets can be combined for a comprehensive analysis (Bispo & Kuula-Luumi, 2017). From the methodological point of view, open data can stimulate new kinds of research and support studies on collection methods and data measuring, which incorporates, for example, automated tools for organization and processing (Uhlir & Schröder, 2007). They also make creating new data sets, information, and knowledge possible by combining data from multiple sources (Uhlir & Schröder, 2007; Wilbanks et al., 2006).

Additionally, open data sharing has a crucial role in archiving and preserving research assets. Here, the Data Management Plan (DMP) plays an essential function in the direction of studies, determining the strategic route for collecting, organizing, storing, and sharing data. The DMP specifies collection methods and those persons who will have access to the data; it clearly outlines the security and ethical protocols that will be adopted. By approaching fundamental issues, such as which technologies will be employed to access the data and the structures for storage and conservation, the DMP also offers detailed guidelines to ensure the integrity and security of data over time. Even though the elaboration of the DMP is not yet a common practice in many areas of research (Gorgolewski et al., 2013), its relevance follows the discussions about open data, and its elaboration has escalated among national and international funding agencies.[2]

As an example, a study highlighted a significant drop in the accessibility of data from 516 ecological studies published between 1991 and 2011 (Vines et al., 2014). The investigation indicated that the probability of access to these data diminished by 17% per year, with only 20% of the studies in the decade of 1990 keeping their data available. Finding authors and receiving answers was quite challenging, with success only in 37% of the cases, and the probability of finding functional e-mail addresses decreased by 7% per year (Vines et al., 2014). More recent findings in other contexts, such as political science, have also reported difficulties in accessing data and complementary materials from published studies and replicating these studies (Avelino et al., 2021).

Open data sharing not only serves as a safe locker for scientific data, but it also offers an opportunity to apply new analysis methods and to explore topics that may not have been foreseen previously (Uhlir & Schröder, 2007), which stimulates new studies and new kinds of research, thus enriching not only the knowledge base but also opening ways for unexpected forms of understanding phenomena and study objects. Besides, the availability of open data aligns with the philosophy of open education, which aims at making knowledge accessible. This is particularly important for forming new research teams, which start to have access to real data to learn and practice approaches, strategies, and methodological tools, thus enhancing the learning activity. Corroborating this perspective, a study about data repositories in the United Kingdom and Finland highlighted that students and employees from educational institutions are the most significant users of open data repositories of a qualitative nature (Bispo & Kuula-Luumi, 2017).

From an optimistic perspective and in theoretical terms, it is also considered that research open data can contribute to a transfer of information, promoting the development of capacities in countries with fewer research resources and encouraging interdisciplinary, intersectoral, interinstitutional, and international research (Uhlir & Schröder, 2007; Wilbanks et al., 2006). Finally, by principle, it is assumed that data sharing as part of open science amplifies the return on public investment in research for society.

SPECIFICITIES OF QUALITATIVE RESEARCH AND CONSIDERATIONS ABOUT (NOT) SHARING DATA

Despite the points we summarized in the previous section, it is fundamental to explicitly recognize the idiosyncrasies of the (more traditional) qualitative research in Science Education as well as the considerations that can restrict the feasibility when sharing these data. The literature on this theme is still scarce in Education research,[3] which instigates us to propose an interdisciplinary dialogue with researchers from other areas and contexts who have been studying the subject longer. The policies and resources related to open data sharing exhibit considerable variations in global terms; however, most of them seem to have been developed with the purpose of satisfying data of a quantitative nature (Mozersky et al., 2020a).

Notwithstanding, an intrinsic part of the discussions about data opening is the principle that they have to be accessible whenever possible but must remain restricted when necessary, preserving, however, the opening of metadata, which contextualizes the nature of the data. Therefore, the remarkable growth of the open science movement suggests that qualitative studies can also be subject to increasing requirements concerning data sharing (Bishop & Kuula-Luumi, 2017; Crosas et al., 2018). It should be noted, also, that this does not imply an indiscriminate and unethical sharing of data, for there are distinct levels of data opening under the open science perspective, as we mentioned previously, and data can be made available during different phases of the research process (Branney et al., 2019). Following, we briefly present some aspects that permeate this debate, intending to expand the dialogue, for this is a discussion in progress that is still incipient in science education research.

