



Measuring Data Reusability in the Open Science Era

Lili Zhang^a

^aComputer Network Information Center, Chinese Academy of Sciences
E-mail: zhll@cnic.cn, 86-10-58812519

Received: 09 August 2024; Accepted: 20 August 2024

Open data, among the key pillars of Open Science, is one of the drivers for science and society. To promote open data, FAIR metrics should be ready to measure the performance of data-sharing works. However, considering the complex process of lifelong data management, how can we guarantee metrics of reusability for data that could objectively and truly reflect the reusability of data? The current data metrics may focus on different aspects of data, which include quality-based metrics, such as data accuracy, completeness, and usability measurement; impact-based metrics, including interest-based metrics, such as visits to databases and tweets in social media, and effort-sparing actions, such as downloads; knowledge-based metrics, such as citations in publications and patents; and value-based metrics, such as those highly praised by their functionality as social capital, or direct monetary return on capital assets driven by open data work. Considering the lifelong data management process, basic principles should be developed to balance different roles and maximize the total benefits. This way, inspired by Ranganathan's laws of library science, considering the intrinsic and extrinsic value of data and the current measuring practices adopted, the basic guiding principles will be developed and discussed. The core values of open science are also mapped to ensure the openness and inclusiveness of the principles and to guarantee a better flow of data across the science community and society.

Keywords: Open data, data metrics, data reusability, evaluation principles, Open Science

Introduction

As the new currency¹ of the big data era, data prevails in the research community and society. As both input and output throughout research lifecycles, data proliferate scientific research by ensuring the competitiveness and innovativeness of science through big data analytics. Data also consolidates research by demonstrating the reproducibility and reliability of deliverables. The value of open data is multifaceted. Boosted by data, the big data era witnesses an emerging data economy, increasing data capital, enhanced data literacy empowerment, and growing efforts to foster a data-sharing culture². However, in facing the data challenges of tackling health crises, natural hazards, and the global digital divide and in achieving the United Nations' Sustainable Development Goals, the call for open data is pressing in all domains. Thus, a better understanding of current open data work, such as through monitoring and evaluation, is an important issue.

Data metrics deploy diversified measuring methods to check data status, evaluate the performance of data-sharing work, thus demonstrating the levels of reusability of data. Open data metrics will ensure a sustained open data world by fully explaining data

reusability through combined qualitative and quantitative methods. Hence, this paper will begin by identifying emerging dilemmas facing data metrics. Then, relevant works with examples of data metrics will be summarized. Thus, gaps to be bridged are analyzed, and possible frameworks for future data metrics development are discussed. The core values of the *UNESCO Recommendation on Open Science*³ are also mapped into the framework to ensure its resilience in tackling open data metrics work now and in the future.

Dilemmas in measuring data reusability

Measuring the performance of data and data work provides insights into data development and thus illuminates future plans for data work. However, current measurements still face diverse challenges in developing appropriate methods and adopting data metrics in complex scenarios.

Open data deficit

Data-sharing principles usually address the responsibilities of sharing data to the greatest extent for public-funded data⁴ since significant investments are expected to promote data sharing. The Data Management Plan⁵ is one tool that demonstrates the

important role of facilitating data sharing in public-funded research. To measure the visible values of data sharing, the European Commission⁶ sets a good example by measuring the direct market size of open data. However, open data investment doesn't necessarily lead to visible outputs in a comparative measurable period. Sometimes, data value generation is subject to the long term. Meanwhile, values, such as social and cultural benefits, are sometimes invisible, either unmonitored or unmeasurable. Thus, to convince funding agencies and incentivize data contributors, data metrics should extend their scope of measurement to unveil scientific, social, and economic value, including visible and invisible, real and promising benefits that particular open data work may bring.

Data credit debt

Metrics usually measure the frequency of data reuse, such as downloads, visits, citations, and others. However, the importance of different reuses is uneven. Visits reflect potential attention to a dataset but do not necessarily lead to real reuse. Compared with visits, downloads are more positive signs of in-depth reuse. Citations go even further, unveiling further reusability within the scientific community. However, reusing value in the same metrics still varies. For example, data could be cited for different purposes or by papers of uneven impacts, indicating different values of data reusability, such as the case in citation inflation⁷. Thus, metrics bring data credit, but different credits are not always comparable. Some might surpass others regarding metrics quality, while bubbles may occur among others.

One-size-fits-all solutions

Data and data work are quite different across domains and regions. Time, geospatial, technological, management factors, and others also matter. In real practices, we usually compromise efficiency and effectiveness by using one-size-fits-all solutions, even though we know situations may change. Efficiency measures the plausibility of the one-size-fits-all solution, while effectiveness measures the appropriateness of fitting into different scenarios. Therefore, we need to leverage efficiency and effectiveness by aligning with other data management strategies to "measure what matters."⁸ Generally, the value of data is central to most metrics⁹. However, different scenarios may address particular issues. For example, "cloud-first" based data security, data

residency, and sovereignty are highlighted in Canada¹⁰. European Commission¹¹ emphasized the demand for open data's promising quality and quantity by economic and social value. Other open data expectations also include guidelines on decision-making and future development in Africa¹², a consolidated foundation for "open and transparent" government development in the US¹³, and boosting scientific innovation and social development in China¹⁴. Thus, the conceptualization of certain open data metrics should reflect the values appreciated in the scenarios.

Trapped in metrics

Once data metrics are fixed up, proactive steps will follow. Like the sunflower effect where sunflowers face the sun¹⁵, metrics may function quite similarly: data work will most likely chase and act as guided by the metrics. Never make open data work trapped in metrics. Considering this situation, the design and development of data metrics should be flexible and inclusive to leverage the interests of stakeholders. Metrics should also be robust and resilient to tackle data challenges, incentivize data contributors, and exemplify all potentials for a better world of shared data.

