

14. BeFAIRandCARE: FAIR AND CARE PRINCIPLES IN RESEARCH DATA MANAGEMENT

Silvana Aparecida Borsetti Gregorio Vidotti¹⁵⁸

Emanuelle Torino¹⁵⁹

Caio Saraiva Coneglian¹⁶⁰

14.1 INTRODUCTION

The current moment in society can be understood from an important evolution of digital technologies, but specifically from an understanding of the role of data in the decision-making process. In general terms, data collection, treatment, analysis, and availability became essential processes so that all sectors of the economy could survive and advance. In addition, the scientific research process itself was strongly impacted by this trend, and started to use, in a very significant way, data analysis and reuse for conducting research.

Such context lead to the development of studies related, for example, to data science, big data and scientific research and communication, which made it possible to understand the role of data in the current moment. Thus, an important aspect for understanding and analyzing the time we live in is the volume of existing data that, when treated and analyzed, are capable of generating great wealth for all processes involved.

It is essential to point out in this context the open data movement, which is mainly linked to government data. This movement demonstrates how the trend towards valuing data can contribute to the transparency of public authorities, as well as to improve the efficiency of services provided to the population. The open data movement is a trend across the planet and, when aligned with people's interests, it is capable of improving the well-being of a community.

In the scientific field, an increasingly valued trend is the open access movement, which aims to provide transparency and openness to the scientific results achieved in the research process. This movement aims for researchers to clearly demonstrate their research results, including the free availability of data collected and results achieved.

Even more broadly, open science is another movement strongly linked to research data availability that, when analyzed from the perspective of publication, demonstrates the potential of making research data available and shared. With the support of data repositories, data journals and data articles, such movement has led to a

158 PhD by the Graduate Program in Education (Unesp). Faculty member in the Department and Graduate Program in Information Science (PPGCI - Unesp) at São Paulo State University - Unesp. E-mail: silvana.vidotti@unesp.br

159 Doctorate student by the Graduate Program in Information (PPGCI-Unesp). Librarian at the Universidade Tecnológica Federal do Paraná (UTFPR). E-mail: emanuelle@utfpr.edu.br

160 PhD by the Graduate Program in Information Science (PPGCI-Unesp). Universidade de Marília (Unimar). E-mail: caio.coneglian@gmail.com

transformation in the way research is carried out, with an impact on the various stages of the scientific process, including funding agencies that have started to demand data management plan.

Based on the trends presented, an appreciation of the aspects involving data in different areas is identified. From government data to research data, performing data management has become essential. In this sense, it is fundamental that all aspects involving the ability of machines to locate, access and interoperate data are understood and well-defined.

Additionally, there are other aspects that must be observed throughout the research process when working with data from specific people and communities, considering the data life cycle. These communities and the research data generated that involve them need to have a special treatment, which ensures data sovereignty, to maximize the benefits and mitigate the impacts, as well as enable adequate treatment for the availability, access and (re) use. Therefore, this book chapter aims to discuss the FAIR and CARE principles for managing human research data from specific communities.

14.2 FAIR PRINCIPLES

FAIR principles (Findable, Accessible, Interoperable and Reusable) were proposed in 2016, as a way of dealing with computational aspects involved in making data available in different contexts, including research data. Such principles were developed understanding a scenario in which, in the same intensity that data become important for the scientific process and the whole Society, there was a difficulty for people and machines to be able to locate and use available data. Thus, “[...] FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals.” (Wilkinson *et al.*, 2016, p. 1).

In the initial proposal of FAIR, Wilkinson *et al.* (2016) point out the importance of computational agents in the environments that store data and highlight the need for these agents to be able to have information about the databases, in an interoperable way and with easy access. In this way, computational agents can help people to locate and use available data. The authors also report that, in environments with a huge number of databases, people depend on computational agents to relate to such data.

Based on this understanding, and having FAIR principles as reference, Australian National Data Service (2020) present FAIR principles in Figure 1.