In the European and Australian context, we observe consolidated policies to stimulate qualitative data sharing, while in the United States and Brazil, the practice is relatively recent. Among the reasons that explain the challenges to the sharing of qualitative data include the contextual and sensitive nature of these data, the lack of institutional support and recognition, adequate funding, and the absence of infrastructure and technologies required for data management (Tenopir et al., 2011; CWTS & Elsevier, 2017; Mozersky et al., 2020a). Challenges related to time and continuous updating also represent substantial obstacles. Thus, we understand that any data-sharing paradigm should consider both the difficulties researchers face in generating data and the added value for end users (Jackson & Pachter, 2023).

The dimensions of support and recognition for researchers who adopt data sharing as a practice have also been on the agenda of discussions on open science, given that contributions and workloads can be recognized unequally between different areas of knowledge. In this context, assigning a Digital Object Identifier (DOI) to datasets deposited in Dataverse can be a significant stimulus for the scientific community since it provides unique and permanent identification, facilitates accurate citations, tracks impact over time and promotes efficient data reuse. Despite this, in the Brazilian context, there still need to be specific fields designated for this purpose on the national Lattes Platform, although the inclusion of these datasets is feasible in the technical production field.

From an ethical viewpoint, it is crucial to highlight that qualitative data often address sensitive and stigmatized issues that can rarely be explored comprehensively only through quantitative approaches. This raises considerable concerns about informed consent, the data property, confidentiality, and the required preserva-

tion of anonymity (DuBois et al., 2018; Prosser et al., 2022). The possibility of reidentification of participants, primarily due to the intrinsic detailing of the qualitative data, can bring about adverse implications, which could transform the sharing of such data into a delicate and complex ethical matter (Mozersky et al., 2020b).

Aiming to examine the issue, Mozersky et al. (2020b) conducted a study in the United States which included interviews with 30 individuals who had participated in sensitive qualitative studies to explore their perspectives and concerns regarding qualitative data sharing. The participants were favorable to data sharing, provided that de-identification procedures were in practice and that experts conducted the sharing; however, they raised concerns about confidentiality and the data's inadequate use. Most of them expressed the expectation of having their data shared, thus demonstrating a possible greater sharing obligation than previously considered (Mozersky et al., 2020b). It is important to note that such studies are still rare, especially concerning vulnerable populations, including children, who frequently appear as participants in studies in education. So, as the discussion about the ethics of qualitative data improves, promoting more ethical practices in sharing and using these data is fundamental to balance the need for transparency and the protection of the participants, mainly in contexts of sensitive research.

The contextual nature of qualitative data shows a significant epistemological challenge. Researchers who originally collected these data have a deep understanding of the context and the underlying interpretation. Therefore, the secondary analysis of qualitative data made by researchers who do not have the contextual expertise and the experience of the specialists who raised the original data can be an overly complex task. Notwithstanding, not all qualitative data is unfeasible to share. Resources such as detailed metadata documents, code books, interview guides, and anonymized transcriptions can be available to verify conclusions from specific qualitative studies. Additionally, secondary analyzes of these data can result in new perceptions, improving the field of qualitative research (DuBois et al., 2018), and they have also become a growing practice, including the development of specific repositories for the deposit of qualitative data,[4] as discussed in the following section.

ARE SHARED DATA REUSED?

There are many ways to share research data, and Dataverse is one of the appropriate platforms for this purpose. Widely used in academic and scientific institutions, Dataverse plays a crucial role in promoting a culture of open data, thus promoting the sharing, publication, and management of scientific data sets. This open-source platform allows institutions to adopt, customize, and contribute to their development, offering a centralized infrastructure designed to store, organize, and provide data. Moreover, it provides detailed information, such as metadata and documentation, with accessibility to the academic community and the general public (King, 2007). Furthermore, the use of Dataverse by entities such as the Scientific Electronic Library Online (SciELO) highlights its importance to the ecosystem of scientific research and the diffusion of knowledge in the context of the Global South.

Data presented by the Dataverse[5] project in December of 2023 indicated the existence of more than 15.9 thousand Dataverses, encompassing 403 thousand data sets (datasets) and almost 5 million digital files. Two sets of information derived from these data stand out and raise reflections for a dialogue. First, the total number of files downloaded can be an initial indicator of the interest generated by these data. In December 2023, about 74.8 million downloads were made, considering the files stored in the monitored Dataverses. Secondly, it is pertinent to highlight the data sharing per knowledge areas. Among the collections of datasets, the areas of Social Sciences and Arts and Humanities, traditionally associated with qualitative research, were responsible for approximately 62% of the datasets shared in the Dataverses monitored by the project.[6]

We need more detailed information about using these data repositories for research purposes, especially in Education. Works evaluating citations of openly shared datasets could indicate the use (and quality) of open data. In this way, as we did in the preceding sections, we bring reflections and studies from other knowledge areas to amplify the discussion about open data sharing and use within the scope of the Science Education research.

