OPINION PAPER

Artificial intelligence in research and development for sustainability: the centrality of explicability and research data management

Erik Hermann[1] · Gunter Hermann[2]

Abstract

Sustainability constitutes a focal challenge and objective of our time and requires collaborative efforts. As artificial intelligence brings forth substantial opportunities for innovations across industry and social contexts, so it provides innovation potential for pursuing sustainability. We argue that (chemical) research and development driven by artificial intelligence can substantially contribute to sustainability if it is leveraged in an ethical way. Therefore, we propose that the ethical principle *explicability* combined with (open) research data management systems should accompany artificial intelligence in research and development to foster sustainability in an equitable and collaborative way.

Keyword Artificial intelligence · Research and development · Sustainability · Explicability · Research data management

1 Introduction

The increasing computational power and availability of big data have substantially empowered artificial intelligence (AI) in recent years [10]. Thereby, AI – in its simplest sense, defined as "ability of computers or other machines in performing activities that require human intelligence" [1] – has become a transformative power across a wide range of industrial, social, and intellectual contexts [11]. It offers vast opportunities for innovations [4, 5], engenders new solutions to complex problems [39], and enables to address major societal issues, including sustainable development, and environmental sustainability [33, 41, 42]. Particularly, the data-intensive research and applications benefit from AI due to the possibility to process and analyze massive amounts of structured and unstructured (big) data [10, 23]. Big data can be leveraged to foster innovation and corporate social performance as well [6, 16]. Thus, AI and big data in concert provide an unprecedented innovation potential. Chemical research and development (R&D) have derived their insights from data from the very beginning. In light

of the massive amounts of (experimental) data on chemical and physical properties, chemical reactions and structures, and biological activities, the development and deployment of methods from computer science – termed artificial intelligence already then – in chemistry dates back to the 1960s [15]. The long-standing history of computational methods to analyze chemical data and their large scale and scope make AI particularly suitable for and prevalent in chemical R&D. That is reflected by the various areas of application such as toxicity prediction [19, 43], drug discovery and design [20, 36, 37], and life-cycle impact assessment [38], in which AI is utilized. We argue that AI in chemical R&D can substantially contribute to environmental sustainability if AI is harnessed in an ethical way.

Sustainability features a central ethical principle and objective related to the development and deployment of AI [14, 21, 41]. Generally, the increasing pervasiveness and impact AI on the individual, economic, and societal level have sparked the debate of ethical principles guiding AI development and use [8, 13, 14, 17, 21, 29, 31]. The rather fragmented AI ethics landscape [21] consists of recurring principles, such as *transparency*, *beneficence*, and *nonmaleficence* that are of high order, normative, and deontological nature [17, 29] and thus require translation into business practice [29, 31]. Hereinafter, we aim at showing that the ethical principle *explicability*, that is, how AI works (*intelligibility*) and who is responsible for the way AI works (*accountability*), in combination with an open research data

✉ Erik Hermann
hermann@ihp-microelectronics.com

1 IHP – Leibniz-Institut für innovative Mikroelektronik, Frankfurt (Oder), Germany

2 QoD Technologies GmbH, Berlin, Germany

management system should be the focal factors accompanying AI in R&D to promote sustainability.

2 Explicability

Leveraging AI in (chemical) R&D for sustainability is at the core of the *beneficence* principle of AI ethics. Ethical frameworks for AI [14] increasingly propagate that AI should promote individual, social, and environmental good and well being (i.e., *beneficence*), while preventing any risks and harm (i.e., *nonmaleficence*). Unifying these objectives to achieve a "dual advantage" for society [14] is central to the AI-for-social-good perspective [8, 13, 14, 39]. The other three ethical principles inherent in the AI-for-good-perspective are *autonomy* (i.e., self-determination, balanced human and AI agency, power to decide), *justice* (i.e., fairness, avoidance of biases, discrimination, and inequality), and *explicability* (transparency, intelligibility, accountability) [14].

