Looking Back to the Future: A Glimpse at Twenty Years of Data Science

LILI ZHANG (iD)

ABSTRACT

This paper carries out a lightweight review to explore the potentials of data science in the last two decades and especially focuses on the four essential components: data resources, technologies, data infrastructures, and data education. Considering the barriers of data science, the analysis has been mapped into four essential components, highlighting priorities and challenges in social and cultural, epistemological, scientific and technical, economic, legal, and ethical aspects. As a result, the future development of data science tends to shift toward datafication, data technicity, infrastructuralism, and data literacy empowerment. The data ecosystem, at the macro level, has also been analyzed under the open science umbrella, providing a snapshot for the future development of data science.

CORRESPONDING AUTHOR:

Lili Zhang

Computer Network Information Center, Chinese Academy of Sciences, CN

zhll@cnic.cn

1. INTRODUCTION

Data science, namely the science of data and science facilitated by data, has significantly changed how we work, study, and live. Over the recent two decades, statistics show a growing interest in data science (see Figure 1) that remained relatively steady early in the new millennium before a constant increase in the most recent decade. Moreover, the incredible power embedded in data science ignites the worldwide digital revolution in which everyone enters the unprecedented big data era (Science International 2015).
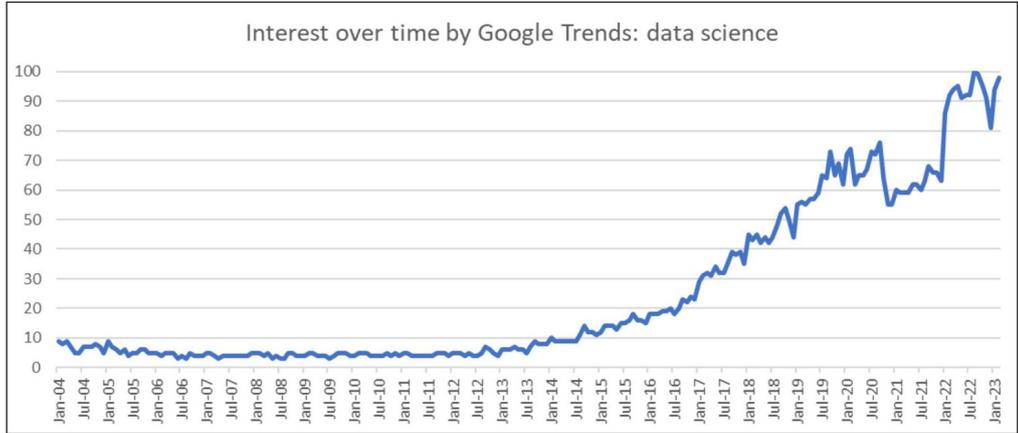
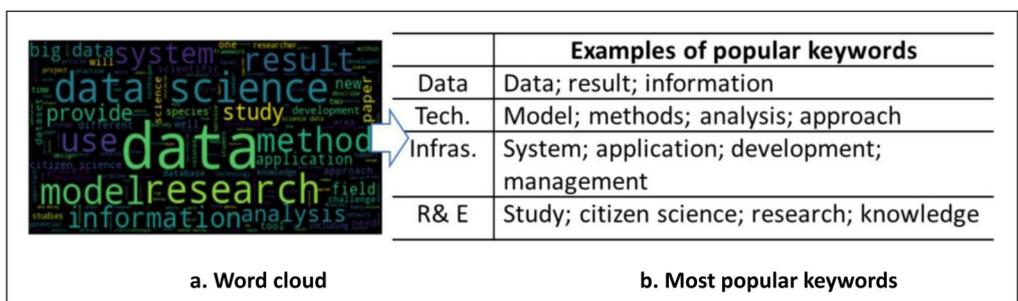**Figure 1** Interest over time by Google Trends: data science around the world (23 Feb. 2023).