As previously emphasized, the implications of using open data in qualitative research involve epistemological, methodological, ethical (Alexander et al., 2019; Childs et al., 2014), and legal issues (see Prosser et al., 2022). More precisely, challenges arise concerning the secondary analysis, which involves reusing qualitative data previously collected in prior studies. As we summarized, this practice can be motivated by searching for transparency, validating results, and creating new knowledge based on preexisting data.

One of the main challenges associated with using open data in qualitative studies is the reuse outside the original context. The qualitative data encapsulate individual experiences through words, images, or behaviors and are shaped through observations, interviews, documents, and artifacts (Chauvette et al., 2019; Weller, 2023). The singularity of the qualitative data emerges from historical, cultural, and social contexts, which gives them an inimitable nature (Creswell & Poth, 2018; Weller, 2023). The relationship between researchers and participants leads to the joint construction of knowledge, thus highlighting the importance of the context of the data. Additionally, qualitative data are intrinsically subjective, influenced by the experiences and perspectives of both the researchers and the participants. In this way, proper contextualization is necessary to reuse qualitative data to ensure its interpretation and meaning.

The nature of the specific methods used in qualitative studies also represents challenges to reusing open data. Some research delimitations are not appropriate for secondary analysis due to the strong influence of experts and the contextual nature of the participants' experiences. In participatory research, for example, data are not limited to simple transcriptions, which makes reuse difficult without the involvement of the people participating (Chauvette et al., 2019).

The ethical and legal complexities of open data sharing in qualitative research shall not be underestimated. Informed consent of participants and privacy and confidentiality issues appear as crucial elements (Jesus, 2019). Data sharing may conflict with the original conditions governing the consent obtained. Besides, the inadvertent identification of participants due to specific details in their experiences is a reason for concern. In small communities or specific contexts, the identity of participants can be reconstituted, compromising privacy (Chauvette et al., 2019). Protecting an individual's identity and anonymity can be challenging and result in the exclusion of essential information and the preservation of useless data.

Given these substantial reflections, a question arises about the pertinence of sharing qualitative data for reuse in contexts different from those conceived initially. Despite the apparent limitations, repositories aimed at making qualitative data available are increasing and are already a reality in countries such as Germany, Australia, Slovenia, Finland, Ireland, Norway, Poland, and the United Kingdom (Bishop & Kuula-Luumi, 2017).

By going deeper into the topic, Bishop and Kuula-Luumi (2017) conducted a study focusing on two repositories in the United Kingdom and Finland, and they found an increase in the use of qualitative data for secondary analysis. The study showed that most data collections were used, thus refuting the premise that data were only archived in repositories and never employed. Besides, the research indicated that (i) primary data can stand out as valuable resources to approach complex issues when the direct answers are hard to obtain due to the sensitivity of the topics, and (ii) a variety of approaches for secondary qualitative analysis came up and expanded over time.

In addition to the study by Bishop and Kuula-Luumi (2017), there is evidence of the increase in the reuse of data in secondary qualitative analysis in entire disciplines, as it has been widely debated in the Psychology area in recent years (DuBois et al., 2018). For some authors, sharing and using qualitative data is already a reality, even if researchers remain unaware or resistant to such discussions (Prosses et al., 2022;

DuBois et al., 2018; Strasser, 2013). Therefore, we understand that the continuity of these dialogues and the promotion of better practices for sharing and using qualitative data are essential to understand further the implications and benefits of this process in many areas of academic research, including Science Education.

CONSIDERATIONS ABOUT DATA SHARING IN SCIENCE EDUCATION RESEARCH BASED ON THE FAIR AND CARE PRINCIPLES

The sharing and reuse of data in scientific research depend on mainly two essential factors: adequate data archiving and the capacity to find them afterward. Under this context, the metadata and readme files work as keys to unblock the data potential. They function as guides to describe the content, the structure, and the context of the datasets so other researchers and interested parties can comprehend them. Where these metadata and readme[7] files are well-designed, they enable the discovery and qualified use of the data and ensure clear, structured, and more readily accessible information (Elger et al., 2016; SciELO, 2021, 2023).