We attach particular importance to the *explicability* principle because it features a proethical condition for enabling or impairing judgments of *beneficence*, *non-maleficence*, *justice*, and *autonomy* [40]. Understanding the functionalities of AI, that is, *intelligibility*, informs evaluations of the other principles by comprehending if and how AI benefits (*beneficence*) or harms (*nonmaleficence*) individuals and society in a fair and unbiased way (*justice*) and by deciding to delegate decisions to AI systems or not (*autonomy*) [14]. *Intelligibility* relates to the epistemological dimension of *explicability* [14] and can be defined as human understanding of a model's function without any need for explaining its internal structure or underlying data processing algorithm [3]. Although it is often used interchangeably with concepts such as *comprehensibility*, *interpretability*, *explainability*, and *transparency*, it is considered as the most appropriate conceptualization [3].

In an AI-driven R&D context, *intelligibility* is multidimensional, since it concerns internal stakeholders within and external stakeholders outside the research institution or company. There is no one-fits-all approach [20], but the audience is decisive [32]. Researchers directly involved in the R&D process need an in-depth understanding of AI models, predictions, outcomes, and underlying data. Although full *intelligibility* of complex AI models might be hard to achieve, researchers should know and understand how AI systems reach predictions (i.e., transparency), why model predictions are acceptable, to what extent they provide new, relevant information, and how reliable they are [20]. When deciding between simpler and more complex, black-box models (i.e., modeling stage), researchers should be able to evaluate the respective AI models' ability to fit data (i.e., predictive accuracy). To foster this evaluation, the number of

model parameters can be reduced (i.e., sparsity), the model prediction-making process can be internally simulated and reasoned (i.e., simulatability) or partitioned (i.e., modularity). At the post hoc analysis stage, fitted/trained models are assessed in terms of what and which relationships they have learned from data (i.e., descriptive accuracy), either on the single (local) prediction level or on the (global) dataset level. Therefore, researchers can look at the importance of certain dataset or variable features for model predictions [32].

However, *intelligibility* cannot be simply transferred from internal to external stakeholders due to proprietary boundaries and intellectual property right restrictions in case of commercial product development [2, 30]. Narrowly related to *intelligibility* is *accountability*, the ethical dimension of *explicability* [14], since judgments about *accountability* require a certain understanding of the underlying processes of AI systems and applications (i.e., *intelligibility*) [7, 26, 28, 31]. *Accountability* can enable a shared responsibility within the organization and towards external stakeholders, which is particularly relevant in the sustainability context. *Accountability* has two temporal perspectives. In hindsight, responsibility is ascribed when something goes wrong. Moreover and even more importantly, AI systems should be responsibly designed and deployed in a forward-looking way [7], which again requires *intelligibility*.

In the following, we present a research data management system that acts on an *explicability* and open science maxim in order to foster collaborative actions in respect to sustainability objectives.