Figure 1- FAIR Principles



Source: Australian National Data Service (2020).

Figure 1 highlights characteristics linked to each one of FAIR principles, showing elements that must be considered when making data available or when releasing them.

According to GOFAIR (2020), the Findable (**F**) principle assumes that in order for data to be used and/or reused it must be located, activated, readable and processable by humans and computer applications. In addition, it is necessary to adopt persistent identifiers for the data, describe them exhaustively through enriched metadata and make them available in an indexed infrastructure. The Accessible (**A**) principle reflects the capacity of a data set to be accessed and the specifications to do so, including the use of communication protocols, authentication, levels of access and metadata persistence, even if data is not available anymore. The Interoperable (**I**) principle aims to optimize the communication between different systems and the integration of different data sets. For this purpose, data, and metadata need to be readable and adequate to recognized standards and vocabularies, enhancing the link with other standards and including qualified references. Finally, the Reusable principle (**R**) aims at optimizing data reuse process. Data reuse is about how well data and metadata are described, including information on right of use, data provenance and context, in a way to allow data to be combined and reused by other instances.

According to FORCE21 (2020, local. 1 preamble):

[...] through the definition and widespread support of a minimal set of community-agreed guiding principles and practices, data providers and data consumers – both machine and human – could more easily discover, access, interoperate, and sensibly re-use, with proper citation, the vast quantities of information being generated by contemporary data-intensive science.

It is noteworthy that FAIR principles apply to the treatment of data, metadata, and infrastructure to maximize the location, access, interoperability, and reuse of data.

14.3 CARE PRINCIPLES

The technological infrastructure and connectivity increase data value, therefore, principles for the processes of collection, storage, and availability are essential. In this matter, and considering indigenous data sovereignty – presented by Stone and Calderon (2019, local. introduction) as “[...] the rights of Indigenous Peoples and nations to govern themselves and the data about them [...]” -, on which only indigenous peoples have primacy to take decisions, according to their interests and values, Global Indigenous Data Alliance established in 2018, at the International Data Week and Research Data Alliance Plenary, CARE Principles for Indigenous Data Governance.

CARE is an acronym for Collective Benefit, Authority to Control, Responsibility, Ethics.

[...] the ‘CARE Principles for Indigenous Data Governance’ were developed by the Research Data Alliance (RDA) International Indigenous Data Sovereignty Interest Group. They aimed to empower Indigenous Peoples, by shifting the focus of data governance from consultation to values-based relationships that promote equitable Indigenous participation in processes of data reuse, which will result in more equitable outcomes, as well as preserving relationships built on trust and respect. (Carroll *et al.*, 2020b, local. introduction).

Such initiative is based, according to Global Indigenous Data Alliance (2019), on the Declaration of the United Nation on the Right of Indigenous People (Nações Unidas, 2008, p. 1), which recognizes the auto-governance and authority rights of indigenous peoples on their cultural heritage. The “languages, knowledge, practices, technologies, natural resources and territories” are considered by them as indigenous data, often expressed orally and considered essential for their development and rights.