The FAIR principles come into play at this point: they establish guidelines that transcend the traditional readme files. They increase the assurance that the data will be easily found (Findable), accessible to interested parties (Accessible), interoperable between different systems and tools (Interoperable), and reusable in several studies (Reusable) (Wilkinson et al., 2016). These principles have been widely adopted to improve the efficiency of scientific data when stored in repositories such as the Dataverses.

The FAIR principles represent a qualitative advance in data sharing since they enable data to be readable and facilitate the collaboration and solution of complex challenges in many disciplines. By adopting the FAIR principles, researchers and institutions promote transparency and collaboration, which aligns with established data management practices in the digital age (Wilkinson et al., 2016). However, while essential, the FAIR principles do not entirely address the complexities inherent to the research data in sensitive areas, such as Education. Ethical and cultural issues play a fundamental role in collecting and using these data. Moreover, the principles of caring for and respecting these data must be equally considered.

In this context, initiatives such as "The Good Data Manifesto" (Trenham & Steer, 2019) and the CARE principles (Collective Benefit, Authority to Control, Responsibility, Ethics) (GIDA, 2018) have been gaining distinction. They try to ensure the responsible use of data, considering the social, cultural, and ethical impact (Borsetti et al., 2021). The CARE principles mainly focus on communities' authority over their data, promoting ethics and responsibility in treating and using this information (GIDA, 2018).

This combination of FAIR and CARE principles has formed the movement for data sharing driven by the slogan #BeCARE&FAIR, which aims to promote an inclusive and ethical approach to data sharing, showing respect for the many requirements and contexts of the involved populations, ensuring that the use of the data is made respectfully and beneficially, taking into consideration their languages, cultures, and visions of the world (Borsetti et al., 2021). With this, we can certify that the research data sharing in sensitive areas like education meets the technical criteria and that essential ethical and cultural principles are observed. In this context, the readme files are a starting point, the CARE principles represent the required ethical and responsible pillars to deal with sensitive information, and the FAIR principles are the pragmatic path to enable the sharing and reusing of data in the research.

PATHS TAKEN BY *ENSAIO: RESEARCH IN EDUCATION AND SCIENCES* JOURNAL TOWARDS SHARING DATA AND SUPPLEMENTARY MATERIALS FOR RESEARCH

In this editorial, we show an evolution of the subjects previously addressed in our editorials about open science practices (see Mendonça & Franco, 2021; Mendonça et al., 2023). This expansion was promoted by periodic studying and training meetings within the journal editorial board, during which we read and further discussed crucial issues, such as guidelines for Open Science and ethics in education research, proposed in organized materials by the National Association of Postgraduate and Research in Education (ANPED, in Brazilian Portuguese). Increasing our understanding of these topics has been paramount in directing our approach on behalf of more responsible and ethical open science in editorial actions.