3 Open research data management

The research data management system illustrated in Fig. 1 consists of the AI-driven (chemical) R&D process which is fed with relevant multifaceted experimental and secondary data. The R&D process ranges from the definition of the required properties, the AI-based molecular design and in silico characterization of the relevant properties to the AI-based ranking of the most promising substances and product candidates. The latter provides the basis for follow-up laboratory research, and the corresponding research findings in turn feed the experimental data bases. Thus, a data life cycle with substantial epistemic and methodological advantages can emerge. That is, a broader and permanently refined data bases allow more comprehensive model training, which eventually increases accuracy and validity of AI model predictions. As AI predictions are as accurate and unbiased as the underlying data, consistent, unbiased and curated data sets are a prerequisite for AI-driven R&D to take full effect [9, 36]. Biases can arise from misleading proxy features [3, 34], sparse (small) data [9, 34], or even researchers themselves due to personal preferences and biases and (chemical)
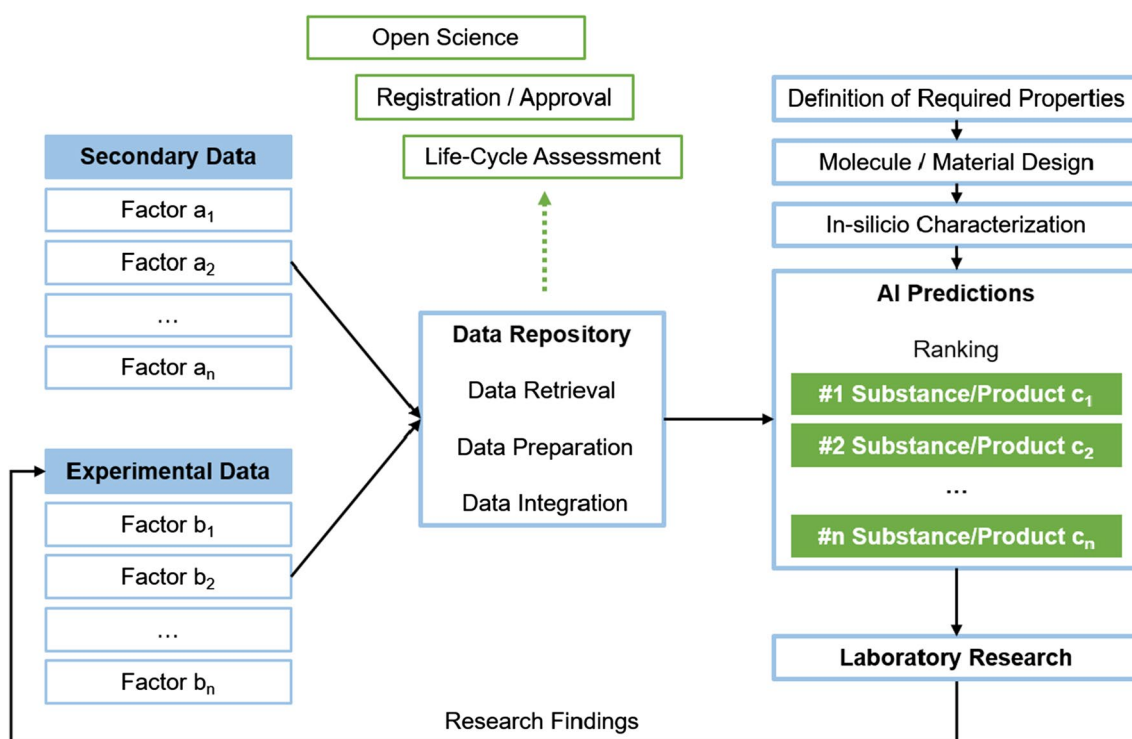
Fig. 1 Research data management system

education [3, 36]. A curated data life cycle can counter these biases.

Moreover, sophisticated data bases and self-learning AI models facilitate to simultaneously investigate a greater amount and variety of research questions as compared to resource-intensive laboratory research that often uses rather inefficient one-parameter-at-a-time methods [36]. That acquires particular importance in light of the complex mixture of thousands of chemicals the environment and humans are exposed to from multiple sources, their interactions, interdependencies, and externalities, and eventually their related risks and adverse effects for humans and the environment [12, 22]. AI-driven R&D processes have the potential to define, study, and detect substances, product candidates, and their properties that are best-suited for areas of applications and surrounding circumstances and environmental factors. Thereby, environmental and social good (*beneficence*) is fostered, while adverse environmental and societal effects are limited (*nonmaleficence*), that is, sustainability is pursued.

Investigating a multitude substances and product candidates further accounts for the *justice* principle, which should also guide environmental sustainability [45]. The *justice* principle entails fairness and the prevention of unwanted/unfair biases [21, 31], sharing benefits and prosperity and fostering solidarity [14]. In the AI-driven R&D context, *justice* implies that the research outcomes (e.g., substances

or products), but also scientific data should equally benefit all stakeholders affected, within and across countries and regions. Given that developing countries are at the higher risks of adverse effects of climate change [18] and detrimental impacts by chemical pollutants [12], *justice* becomes particular important for R&D in the sustainability context. Therefore, countries' environmental and social circumstances and idiosyncrasies should be taken into consideration—an objective that is better pursuable and achievable by means of AI-driven R&D as compared to conventional laboratory research.