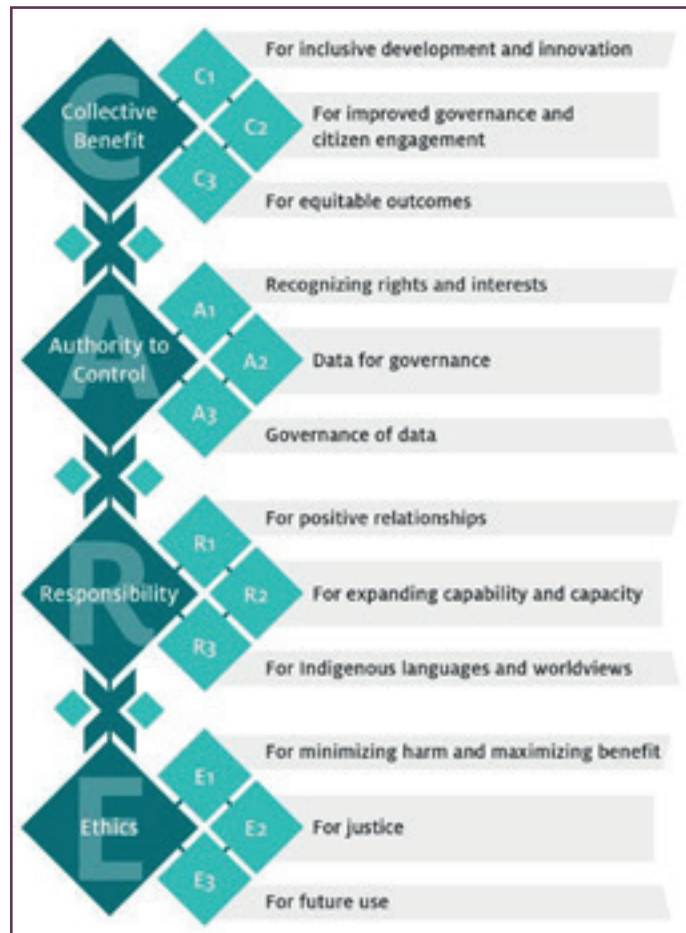
In this way, it has become necessary for indigenous people to establish principles that make it possible to indiscriminately make decisions about their data, since the global movement around open data, whether governmental or research, does not compromise with the aforementioned interests and that data exchange favored by the movement and structured on principles such as FAIR do not mention ethical and cultural characteristics, and/or characteristics from historical context. According to Carroll *et al.* (2020a), it happens due to the tension of indigenous community in protecting their data and interests and supporting initiatives such as openness, data sharing and machine learning, aiming for researchers, managers, and data users to be “fair and care”.

Therefore, CARE principles seek to establish governance over people-drive data. “These principles complement the existing FAIR principles, encouraging open and other data movements to consider both people and purpose in their advocacy and pursuits.” (Global Indigenous Data Alliance, 2019, p. 1). Carroll *et al.* (2020a) emphasize that CARE principles

were designed to complement FAIR principles, aiming to include indigenous peoples so that they can be implemented together; in addition, they emphasize that the governance of indigenous data encompasses the administration and control over data, which includes the processes of collection, storage, analysis, use, and reuse.

The four CARE principles are structured into 12 sub principles associated to it, according to Figure 2.

Figure 2- CARE Principles for Indigenous Data Governance



Source: Carroll *et al.* (2020, p. 5).

The first principle, Collective Benefit (C), establishes that “Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data.” (Global Indigenous Data Alliance, 2019, p. 2). For that purpose, **C1** - For inclusive development and innovation – governments and institutions must support data (re)use by indigenous peoples and communities, aiming at innovation, value generation and local development; **C2** - For improved governance and citizen engagement – data enables the involvement of governments, institutions and citizens, provide transparency and assist in planning, evaluating and decision-making, in addition to providing information on the indigenous peoples’ interest; **C3** - For equitable outcomes – indigenous data are related to their values and can be extended to society, in a way that all must benefit indigenous peoples.

The second principle Authority to Control (**A**) consists in:

Indigenous Peoples' rights and interests in Indigenous data must be recognized and their authority to control such data be empowered. Indigenous data governance enables Indigenous Peoples and governing bodies to determine how Indigenous Peoples, as well as Indigenous lands, territories, resources, knowledge and geographical indicators, are represented and identified within data. (Global Indigenous Data Alliance, 2019, p. 3).

In **A1** - Recognizing rights and interests – indigenous peoples must have their rights and interests in their data and knowledge recognized, which is done through free, prior and manifest consent during data collection, including the data uses, data policies and protocols used during collection; **A2** - Data for governance – indigenous peoples must exercise governance on their data, which must be available and accessible to them; **A3** - Governance of data – indigenous people can develop protocols of governance and access to their data, especially those regarding indigenous knowledge.