This paper carries out a lightweight literature review to unveil data science in the last 20 years. It employs scientometric methods to identify data science essentials and introduces examples illustrating multidimensional challenges. It also includes an analysis of future trends and proactive actions. Hopefully, this paper will help reflect on data science work and support its future development under open science guidance (UNESCO 2021).

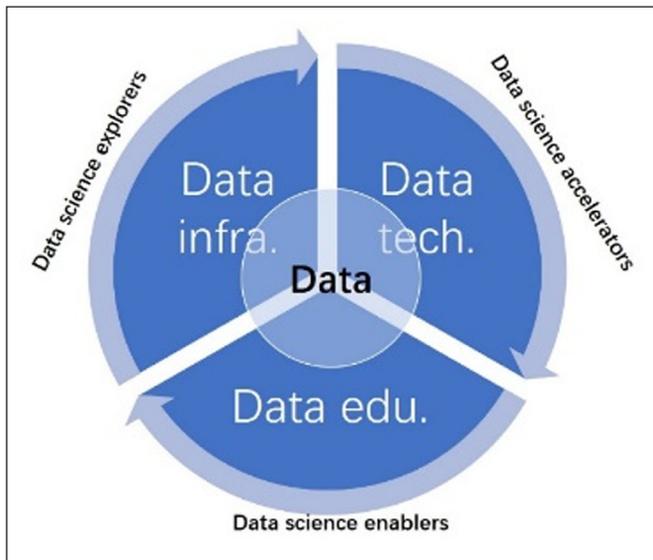# 2. EXPLORING THE FULL POTENTIAL OF THE BIG DATA ERA

From no data and little data to big data (Borgman 2015), the new era brings the most significant advances to data science. According to Google Trends, the most relevant data science topics searched fall into different categories, such as data, technology, infrastructure, and education (see Table 1). Another word frequency analysis of the data science publications on the Web of Science platform shows similar results (see Figure 2). Therefore, we employ the four essential components (see Figure 3), namely data resources, technologies, data infrastructures, and data education, to depict data science trends.

CATEGORIES	TOP 25 MOST POPULAR TOPICS RELATED TO DATA SCIENCE (GOOGLE TRENDS, 23 FEB. 2023)
Data	data; big data
Technology	Python; analytics; learning; data analysis; machine learning; analysis; artificial intelligence
Infrastructure	computer; machine; project; engineering
Education	data science; course; university; job; master's degree; master of science; bachelor's degree; salary
Others	science; computer science; statistics; business

Table 1 Relevant data science topics searched on Google.

Note: Data are captured from Google Trends and manually cleaned.

	Examples of popular keywords
Data	Data; result; information
Tech.	Model; methods; analysis; approach
Infras.	System; application; development; management
R& E	Study; citizen science; research; knowledge

a. Word cloud   **b. Most popular keywords**

**Figure 2** Relevant data science topics in Web of Science publications.

Figure 3 Four essential components of data science.

As the new underpinning of the digital revolution, data have become hot spots in different areas. Many types of research, far away in the luminous galaxies (Combes 2021) or up close in the micro-ecosystems (ISC 2021), are driven by data-related research resources. Data-centric research drives paradigm shifts to address transparency and openness throughout the data life cycle. Data improve the performance of global governance by steadily aggregating social capital (Malgonde & Bhattacherjee 2014) and proliferate the digital economy by reducing costs and facilitating value appreciation of data-relevant assets.

Nevertheless, data are not alone. The supporting information and communication technologies (ICT) provide opportunities to facilitate the full exploitation of data value. From analysis to analytics (Cao 2017), technologies transfer data from facts to insights. From data mining to artificial intelligence, technical development drives data to understand the past and envisage the future (Provost & Fawcett 2013). From centralized data control to cloud solutions, technologies support better data management for scalable research. Thus, as catalysts and boosters, technologies improve data performance by adding value to data within the whole life cycle and every workflow.