Again, intelligible AI methods and effective research data management that transparently document final and preliminary research findings for follow-up research as well as for subsequent life-cycle assessments, registration, and approval processes (see Fig. 1) can be cornerstones for equitable and collaborative sustainability efforts. Eventually, AI methodologies and respective research findings on behalf of (global) the environmental sustainability should be shared with other researchers and/or made entirely publicly available. That is, open data/science approaches and policies [27, 35] should gain center stage. To facilitate global data utilization, research data management systems should follow the FAIR principles, that is, Findable, Accessible, Interoperable, Reusable [44]. Of course, such a transparency and open science policy is an ideal that is limited by proprietary and professional boundaries [2, 30], (perceived) incentives for

researchers [25], and governance and technical requirements [27]. However, *explicability* of the R&D process and data accessibility seem imperative to collaboratively and substantially increase the scale in endeavors to purse sustainability.

4 Conclusion

AI combined with the big data offer unprecedented innovation potential. Backing (chemical) R&D with both provides substantial opportunities to investigate and identify solutions that underpin and accelerate sustainability, that is, AI can be a game changer [24]. The global inequality of environmental pollution and negative side effects require world spanning and collaborative sustainability efforts. Therefore, innovative AI-driven R&D processes and the respective research findings and knowledge should not be exclusively intelligible and accessible by a privileged minority or elite. Since pursuing sustainability is also a highly ethical objective—generally [45] and in respect to the development and deployment of AI [14]—AI-driven R&D for sustainability should be guided by *explicability* and open data/science policies. Eventually, not only the environment constitutes a social good, but also methods and insights aiming at protecting and sustaining it.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Ali, J.M., Hussain, M.A., Tade, M.O., Zhang, J.: Artificial Intelligence techniques applied as estimator in chemical process systems – A literature survey. Expert Syst. Appl. **42**(14), 5915–5931 (2015). https://doi.org/10.1016/j.eswa.2015.03.023

2. Ananny, M., Crawford, K.: Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Med. Soc. **20**(3), 973–989 (2018). https://doi.org/10.1177/1461444816676645

3. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Benneto, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F.: Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion **58**, 82–115 (2020). https://doi.org/10.1016/j.inffus.2019.12.012

4. Borges, A.F.S., Laurindo, F.J.B., Spínola, M.M., Gonçalves, R.F., Mattos, C.A.: The strategic use of artificial intelligence in the digital era: systematic literature review and future research directions. Int. J. Inf. Manage **57**, 102225 (2021). https://doi.org/10.1016/j.ijinfomgt.2020.102225

5. Brynjolfsson, E., Mitchell, T.M.: What can machine learning do? Workforce implications. Science **358**(6370), 1530–1534 (2017). https://doi.org/10.1126/science.aap8062

6. Calic, G., Ghasemaghaei, M.: Big data for social benefits: Innovation as a mediator of the relationship between big data and corporate social performance. J. Bus. Res. **131**, 391–401 (2021). https://doi.org/10.1016/j.jbusres.2020.11.003

7. Coeckelbergh, M.: Artificial intelligence, responsibility attribution, and a relational justification of explainability. Sci. Eng. Ethics **26**(4), 2051–2068 (2020). https://doi.org/10.1007/s11948-019-00146-8

8. Cowls, J., Tsamados, A., Taddeo, M., Floridi, L.: A definition, benchmark and database of AI for social good initiatives. Nat. Mach. Intell. **3**(2), 111–115 (2021). https://doi.org/10.1038/s42256-021-00296-0

9. de Almeida, A.F., Moreira, R., Rodrigues, T.: Synthetic organic chemistry driven by artificial intelligence. Nat. Rev. Chem. **3**(10), 589–604 (2019). https://doi.org/10.1038/s41570-019-0124-0

10. Duan, Y., Edwards, J.S., Dwivedi, Y.K.: Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. Int. J. Inf. Manage **48**, 63–71 (2019). https://doi.org/10.1016/j.ijinfomgt.2019.01.021

11. Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P.V., Janssen, M., Jones, P., Kar, A.K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., Williams, M.D.: Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int J Inf Manage **57**, 101994 (2021). https://doi.org/10.1016/j.ijinfomgt.2019.08.002

12. Escher, B.I., Stapleton, H.M., Schymanski, E.L.: Tracking complex mixtures of chemicals in our changing environment. Science **367**(6476), 388–392 (2020). https://doi.org/10.1126/science.aay6636

13. Floridi, L., Cowls, J., King, T.C., Taddeo, M.: How to design AI for social good: seven essential factors. Sci Eng Ethics **26**(3), 1771–1796 (2020). https://doi.org/10.1007/s11948-020-00213-5

14. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayena, E.: AI4People – An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach. **28**(4), 689–707 (2018). https://doi.org/10.1007/s11023-018-9482-5

15. Gasteiger, J.: Chemistry in times of artificial intelligence. ChemPhysChem **21**(20), 2233–2242 (2020). https://doi.org/10.1002/cphc.202000518

16. Ghasemaghaei, M., Calic, G.: Assessing the impact of big data on firm innovation performance: big data is not always better data. J. Bus. Res. **108**, 147–162 (2020). https://doi.org/10.1016/j.jbusres.2019.09.062

17. Hagendorff, T.: The ethics of AI ethics: an evaluation of guidelines. Minds Mach. **30**(1), 99–120 (2020). https://doi.org/10.1007/s11023-020-09517-8

18. IPCC: Global warming of 1.5 °C. An IPCC Special Report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. IPCC (Intergovernmental Panel on Climate Change) (2018).

19. Idakwo, G., Luttrell, J., Chen, M., Hong, H., Zhou, Z., Gong, P., Zhang, C.: A review on machine learning methods for in silico toxicity prediction. J. Environ. Sci. Health C Toxicol. Carcinog. **36**(4), 169–191 (2019). https://doi.org/10.1080/10590501.2018.1537118

20. Jiménez-Luna, J., Grisoni, F., Schneider, G.: Drug discovery with explainable artificial intelligence. Nat. Mach. Intell. **2**(10), 573–584 (2020). https://doi.org/10.1038/s42256-020-00236-4

21. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**(9), 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2

22. Johnson, A.C., Jin, X., Nakada, N., Sumpter, J.P.: Learning from the past and considering the future of chemicals in the environment. Science **367**(6476), 384–387 (2020). https://doi.org/10.1126/science.aay6637

23. Jordan, M.I., Mitchell, T.M.: Machine learning: trends, perspectives, and prospects. Science **349**(6245), 255–260 (2015). https://doi.org/10.1126/science.aaa8415

24. Kaplan, A., Haenlein, M.: Rulers of the world, unite! The challenges and opportunities of artificial intelligence. Bus. Horiz. **63**(1), 37–50 (2020). https://doi.org/10.1016/j.bushor.2019.09.003

25. Kwon, S., Motohashi, K.: Incentive or disincentive for research data disclosure? A large-scale empirical analysis and implications for open science policy. Int. J. Inf. Manage **60**, 102371 (2021). https://doi.org/10.1016/j.ijinfomgt.2021.102371

26. Lepri, B., Oliver, N., Letouzé, E., Pentland, A., Vinck, P.: Fair, transparent, and accountable algorithmic decision-making processes. The premise, the proposed solutions, and the open challenges. Philos. Technol. **31**(4), 611–627 (2018). https://doi.org/10.1007/s13347-017-0279-x

27. Link, G., Lumbard, K., Conboy, K., Feldman, M., Feller, J., George, J., Germonprez, M., Goggins, S., Jeske, D., Kiely, G., Schuster, K., Willis, M.: Contemporary issues of open data in information systems research: considerations and recommendations. Commun. Assoc. Inf. Syst. **41**(25), 587–610 (2017). https://doi.org/10.1016/j.anireprosci.2016.02.027