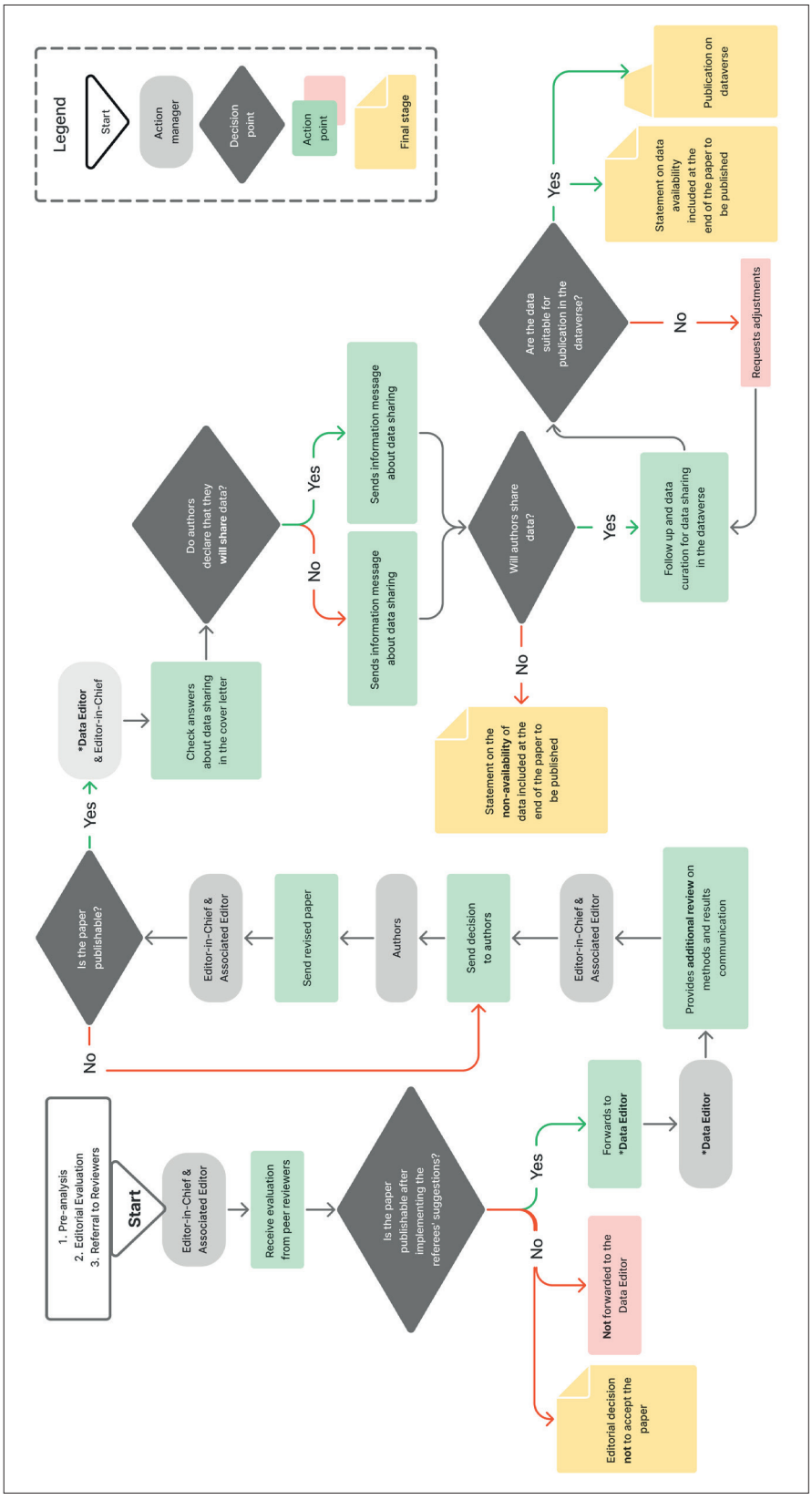
Another significant achievement was the expansion of the data sets published in the Dataverse of *Ensaio: Research in Education and Sciences*, organized according to the guidelines of SciELO (SciELO, 2021, 2023). Additionally, our continuous collaboration with the team of SciELO Data has enabled a constructive and valuable dialogue that contributes to the advancement of data sharing in our research area and helps our editorial team provide better support for the authors who want to publish in *Ensaio*.

We also established and structured a Data Editorship to constantly review and enhance our data-sharing editorial policy (Figure 1). Its activities comprise the evaluation of the possibility of sharing data, including the definition of the opening level and the form of availability, and verifying any pending issues in the communication of data methodology and representation. In addition, the Data Editorship points out possible complementary reviews to the Associate Editors and, when required, invites authors to share their data. After that, it performs the data curation and assists the authors in the sharing process, culminating with the data publication in the Dataverse of the journal.[8]

Through this editorship, we aim to maintain close and constant communication with the authors throughout the editing process, trying to understand their doubts and concerns. This continuous interaction has enriched our comprehension of open science's complexities and motivated us to prepare this editorial. In this process, we observed a great potential to increase our understanding of the subject, and we attempted to contribute to strengthening the conversations about open data issues beyond the pages of this text. We have prepared additional resources, such as the *Guide to preparing a dataset for deposit in a repository: guidelines for the science education research*, to facilitate the preparation of datasets and their submission to the journal's Dataverse.

Our commitment to Open Science transcends the borders of the journal. We actively participate in events promoted by SciELO as guests for relevant debates, for example, the SciELO 25 Years Seminar, where we discussed editorial policies for data in journals of SciELO Brazil.[9] In addition, we also invested in creating modules dedicated to Open Science and ethics in education research, interconnected in disciplines offered to postgraduate students in Brazilian public universities (such as UFOP and UFMG). This initiative intends to prepare new researchers and scientific editors with the required skills and knowledge to promote more open and ethical research and publication practices in science education.