Further, data infrastructures facilitate the deployment of technologies, thus supporting data use and reuse. Technical infrastructures provide connectivity for data exchange, capacities for data storage, computing for data processing, algorithms and software for data analysis, portals and technical support for the accessibility and reusability of data services. Thus, data and technologies evolve together with the supporting data infrastructures. Take the Scientific Data Program of the Chinese Academy of Sciences (CAS SDP), for example (Figure 4) (Zhang et al. 2021): initiated in the 1980s and sustained for over 30 years, the CAS SDP has established data infrastructures to support disciplinary research. The data infrastructures have been growing from early databases to the data grid, data engineering, data cloud, and an open data system of systems, as well as leading the exploration of co-building a Global Open Science Cloud (CODATA 2021; CSTCloud 2021). The evolving construction of the supporting e-infrastructures reflects the iteration of data and technologies that has adapted to the dynamic research scenarios of different research periods.
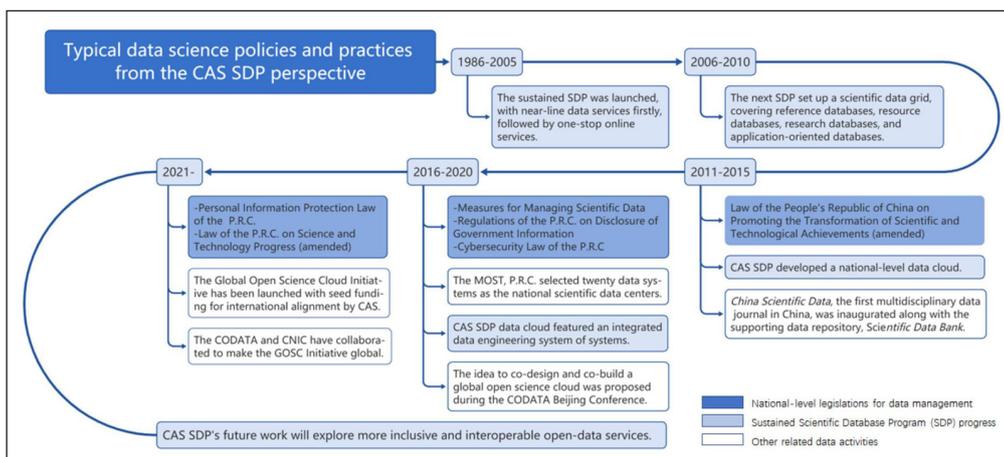
Figure 4 Typical data science policies and practices from the CAS SDP perspective.

Moreover, data science education includes popular colleague programs, such as the iSchool applied data science master program and the data science training for citizens, such as the CODATA-RDA summer school and online training on the MOOC, Coursera, and EdX platforms, etc. Building data capacities throughout different programs has enabled better use of data, technologies, and data infrastructures. In China, for example, the Ministry of Education issued the *Action Plan for Promoting the Development of Big Data* (2015), followed by a steadily increasing number of university data science majors and a boom in data-relevant careers nationwide. Data science also helps create new types of professionals, such as data scientists (Dhar 2013), data engineers, data architects, and others. Therefore, data education sustains data science careers and helps cultivate citizens with better literacy to surf the wave of the digital revolution.

Above all, these essential data science components work together, contributing to the scientific research, societal development, digital economy, and better alignment within and across communities, domains, and regions.

3. DATA SCIENCE CHALLENGES AND PRIORITIES

In addition to the remarkable contributions to the digital revolution, data science also raises grand challenges in social and cultural, epistemological, scientific and technical, economic, legal, and ethical dimensions.

First, several data factors contribute to the challenging barriers, such as the multiple sources, the enormous scale, the diversified formats, the intrinsic nature, the uncertain value behind them, and the way of communication. Social and cultural contexts complicate data interpretation (Shanks & Corbitt 1999), making data abuse a potential threat to data reuse. Furthermore, data address rights protection, such as privacy, security, ownership, and intellectual property (Taylor 2017; Calzada & Almirall 2020). Thus, necessary closeness and default openness should clarify their boundaries to explore the full potentials of open data. Moreover, nourished culture, evolving governing rules, and data ethics (Floridi & Taddeo 2016; Gundersen 2017; Vydra et al. 2021) should work together to bridge digital gaps, promote research integrity, and support future envisions vigorously.