28. Martin, K.: Ethical implications and accountability of algorithms. J. Bus. Ethics **160**(4), 835–850 (2019). https://doi.org/10.1007/s10551-018-3921-3

29. Mittelstadt, B.D.: Principles alone cannot guarantee ethical AI. Nat. Mach. Intell. **1**(11), 501–507 (2019). https://doi.org/10.1038/s42256-019-0114-4

30. Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L.: The ethics of algorithms: mapping the debate. Big Data Soc. **3**(2), 1–21 (2016). https://doi.org/10.1177/2053951716679679

31. Morley, J., Floridi, L., Kinsey, L., Elhalal, A.: From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Sci.

Eng. Ethics **26**(4), 2141–2168 (2020). https://doi.org/10.1007/s11948-019-00165-5

32. Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., Yu, B.: Definitions, methods, and applications in interpretable machine learning. Proc. Natl. Acad. Sci. **116**(44), 22071–22080 (2019). https://doi.org/10.1073/pnas.1900654116

33. Nishant, R., Kennedy, M., Corbett, J.: Artificial intelligence for sustainability: challenges, opportunities, and a research agenda. Int. J. Inf. Manage **53**, 102104 (2020). https://doi.org/10.1016/j.ijinfomgt.2020.102104

34. Rich, A.S., Gureckis, T.M.: Lessons for artificial intelligence from the study of natural stupidity. Nat. Mach. Intell. **1**(4), 174–180 (2019). https://doi.org/10.1038/s42256-019-0038-z

35. Rüegg, J., Gries, C., Bond-Lamberty, B., Bowen, G.J., Felzer, B.S., McIntyre, N.E., Soranno, P.A., Vanderbilt, K.L., Weathers, K.C.: Completing the data life cycle: using information management in macrosystems ecology research. Front. Ecol. Environ. **12**(1), 24–30 (2014). https://doi.org/10.1890/120375

36. Schneider, G.: Automating drug discovery. Nat. Rev. Drug Discov. **17**(2), 97–113 (2018). https://doi.org/10.1038/nrd.2017.232

37. Schneider, G.: Mind and machine in drug design. Nat. Mach. Intell. **1**(3), 128–130 (2019). https://doi.org/10.1038/s42256-019-0030-7

38. Song, R., Keller, A.A., Suh, S.: Rapid life-cycle impact screening using artificial neural networks. Environ. Sci. Technol. **51**(18), 10777–10785 (2017). https://doi.org/10.1021/acs.est.7b02862

39. Taddeo, M., Floridi, L.: How AI can be a force for good. Science **361**(6404), 751–752 (2018). https://doi.org/10.1126/science.aat5991

40. Turilli, M., Floridi, L.: The ethics of information transparency. Ethics Inf. Technol. **11**(2), 105–112 (2009). https://doi.org/10.1007/s10676-009-9187-9

41. van Wynsberghe, A.: Sustainable AI: AI for sustainability and the sustainability of AI. AI Ethics (2021). https://doi.org/10.1007/s43681-021-00043-6

42. Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S.D., Tegmark, M., Fuso Nerini, F.: The role of artificial intelligence in achieving the sustainable development goals. Nat. Commun. **11**, 233 (2020). https://doi.org/10.1038/s41467-019-14108-y

43. Vo, A.H., Van Vleet, T.R., Gupta, R.R., Liguori, M.J., Rao, M.S.: An overview of machine learning and big data for drug toxicity evaluation. Chem. Res. Toxicol. **33**(1), 20–37 (2020). https://doi.org/10.1021/acs.chemrestox.9b00227

44. Wilkinson, M.D., Dumontier, M., Aalsbersberg, I.J., Appleton, G., Axton, M., Mons, B.: The FAIR guiding principles for scientific data management and stewardship. Sci. Data **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

45. Zagonari, F.: Environmental sustainability is not worth pursuing unless it is achieved for ethical reasons. Palgrave Commun. **6**, 108 (2020). https://doi.org/10.1057/s41599-020-0467-7