Figure 1. Flowchart of the simplified *Ensaio* journal editorial process, highlighting the actions of the Data Editorship



In the face of the reflections we presented in this editorial and due to the scarcity of discussions about open data in Science Education research, we reiterate the importance of promoting an open and comprehensive debate about this topic in different spaces and formats. We invite the Science Education community to join us in this dialogue. The implications derived from the studies presented here emphasize the urgency in adopting policies for open data sharing not only as a protection for valuable information but also as a catalyst for sustainable knowledge development in research. Complementarily, to sustain this increasing tendency towards qualitative open data, it is fundamental that the epistemological, methodological, legal, and ethical implications are appropriately addressed, providing a sound basis for this emerging practice.

Inspired by the discussions already in progress in areas like Psychology (e.g., DuBois et al., 2018; Roller & Lavrakas, 2018), we consider that the significant impact of data sharing in the Science Education area will depend mainly on the decisions taken by the researchers in terms of the materials and resources chosen to be shared openly, which can result in remarkable methodological benefits when in comparison to a scenario in which the data sharing is scarce.

In conclusion, we hope this editorial and its supplementary material have clarified the types, characteristics, and purposes of open data in research and encouraged the ethical and responsible practice of data sharing in this quadrennium, as expressed in this text and the editorials published in 2021, 2022, and 2023, we are advancing our open science policies, intending to give more transparency to scientific communication based on the practices of publishing original articles associated with research data.

ACKNOWLEDGEMENTS

Federal University of Ouro Preto, for the resources for translating the editorial into English and Spanish. To CNPQ agency for the research productivity grant for the second author.

DECLARATION OF AUTHORSHIP

NATHÁLIA HELENA AZEVEDO: Formal analysis, Conceptualization, Writing – First version, Writing – Review and Editing, Visualization.

PAULA CRISTINA CARDOSO MENDONÇA: Formal analysis, Writing – Review and Editing, Fund raising, Visualization.

REFERENCES

Albagli, S., Clinio, A., & Raychtock, S. (2014). Ciência Aberta: correntes interpretativas e tipos de ação. Liinc em revista, 10(2).

Alexander, S. M., Jones, K., Bennett, N. J., Budden, A., Cox, M., Crosas, M., ... Weber, N. (2019). Qualitative data sharing and synthesis for sustainability science. Nature Sustainability. doi:10.1038/s41893-019-0434-8

Avelino, G., Desposato, S., & Mardegan, I. (2021). Transparency and Replication in Brazilian Political Science: A First Look. Dados, 64.

Bishop, L., & Kuula-Luumi, A. (2017). Revisiting qualitative data reuse: A decade on. SAGE open, 7(1), 2158244016685136.

Borsetti, S., Vidotti, G., Torino, E., & Coneglian, C. (2021). #SejaJUSTOeCUIDADOSO: princípios FAIR e CARE na gestão de dados de pesquisa. In: SALES, Luana Farias; VEIGA, Viviane dos Santos; HENNING, Patrícia; SAYÃO, Luís Fernando (org.). Princípios FAIR aplicados à gestão de dados de pesquisa. Rio de Janeiro: Ibict, 2021. p. 201-214.

Branney, P., Reid, K., Frost, N., Coan, S., Mathieson, A., & Woolhouse, M. (2019). A context-consent meta-framework for designing open (qualitative) data studies. Qualitative Research in Psychology, 16(3), 483-502.

Chauvette, A., Schick-Makaroff, K., & Molzahn, A. E. (2019). Open data in qualitative research. International Journal of Qualitative Methods, 18, 1609406918823863.

Childs S., McLeod J., Lomas E., Cook G. (2014). Opening research data: Issues and opportunities. Records Management Journal, 24, 142.

Corti L., Fielding N., Bishop L. (2016). Editorial for special edition, digital representations: Re-using and publishing digital qualitative data. SAGE Open, 6, 1.

Creswell J. W., Poth C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). Thousand Oaks, CA: Sage.

Crosas, M., Gautier, J., Karcher, S., Kirilova, D., Otalora, G., & Schwartz, A. (2018). Data policies of highly-ranked social science journals. Retrieved from https://dataverse.org/sites/projects.iq.harvard.edu/files/dataverseorg/files/data_policies_of_highly-ranked_social_science_journals.pdf

Centre for Science and Technology Studies (CWTS) & Elsevier. (2017). Open Data: the researcher perspective. https://www.elsevier.com/__data/assets/pdf_file/0004/281920/Open-data-report.pdf

DuBois, J. M., Strait, M., & Walsh, H. (2018). Is it time to share qualitative research data?Qualitative Psychology, 5(3), 380–393.