Second, as the essential game-changers for science, promising and robust technologies support cutting-edge data exploration by combining certain data into tailored scenarios. For example, long-tailed value exploration for massive data, lack of data, and sensitive data coexist in scenarios. Thus, any data curation should be set in the lifelong data cycle to save chaos, such as real-time massive data integration in Earth Sciences for Sustainable Development Goals (SDGs) research (Guo et al. 2021). Moreover, epistemological thinking (Kempeneer 2021) should be involved in the technical design, pulling data results out of the black box of technologies. Other important concerns also include technology neutrality, algorithm ethics, legislation redesigns, and global alignments (George & Walsh 2022) to better serve the public interests, like in the case of ChatGPT (van Dis et al. 2023).

Third, sustained governance models with social, cultural, and economic considerations are critical for successful data science, especially for data infrastructures. Infrastructures are comprehensive systems that combine data, technologies, hardware, software, and others (Mayernik et al. 2017) for service delivery. Potential players in a data infrastructure ecosystem may include the infrastructures (service providers), resource suppliers (both for data and technologies), users, and funding agencies. Therefore, data infrastructures rely on robust business models to balance all possible stakeholders' interests and ensure data science components running systematically and healthily. Considering the 'Matthew effect' (Merton 1968; Merton 1988; Bol, de Vaan & van de Rijt 2018), chances of future opportunities may be amplified by accumulative advantages, such as current construction scale, prestige, and popularity. Thus, vulnerabilities may lie in the future opportunities to raise funding as newly established infrastructures, to share data as sovereign resources suppliers, and to access data services as niche-demand users. Technical concerns may include exploring systematical design for flexible and extensible services, effective functionality deployment for efficient data curation, and environmentally friendly development.

Fourth, data education (NASEM et al. 2018; Wise 2020) should build capacities for specialists and citizens. Data science overlaps with computer science and statistics and focuses on real-world problem solving. Thus, data professionals' challenges may include establishing proper social identity, connecting with other relevant societal roles, and mapping real problems into tailored curricula, such as 'precision education and individualized learning' (Luan et al. 2020). Cultivating citizens may also be popular, such as coding for data cleaning, analysis, and visualization, but that is not enough. Data science education, both professional and amateur training, should aim bigger, covering the whole life cycle of data in an open-science manner to embrace inclusive and responsible data science in the future.

Specifically, no data challenges come alone, and no data science essentials can break the silos independently. Instead, data challenges require collaborative support from technologies. Moreover, infrastructures will facilitate data sharing and technology implementation. Educating data professionals and the social community will empower data fully by "leaving no one behind" (UN Sustainable Development Group 2022). Therefore, solutions should consider the four data science essentials together, not segregated ones.

4. FUTURE VISIONS AND NEXT STEPS

Challenges also bring opportunities. Considering data as the global public good (CODATA et al. 2019), data science may assume many future responsibilities, with inevitable trends toward datafication, data technicity, infrastructuralism, literacy empowerment, and others. The following subsections elaborate on each of these issues.

From data to datafication (Mayer-Schönberger & Cukier 2013; Van Es & Schäfer 2017; Mejias & Couldry 2019), quantitated data activities prevail in the big data world. Datafication pulls data from traditional statistics to big data analytics with facts, knowledge, and wisdom. Deeply rooted in the digital revolution, datafication contributes to the booming digital economy while encountering societal and cultural conflicts. To better harness the power of datafication, we should embrace open science (UNESCO 2021) more than ever. A series of guidelines should be followed, such as the FAIR principles (Wilkinson et al. 2016), the CARE principles (Carroll et al. 2020), and the TRUST principles (Lin et al. 2020), as well as others. Open data involves sharing for reuse and closing for protection, commercial and noncommercial models, scientific and pragmatic explorations, and close connections among the research community, social enterprises, and citizens. To open or to close may not be contradictory, but there are inevitably frictions and gaps. For the sake of good science, rules should clarify data boundaries, especially highlighting ways for grey data reuse (Borgman 2018). Research integrity and data ethics are also necessary for responsible open science. For example, Indigenous data sovereignty (Carroll, Rodriguez-Lonebear & Martinez 2019), democratic accountability (Gurumurthy, Chami, & Bharthur 2016), and other moral aspects of data are to balance the interests of potential stakeholders for sustained lifelong data curation.