The third principle, Responsibility (**R**), establishes that:

Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples' self-determination and collective benefit. Accountability requires meaningful and openly available evidence of these efforts and the benefits accruing to Indigenous Peoples." (Global Indigenous Data Alliance, 2019, p. 4).

In this matter, **R1** - For positive relationships – indigenous data use is possible when based on relationships of respect, reciprocity and trust, so that the researcher is responsible for ensuring that data collected, their interpretation and use guarantee and respect the dignity of indigenous peoples; **R2** - For expanding capability and capacity – the use of indigenous data requires reciprocal responsibility, competence in data with indigenous people and the development of digital infrastructures that enable data collection, management, security, and governance; **R3** - For indigenous languages and worldviews - such resources must generate data based on languages, experiences, values, principles and world perspectives of indigenous peoples.

The fourth principle, Ethics (**E**), "Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem." (Global Indigenous Data Alliance, 2019, p. 5). consists in: **E1** - For minimizing harm and maximizing benefit – minimize damages that can derive from stigmas or deficits related to the indigenous peoples, guiding data collection, treatment, and use in ethical presets, in line with indigenous ethical structures and with rights established through the United Nations Declaration on the Rights of Indigenous Peoples; **E2** - For justice – use ethical processes to address imbalances in power and resources, as well as the way they affect indigenous and human rights; **E3** - For future use – data governance must consider the potential future use based on ethical bases, values, and principles of indigenous peoples, and also express provenance, purpose, rights of use, including limitations, obligations in use and consent in the metadata.

14.4 #BeFAIRandCARE: DISCUSSIONS AND NOTES

Global Indigenous Data Alliance (GIDA)¹⁶¹ expresses, through “#BeFAIRandCARE”, that FAIR and CARE principles are complementary when considering technologies, purposes, and people in open data movements. Thus, while FAIR mainly emphasizes the computational aspects given their relevance so that computational agents can help humans in the face of the expressive volume and complexity of data; CARE, as principles for the governance of indigenous data, emphasizes, above all, people and purpose, considering the relevance of data for the advance, self-determination, and sovereignty of indigenous peoples.

Regarding CARE, although it was designed for the governance of indigenous data, Carroll *et al.* (2020a) highlight that indigenous peoples, nations, people, and communities are actors in contemporary global societies. Therefore, CARE principles address aspects relevant to different populations, such as social minorities, communities and groups, who want or need to maintain different levels of treatment and responsibility for their data use. And, among these aspects, the authors highlight privacy, use, reuse and management, which can constitute elements for the establishment of standards, policies, and agendas.

When considering the activity of scientific research, it is evident the convergence of different movements that determine standards, principles and general practices and their domains, whose peculiarities must be met. Stone and Calderon (2019, local. a journey) reaffirm that “The CARE principles certainly prompt us all to consider the people reflected in data and how our actions with it may impact on them.”

The need to manage research data since planning through the research data management plan is an action increasingly required from researchers by education and research institutions, funding agencies, repositories and scientific and data journals. And, although it may seem like a bureaucratization of scientific work, it allows the researcher and others involved in a research project to plan and record the processes and decisions made during the investigation to document each step, which allows for an optimized management, considering the life cycle of research and research data.

In most cases, institutions list a set of tools and templates to support the researcher in the elaboration of research data management plan, which are important elements for data management, availability and future use. However, such templates are generic and open, When They could guide the researcher in the planning and execution of the research stages already based on guiding principles, among which we highlight FAIR and CARE. This way, the necessary elements are treated during the research process and ensure that data opening is properly done.

Throughout the research data management process, since planning, collection, availability and reuse, it is essential to pay attention to the relevance of these data treatment so that they can respect technological principles that make them are easily processable by computational applications, and, therefore, be reused in robust infrastructures and by humans. On the other hand, human principles linked to people, purposes, and the consequences

161 Available in: <https://www.gida-global.org/care>. Access on: 30 sept, 2024.

that the collection, storage, and availability of data from specific communities can bring to the development of the communities themselves must be respected.