Elger, K., Biskaborn, B. K., Pampel, H., & Lantuit, H. (2016). Open research data, data portals and data publication–an introduction to the data curation landscape. Polarforschung, 85(2), 119-133.

Foster. (s.d.). Managing and Sharing Research Data. https://www.fosteropenscience.eu/

Freese, J., & Peterson, D. (2017). Replication in social science. Annual Review of Sociology, 43, 147-165.

GIDA, The Global Indigenous Data Alliance. (2018). CARE principles for Indigenous data governance http://gida-global.org/

Gorgolewski, K. J., Margulies, D. S., & Milham, M. P. (2013). Making data sharing count: a publication-based solution. Frontiers in neuroscience, 7, 9.

Jackson, K. C., & Pachter, L. (2023). A standard for sharing spatial transcriptomics data. Cell Genomics, 3(8).

Janz, N. (2016). Bringing the gold standard into the classroom: replication in university teaching. International Studies Perspectives, 17(4), 392-407.

Jaunsen, A. (2018). The state of Open Science in the Nordic countries: Enabling data-driven science in the Nordic countries. Nordic Co-operation. http://norden.diva-portal.org/smash/record.jsf?pid=diva2%3A1257306&dswid=3471

Jesus, D. M. (2019). Integridade na coleta, na produção e na análise de dados. In ANPED. Ética e pesquisa em Educação: subsídios, vol 1. 127-132.

King, G. (1995). Replication, Replication. PS: Political Science and Politics, vol. 28, n.03, pp. 444-452.

King, G. (2007). An Introduction to the Dataverse Network as an Infrastructure for Data Sharing. Sociological Methods and Research, 36, Pp. 173–199.

Mendonça, P. C. C., Franco, L. G. (2021). Editorial: A ciência aberta e a área de Educação em Ciências – perspectivas e diálogos. *Ensaio Pesquisa em Educação em Ciências*, 23, https://doi.org/10.1590/1983-21172021230102

Mendonça, P. C. C., Franco, L. G., Massi, L., & Coelho, G. R. (2023). Editorial: Experiências da revista Ensaio Pesquisa em Educação em Ciências com avaliação por pares aberta. *Ensaio Pesquisa em Educação em Ciências*, 25, https://doi.org/10.1590/1983-21172022240137

Mozersky, J., Walsh, H., Parsons, M., McIntosh, T., Baldwin, K., & DuBois, J. M. (2020a). Are we ready to share qualitative research data? Knowledge and preparedness among qualitative researchers, IRB members, and data repository curators. IASSIST quarterly, 43(4).

Mozersky, J., Parsons, M., Walsh, H., Baldwin, K., McIntosh, T., & DuBois, J. M. (2020b). Research participant views regarding qualitative data sharing. Ethics & human research, 42(2), 13-27.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., ... & Ioannidis, J. (2017). A manifesto for reproducible science. Nature human behaviour, 1(1), 1-9.

National Academies of Sciences, Engineering, and Medicine. (2019). Reproducibility and replicability in science.

Organization for Economic Cooperation and Development. (2007). OECD principles and guidelines for access to research data from public funding. Retrieved from http://www.oecd.org/sti/sci-tech/38500813.pdf

Open Data Handbook. (2013). What is open data? *Open Knowledge International*. Retrieved from http://opendatahandbook.org/guide/en/what-is-open-data/

Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2020). Analysis of open data and computational reproducibility in registered reports in psychology. *Advances in Methods and Practices in Psychological Science*, 3(2), 229–237.

Prosser, A.M., Hamshaw, R.J., Meyer, J., Bagnall, R., Blackwood, L., Huysamen, M., ... & Walter, Z. (2022). When open data closes the door: A critical examination of the past, present and the potential future for open data guidelines in journals. *British Journal of Social Psychology*.

Roller, M. R., & Lavrakas, P. J. (2018). A total quality framework approach to sharing qualitative research data: Comment on Dubois et al. (2018). Qualitative Psychology, 5(3), 394–401.