Furthermore, constitutive technicity (Gallope 2011; Ash 2012; Ducassé & Lee 2014; Wiktionary 2022) dramatically tightens data and technologies. 'Technicity' depicts the prevalence of technology deployments in data management. 'Constitutive' emphasizes that these technologies transit from outsiders to those closely engaged with lifelong data management intrinsically. Data technicity fastens the pace of value extraction from data and even exceeds human visions in many cases (Silver et al. 2017). The evolving technical design should be interoperable across machines and inclusive to humans. Enhanced collaborative research models and international alignment are to follow, affirming the transparency, flexibility, robustness, and intelligibility in technical development to face the ever-growing data deluge locally and globally. And the FAIR use of technologies should follow the open science paradigm, such as the cases in tackling natural hazards, health crises, and climate change and achieving the UN SDGs.

Data infrastructuralism tends to merge data and technology into streaming and scalable services. 'Infrastructuralism' (Breu & Leo 2022; Brehm 2022), adopted here as a neutral concept, refers to the centrality and materiality of data infrastructures. Future data infrastructures will be incredibly important in integrating multiple-sourced data and complex technologies for user-friendly services. Furthermore, infrastructuralism highlights the predominant roles of data infrastructures in coordinating data science essentials. Guided by open science, data

infrastructuralism will overturn the traditional business models, and provide a reciprocal environment for research and innovation with everyone involved. To better serve as the engines of global research, future facilities should fit into the growing need for enhanced and open data infrastructures, assembling cutting-edge technologies and optimized data resources. Furthermore, these infrastructures should follow interest-balanced and cost-effective models to leverage the responsibilities and rights of all potential stakeholders. Future data infrastructures should also pinpoint collaboration, openness, interconnection, and inclusiveness to reshape trustworthy and reliable worldwide science and technology.

In addition, data education extends literacy empowerment to proliferate data science and cultivate the whole society. Skills training (such as knowledge of programming, algorithms, and systems) is useful to capture the exponentially increasing value of data. At the same time, data literacy (Wolff et al. 2016; Gummer & Mandinach 2015) may also consider epidemiology, policies, participants, impacts, and sustainability to serve particular training objectives. Possible courses should be diversified, covering data culture, rights, and ethics for decent data reuse; data epidemiology, critical thinking, and data skills for efficient data reuse; and data accountability, metrics, and data audits for reliable data reuse. Different training pieces should educate data professionals and citizens, thus establishing shared values on data science. Data education will also endure interactive processes to help everyone get ready for the changing world.

Meanwhile, besides the four essentials, future data science also calls for a full data picture at the macro level. Accordingly, the data ecosystem should be established based on mutual trust. Under the umbrella of open science, future data science will work efficiently and systematically as an ecosystem, with data, technology, education, and others integrated through infrastructures. Thus, key actions may include establishing and maintaining an open science environment for the data ecosystem, involving potential stakeholders under sound management strategies (i.e., interest-balanced models) and sustained models (i.e., fair and efficient reward systems), opening dialogues between communities, and collaborating on open science and data initiatives. In addition, future data ecosystems should encourage the data-sharing culture and enhance the global alignment between physical and virtual facilities to support the data flow of enormous research scopes across domains and regions. Surely, based on a harmonized data ecosystem, data science and the essentials will help us prepare deeply and widely for the adventurous data journey forward.