We emphasize, from Carroll *et al.* (2020a), that although CARE principles have been defined for the governance of indigenous data, whose context is recognized as relevant, it is possible to expand them to other specific communities that may also need them so that they have governance and can develop from their data. In this matter, we highlight other social minorities, such as quilombo communities, riverside communities, social settlements, suburbs and LGBTQ+ communities.

Researchers, institutions, funding agencies, governments, and policymakers are encouraged to ensure that the planning and execution of research, as well as the disclosure of its results, are properly carried out, especially when materialized in open research data, in addition to being based on different established and validated principles, among which we highlight FAIR and CARE

Advances in the sense of operationalizing such principles in a computational way are the object of study, involving domains, communities and correlated instances. According to Carroll *et al.* (2020b), Research Data Alliance, through the International Indigenous Data Sovereignty Interest Group and the FAIR Data Maturity Model Work Group, has already started the necessary discussions to operationalize CARE principles together with FAIR principles. From this perspective, they highlight that one of the challenges is the need to apply CARE at all stages of the data cycle. Thus, issues related to the optimization of computational processes for the location, access and (re)use of data, as well as the sovereignty and equitable treatment of data, can be adequately addressed.

However, for this to occur, in addition to the proper treatment of research data, it is necessary for the researcher to adopt daily practices throughout the research process and throughout the data life cycle. This attitude will ensure that the processes take place fairly and carefully.

Therefore, in the context of scientific research and research data, #BeFAIRandCARE can be:

- a. FAIR: make data available in an open way and in accordance with FAIR principles that favor location, access, interoperability and use, with:
 - » adequate and exhaustive use of metadata in data sets and for representation;
 - » adoption of standardized vocabularies;
 - » adoption of persistent identifiers;
 - » selection of proper digital environment for data availability and/or release;
 - » use of open formats, protocols, and standards;
 - » determination of use license;

- » maintenance of reasonable periods of embargo;
- » adequate indication of data provenance;
- » data versioning;
- » data preservation;
- » acknowledgement of ownership when using research data collected by third parties;
- » use of data in accordance with the provisions of the use license.

b. CARE: respect CARE principles, considering the hegemony of specific communities, their world perspective, sovereignty, and governance on data, with:

- » focus on inclusive development;
- » establishment of equitable relationships of trust, reciprocity, and respect;
- » compliance with ethical and legal precepts in the collection, treatment, storage, and availability of data;
- » data identification;
- » protection of rights, interests, values, and culture;
- » participation in governance and control over data;
- » improvement in data representation;
- » training to use data.

Being fairly consists in formalizing data and making it available following good practices and principles so that they can actually be reused by humans and computational applications. And being carefully includes and expands on the previous aspects by properly dealing with research involving human beings, especially specific communities and social minorities. Thus, #BeFAIRandCARE practices must clearly be present in the daily routine of those who work, directly or indirectly, with processes related to the collection, analysis, treatment, storage, and availability of data, especially when they involve human beings.

14.5 FINAL CONSIDERATIONS

FAIR and CARE principles seek to build reliable, fair and responsible data practices, in both management and governance processes, as well as in results and in the quality of available data sets.

It is worth noting that CARE principles are involved in the whole data life cycle, starting in data management plan, passing through the processes of collection, representation, storage, and potential data availability and reuse, respecting the collective benefits, the authority to control, responsibility and ethics. On the other hand, FAIR principles are also linked to the life cycle; however, they focus on technological infrastructure so that data can be findable, accessible, interoperable, and reusable.

Therefore, it is recommended to adapt data management plans to the FAIR and CARE principles, and the adoption by researchers of #BeFAIRandCARE practices throughout the research process and in the data life cycle, especially in research related to people from specific communities for the equitable treatment of data.