Ruusalepp R. (2008). A comparative study of international approaches to enabling the sharing of research data. Retrieved from https://www.era.lib.ed.ac.uk/bitstream/handle/1842/3361/Ruusalepp%20Data_Sharing_Report.pdf;jsessionid=AEA3F674EFD77EB4667BA978EC45EE48?sequence=1

SciELO. Guia de preparação de dados de pesquisa [online]. SciELO, 2021. Available from https://wp.scielo.org/wp-content/uploads/Guia_preparacao_pt.pdf

SciELO. Guia de depósito de dados de pesquisa [online]. SciELO, 2023. Available from https://wp.scielo.org/wp-content/uploads/Guia_deposito_pt.pdf

Strasser, C. (2013). Some answers to the most common misconceptions about sharing research data. London School of Economics. https://blogs.lse.ac.uk/impactofsocialsciences/2013/05/06/answers-to-common-misconceptions-about-data-sharing/

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M. & Frame, M. (2011). Data sharing by scientists: practices and perceptions, PLoS One, 6, e21101.

Trenham, C., & Steer, A. (2019). The good data manifesto. In Angela Daly, S. Kate Devitt and Monique Mann. Good Data, 37-53.

Uhlir, P.F., & Schröder, P. (2007). Open Data for Global Science. Data Science

UNESCO (2021). UNESCO Recommendation on Open Science. https://doi.org/10.54677/MNMH8546

van Dijk, W., Schatschneider, C., & Hart, S. A. (2021). Open science in education sciences. Journal of Learning Disabilities, 54(2), 139-152.

Vines, T. H., Albert, A. Y., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., ... & Rennison, D. J. (2014). The availability of research data declines rapidly with article age. Current biology, 24(1), 94-97.

Weller, S. (2023). Fostering habits of care: Reframing qualitative data sharing policies and practices. Qualitative Research, 23(4), 1022-1041.

Wilbanks, J., Boyle, J., Reynolds, W.N. (2006). Introduction to Science Commons. Retrieved from http://sciencecommons.org/wpcontent/uploads/ScienceCommons_Concept_Paper.pdf

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9.

NOTES

1 The terms transparency, reproducibility, and replication often appear in the literature as interchangeable. In this text, we follow the definition proposed by Janz (2016) and King (1995), in which transparency refers to the availability of research materials to the academic community, reproducibility refers to the ability to reproduce a study using available data, while replication is understood as the ability to reproduce a study considering new variables or hypotheses.

2 There is a series of resources and tools online to facilitate the creation of a DMP. FAPESP has already performed a curation of the theme, available at https://fapesp.br/gestaodedados-documentosinteress

3 We highlight works related to the learning sciences, such as van-Dijk et al. 2021.

4 For example (i) https://discover.ukdataservice.ac.uk/qualibank and (ii) https://qdr.syr.edu/.

5 https://dataverse.org/

6 International projects monitoring other repositories in Europe and the United States (e.g., https://www.re3data.org/) indicate the existence of specific repositories dedicated to Life and Nature Sciences, as well as repositories that indicate acceptance of Education area data.

7 Complementing this text, we have prepared the *Guide to preparing a dataset for deposit in a repository: guidelines for the science education research*, in which we cover in detail the technical characteristics of datasets and the readme document. The guide is available on the journal's dataverse https://data.scielo.org/dataverse/brepec.

8 https://data.scielo.org/dataverse/brepec

9 SciELO Data: editorial data policies in SciELO Brazil journals https://25.scielo.org/en/seminars/scielo-data/

Nathália Helena Azevedo
Postdoctoral researcher (Institute for Science Education and Communication), University of Groningen, Groningen, NL. Ph.D. in Science Education (Interunit Program in Science Education, USP, Brazil).
E-mail: helena.nathalia@gmail.com

Paula Cristina Cardoso Mendonça
Associate Professor at the Federal University of Ouro Preto (UFOP), Ouro Preto, Brazil. Ph.D. in Education (Federal University of Minas Gerais, UFMG, Brazil).
E-mail: paulamendonca@ufop.edu.br

Institutional Address:
Centro de Ensino de Ciências e Matemática de Minas Gerais – CECIMIG
Faculdade de Educação | Universidade Federal de Minas Gerais
Av. Pres. Antônio Carlos, 6627 - Pampulha
Belo Horizonte - MG | Brazil
CEP 31270-901
revistaepec@gmail.com

CECIMIG would like to thank National Council for Scientific and Technological Development (CNPq) and Research Support Foundation of the State of Minas Gerais (FAPEMIG) for the funding forediting this article.