5. CONCLUSIONS

The recent two decades of data science have been long and exciting, full of difficulties and boundless potential. Looking back into the human history of thousands of years, two decades of data science is extremely short. However, it is of great significance. The transit to the fourth paradigm of scientific research, the global wave of the digital economy accompanied by the rapid rise of many developing economies, the dramatic development of the global village, and the polarization of digital integration and the digital divide are just a few examples. Nevertheless, the great charm of data science lies in the science and permeates the social lives of everyone every day. The power to master the double-edged data sword will advance data science explosively.

As this paper elaborates, data, technology, infrastructure, and education contribute jointly to the four-wheeled wagon of data science. The four essentials will together effectively consolidate and enhance the construction and development of future data science. Meanwhile, the call for open science provides rich soil for the healthy development of the whole data ecosystem. Therefore, looking into the future, we will embrace a better world driven by open data resources, responsible technologies, open infrastructures, and inclusive data education.

DATA ACCESSIBILITY STATEMENT

Data are captured from the Web of Science with selected articles entitled 'data science.' There are no strict constraints on publication times to trace the theme evenly. As a result, 4,131 pieces of records are returned, including 88 records earlier than 2000. Among them, 3,490 non-null-value publication abstracts are taken as valid textual results for word frequency

analysis and word cloud visualization by Python. Data are available at www.webofscience.com [Last accessed 24 February 2023].Refined data and python code are available at: https://doi.org/10.57760/sciencedb.07847.

ACKNOWLEDGEMENTS

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR AFFILIATION

Lili Zhang orcid.org/0000-0003-1847-6683
Computer Network Information Center, Chinese Academy of Sciences, CN

REFERENCES

Ash, J. 2012. Technology, technicity, and emerging practices of temporal sensitivity in videogames. *Environment and Planning A: Economy and Space*, 44(1): 187–203. DOI: https://doi.org/10.1068/a44171

Bol, T, de Vaan, M and **van de Rijt, A.** 2018. The Matthew effect in science funding. *Proceedings of the National Academy of Sciences*, 115(19): 4887–4890. DOI: https://doi.org/10.1073/pnas.1719557115

Borgman, CL. 2015. *Big Data, Little Data, No Data*. Cambridge, MA: MIT Press. DOI: https://doi.org/10.7551/mitpress/9963.001.0001

Borgman, CL. 2018. Open data, grey data, and stewardship: Universities at the privacy frontier. *Berkeley Technology Law Journal*, 33(2): 365–412.

Brehm, W. 2022. Podcasting and education: Reflections on the case of FreshEd. *ECNU Review of Education*, 5(4): 784–791. DOI: https://doi.org/10.1177/20965311221094860

Breu, C and **Leo, JRD.** 2022. Infrastructuralism. *symplokē*, 31(1–2). Available at: https://www.symploke.org/infrastructuralism/ [Last accessed 15 February 2023].

Calzada, I and **Almirall, E.** 2020. Data ecosystems for protecting European citizens' digital rights. *Transforming Government: People, Process and Policy*, 14(2): 133–147. DOI: https://doi.org/10.1108/TG-03-2020-0047

Cao, L. 2017. Data science: A comprehensive overview. *ACM Computing Surveys*, 50(3), Article 43: 1–42. DOI: https://doi.org/10.1145/3076253

Carroll, SR, Garba, I, Oscar, L, et al. 2020. The CARE principles for Indigenous data governance. *Data Science Journal*, 19(1): 43. DOI: https://doi.org/10.5334/dsj-2020-043

Carroll, SR, Rodriguez-Lonebear, D and **Martinez, A.** 2019. *Indigenous data governance: Strategies from United States Native nations*. *Data Science Journal*, 18(1): 31. DOI: https://doi.org/10.5334/dsj-2019-031

CODATA. 2021. Global Open Science Cloud. Available at: https://codata.org/initiatives/decadal-programme2/global-open-science-cloud/ [Last accessed 6 December 2022].