In this chapter, the discussion begins by pointing out ways that can be followed in working with research data involving human beings. Such discussions can be in-depth, for example, based on the United Nations (UN) recommendations for human rights, the Sustainable Development Goals (SDGs) and national and international legislation for the management of personal data.

REFERENCES

AUSTRALIAN NATIONAL DATA SERVICE. **FAIR data training and resources**. [S. l.]: ARDC, 2024. Available from: <https://ardc.edu.au/resource/fair-data-training-resources/>. Access on: 30 sept. 2024.

CARROLL, Stephanie Russo; GARBA, Ibrahim; FIGUEROA-RODRÍGUEZ, Oscar L, *et al.* The CARE Principles for Indigenous Data Governance. **Data Science Journal**, [S. l.], v. 19, n. 1, p. 43, nov. 2020a. DOI: <http://doi.org/10.5334/dsj-2020-043>. Available from: <https://datascience.codata.org/articles/10.5334/dsj-2020-043/#:~:text=The%20CARE%20Principles%20are%20a,value%20of%20data%20for%20reuse>. Access on: 5 feb. 2021.

CARROLL, Stephanie Russo; HUDSON, Maui; HOLBROOK, Jarita; MATERECHERA, Simeon; ANDERSON, Jane. **Working with the CARE principles**: operationalizing Indigenous data governance. London: Ada Lovelace Institute, nov. 2020b. Available from: <https://www.adalovelaceinstitute.org/blog/care-principles-operationalising-indigenous-data-governance/>. Access on: 6 feb. 2021.

FORCE21. **Guiding principles for findable, accessible, interoperable and re-usable data publishing version b1.0**. Calfórnia: Force21, 2020. Available from: <https://www.force11.org/fairprinciples#Annex1-1>. Access on: 7 feb. 2021.

GLOBAL INDIGENOUS DATA ALLIANCE. **CARE Principles for Indigenous Data Governance**. [S. l.]: Data Alliance, sept. 2019. Available from: <https://static1.squarespace.com/static/5d3799de845604000199cd24/t/>

5da9f4479ecab221ce848fb2/1571419335217/CARE+Principles_One+Pagens+FINAL_Oct_17_2019.pdf. Access on: 5 feb. 2021.

GOFAIR. **FAIR principles**. [S. l.]: GOFAIR, 2020. Available from: <https://www.go-fair.org/fair-principles/>. Access on: 7 feb. 2021.

NAÇÕES UNIDAS. **Declaração das Nações Unidas sobre os direitos dos povos indígenas**. Rio de Janeiro: Nações Unidas, mar. 2008. Available from: https://www.un.org/esa/socdev/unpfii/documents/DRIPS_pt.pdf. Access on: 5 feb. 2021.

STONE, Paul; CALDERON, Ania. **[Spotlight] CARE Principles**: unpacking indigenous data governance. [S. l.]: opendatacharter, nov. 2019. Available from: <https://opendatacharter.medium.com/spotlight-care-principles-f475ec2bf6ec>. Access on: 6 feb. 2021.

WILKINSON, M., DUMONTIER, M., AALBERSBERG, I. *et al.* The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, [S. l.], v. 3, n. 160018, mar. 2016. DOI: <https://doi.org/10.1038/sdata.2016.18>. Available from: <https://www.nature.com/articles/sdata201618>. Access on: 6 feb. 2021.

How to cite this chapter: VIDOTTI, Silvana Aparecida Borsetti Gregorio; TORINO, Emanuelle; CONEGLIAN, Caio Saraiva. BeFAIRandCARE: FAIR and care principles in research data management. *In*: SALES, Luana Faria; VEIGA, Viviane Santos de Oliveira; VIDOTTI, Silvana Aparecida Borsetti Gregório; HENNING, Patrícia; SAYÃO, Luís Fernando (org.). **FAIR Principles Applied To Research Data Management: Brazilian Experiences**. Brasília, DF: Editora Ibict, 2024. cap. 14, p. 166-176. DOI: 10.22477/9788570131959.cap14.