CODATA, CODATA IDPC, CODATA and CODATA China High-level International Meeting on Open Research Data Policy and Practice, et al. 2019. The Beijing Declaration on Research Data. *Zenodo*. DOI: https://doi.org/10.5281/zenodo.3552330

Combes, F. 2021. Science with SKA. arXiv:2107.03915 [astro-ph.CO]. SF2A 2021: 238–242. DOI: https://doi.org/10.48550/arXiv.2107.03915

CSTCloud. 2021. Global Open Science Cloud. Available at: https://www.cstcloud.net/gosc.htm [Last accessed 6 December 2022].

Dhar, V. 2013. Data science and prediction. *Communications of the ACM*, 56(12): 64–73. DOI: https://doi.org/10.1145/2500499

Ducassé, P and **Lee, D.** 2014. Technics and the philosopher. *Diacritics*, 42(1): 25–44. DOI: https://doi.org/10.1353/dia.2014.0002

Floridi, L and **Taddeo, M.** 2016. What is data ethics? *Philosophical Transactions of the Royal Society A*, 374(2083): 20160360. DOI: https://doi.org/10.1098/rsta.2016.0360

Gallope, M. 2011. Technicity, consciousness, and musical objects. In: Clarke, D and Clarke, E (eds.), *Music and Consciousness: Philosophical, Psychological, and Cultural Perspectives*, 47–64. 1st ed. New York: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199553792.003.0030

George, A and **Walsh, T.** 2022. Artificial intelligence is breaking patent law. *Nature*, 605: 616–618. DOI: https://doi.org/10.1038/d41586-022-01391-x

Gummer, ES and **Mandinach, EB.** 2015. Building a conceptual framework for data literacy. *Teachers College Record*, 117(4): 1–22. DOI: https://doi.org/10.1177/016146811511700401

Gundersen, LC. 2017. Scientific integrity and ethical considerations for the research data life cycle. In: Gunderson, LC (ed.), *Scientific Integrity and Ethics in the Geosciences*, 133–153. Washington, DC: American Geophysical Union and John Wiley and Sons. DOI: https://doi.org/10.1002/9781119067825.ch9

Guo, H, Chen, F, Sun, Z, Liu, J and **Liang, D.** 2021. Big Earth Data: A practice of sustainability science to achieve the Sustainable Development Goals. *Science Bulletin (Beijing)*, 66(11): 1050–1053. DOI: https://doi.org/10.1016/j.scib.2021.01.012

Gurumurthy, A, Chami, N and **Bharthur, D.** 2016. Democratic accountability in the digital age. IT for Change. Available at: https://www.ids.ac.uk/publications/democratic-accountability-in-the-digital-age/ [Last accessed 14 February 2023]. DOI: https://doi.org/10.2139/ssrn.3875297

International Science Council (ISC). 2021. *Science and Society in Transition: ISC Action Plan 2022–2024*. Available at: https://council.science/actionplan/ [Last accessed 3 December 2022].

Kempeneer, S. 2021. A big data state of mind: Epistemological challenges to accountability and transparency in data-driven regulation. *Government Information Quarterly*, 38(3): 101578. DOI: https://doi.org/10.1016/j.giq.2021.101578

Lin, D, Crabtree, J, Dillo, I, et al. 2020. The TRUST principles for digital repositories. *Scientific Data*, 7(1): 144. DOI: https://doi.org/10.1038/s41597-020-0486-7

Luan, H, Geczy, P, Lai, H, et al. 2020. Challenges and future directions of big data and artificial intelligence in education. *Frontiers in Psychology*, 11: 580820. DOI: https://doi.org/10.3389/fpsyg.2020.580820

Malgonde, O and **Bhattacherjee, A.** 2014. Innovating using big data: A social capital perspective. *Twentieth Americas Conference on Information Systems*, Savannah, Georgia.

Mayer-Schönberger, V and **Cukier, K.** 2013. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. Boston, MA: Houghton Mifflin Harcourt.

Mayernik, MS, Hart, DL, Maull, KE, et al. 2017. Assessing and tracing the outcomes and impact of research infrastructures. *Journal of the Association for Information Science and Technology*, 68: 1341–1359. DOI: https://doi.org/10.1002/asi.23721

Mejias, UA and **Couldry, N.** 2019. Datafication. *Internet Policy Review*, 8(4). DOI: https://doi.org/10.14763/2019.4.1428

Merton, RK. 1968. The Matthew effect in science. *Science*, 159(3810): 56–63. DOI: https://doi.org/10.1126/science.159.3810.56

Merton, RK. 1988. The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *isis*, 79(4): 606–623. DOI: https://doi.org/10.1086/354848

National Academies of Sciences, Engineering, and Medicine (NASEM), Division of Behavioral and Social Sciences and Education, Board on Science Education, et al. 2018. *Data Science for Undergraduates: Opportunities and Options*. Washington, DC: National Academies Press. Available at: https://www.ncbi.nlm.nih.gov/books/NBK532765/ [Last accessed 13 February 2023].

Provost, F and **Fawcett, T.** 2013. Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1): 51–59. DOI: https://doi.org/10.1089/big.2013.1508

Science International. 2015. *Open Data in a Big Data World*. Paris: International Council for Science (ICSU), International Social Science Council (ISSC), the World Academy of Sciences (TWAS), InterAcademy Partnership (IAP).

Shanks, G and **Corbitt, B.** 1999. Understanding data quality: Social and cultural aspects. In: *Proceedings of the 10th Australasian Conference on Information Systems*, 785. New Zealand: Victoria University of Wellington.

Silver, D, Schrittwieser, J, Simonyan, K, et al. 2017. Mastering the game of Go without human knowledge. *Nature*, 550(7676): 354–359. DOI: https://doi.org/10.1038/nature24270

Taylor, L. 2017. What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2): 2053951717736335. DOI: https://doi.org/10.1177/2053951717736335

UN Sustainable Development Group. 2022. *Operationalizing Leaving No One Behind*. Available at: https://unsdg.un.org/resources/leaving-no-one-behind-unsdg-operational-guide-un-country-teams [Last accessed 12 February 2023].

UNESCO. 2021. *UNESCO Recommendation on Open Science*. Available at: https://en.unesco.org/science-sustainable-future/open-science/recommendation [Last accessed 5 December 2022].

van Dis, EAM, Bollen, J, Zuidema, W, et al. 2023. ChatGPT: Five priorities for research. *Nature*, 614(7947): 224–226. DOI: https://doi.org/10.1038/d41586-023-00288-7

Van Es, K and **Schäfer, MT.** 2017. *The Datafied Society: Studying Culture through Data*. Amsterdam, Netherland: Amsterdam University Press. DOI: https://doi.org/10.1515/9789048531011

Vydra, S, Poama, A, Giest, S. 2021. Big data ethics: A life cycle perspective. *Erasmus Law Review*, 14(1): 24. DOI: https://doi.org/10.5553/ELR.000190

Wiktionary. 2022. technicity. Available at: https://en.wiktionary.org/wiki/technicity [Last accessed 15 February 2023].

Wilkinson, M, Dumontier, M, Aalbersberg, I, et al. 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Nature Scientific Data*, 3: 160018. DOI: https://doi.org/10.1038/sdata.2016.18

Wise, AF. 2020. Educating data scientists and data literate citizens for a new generation of data. *Journal of the Learning Sciences*, 29(1): 165–181. DOI: https://doi.org/10.1080/10508406.2019.1705678

Wolff, A, Gooch, D, Cavero Montaner, JJ, et al. 2016. Creating an understanding of data literacy for a data-driven society. *Journal of Community Informatics*, 12(3): 9–26. DOI: https://doi.org/10.15353/joci.v12i3.3275

Zhang, L, Downs, RR, Li, J, et al. 2021. A review of open research data policies and practices in China. *Data Science Journal*, 20(1): 3. DOI: https://doi.org/10.5334/dsj-2021-003