Shadow data sharing

Shadow banking¹⁶ refers to derivative financial institutions that take responsibility for financial banks without sufficient supervision. Similarly, there might be active data reuse outside current monitoring under legal frameworks. For example, data might be reused for black-box situations that are excluded under the current evaluation monitoring system, and that is why the trust crisis of AI¹⁷ debate arises recently. Robust measurement should consider such situations and develop technical and ethical standards, protocols, and others to guarantee the decent sharing of data and proper acknowledgment to the data owners for their contribution.

Incentives failure

Inappropriate design, development, adoption, or interpretation of data metrics may lead to serious incentive failure. For instance, inappropriate methods design cannot capture all valuable dimensions of data sharing, thus leading to negative data development as well; a lack of responding actions regarding measuring results will weaken the incentives for data-sharing efforts; unfair exemplars of data champions and data models may lead to stereotypes, aggregating

data divide among current practices. Therefore, incentives may work only if the data metrics are managed and processed along with the full data lifecycle.

Relevant works

To deal with diversified dilemmas, current data measurements are generally categorized into two groups. One group focuses more on measuring the quality of data, such as adopting FAIR metrics and others. Serving as a prerequisite for open data, such metrics also unveil the potential capabilities of future data reusability. Quality metrics may also involve the assessment of affiliated algorithms, tools, platforms, and collaborative work in preparing open data, such as the practices in CoreTrustSeal¹⁸ and e-Government Openness Index¹⁹, among others.

Other metrics focus more on the interactive responses from potential stakeholders, which is the data impact. According to particular value of reusing, impact metrics can be sub-categorized as follows: **① interest-based data metrics**, which usually focus on the attention to datasets via measuring the frequency of visits, page views, downloads, and others involving social interests. These assessments can tell people’s interests in datasets, somewhat unveiling the social value to citizens, but without a clear description of the data reusability. **② Knowledge-based data metrics**. Data citation and data cited in papers linked datasets with new knowledge discovery. Citation-based data reusability depicts the knowledge reproduction roadmap, thus demonstrating the scientific value of data, which is worth further exploration, such as the practices in data citation index²⁰ and others. **③ value-based data metrics**. These metrics usually measure the economic value of data, such as the

open data contribution to the growth of the digital economy, revenue, and the job market.

However, in real-data assessment work, quality-based and impact-backed data metrics are sometimes independent or merged, as Table 2 exemplifies. Quality metrics focus on intrinsic and extrinsic characteristics of data to ensure the smooth sharing of data. Quality factors include data’s completeness, accuracy, readiness, usability, and operational quality control, such as data rights protection, data resources, governance, ICT maintenance, and portal usability. Quality metrics heavily rely on the performance of data generated, cured, and shared by data contributors. Besides, impact metrics extensively focus on the response of potential stakeholders, such as monitoring the scale of reuse in scientific and social domains, and others, such as participation and collaboration centered around data, guiding data policies, and other involvement regarding the data ecosystem.

Implementing data metrics through the streaming process

Above all, different measurements unveil the capabilities and potentials of open data across scientific, economic, and social dimensions. However, the key to a successful open data measurement relies on several factors, such as the design of metrics methods that align with particular objectives, the adoption of metrics in scenarios, the interpretation of metrics results, and the lessons learned for further data sharing and data metrics work. Fig. 1 shows the full process of implementing data metrics.

Conceptualize innovative open data models and measuring methods. Measuring the performance of data sharing is not the ultimate goal, but enhanced capabilities for reuse are. Therefore, the objectives of



Fig. 1 — Process to implement data metrics

Categories		Examples of measurable variables
Quality metrics	Data quality Operational quality	Data availability, data completeness, data accuracy, data readiness, data usability, FAIRness ²¹⁻²³ Accessibility, the usability of affiliated tools, portals ²⁴ , and platforms ¹⁸
Impact metrics	Interests-based metrics Knowledge-based metrics	Visits ²⁵⁻²⁶ , downloads ²⁷⁻²⁸ , social media attention (i.e. Almetrics ²⁹⁻³⁰) Data citations (citation motivation ³¹ , citation content analysis ³² , citation graph ³³⁻³⁴ , data index ³⁵ , data citation index ²⁰ , informal and formal data citations ³⁶ , etc.)
	Value-based metrics	Data capital ³⁷ , data economy (i.e., market size, price ³⁸ , value), legal rights ³⁹ , etc.

Table 2 — Examples of measuring data reusability

Author(s)	Metrics	Essentials	Categories
FAIRplus ⁴⁰⁻⁴¹	Dataset Maturity Model	<ul style="list-style-type: none"> · Content-related · Representation and format · Hosting environment capabilities 	Data quality
Moraga <i>et al.</i> , 2009 ⁴²	SPDQM	<ul style="list-style-type: none"> · Intrinsic: Data have quality in their own right; · Operational: The system must be accessible but secure; · Contextual: Data quality must be considered within the context of the task at hand; · Representational: Interpretable, easy to understand, concisely, and consistently represented · Quality characteristics: Traceability; Currentness; Expiration; Completeness; Compliance; Understandability; Accuracy 	Data quality
Çaldağ & Gökalp, 2022 ⁴³	Model for Open Data Capability Maturity Assessment	<ul style="list-style-type: none"> · Process dimensions including strategic governance, organizational management, stakeholder engagement and data governance (metadata, data integration, privacy, and security, data storage, and archive, data quality), open data management (business understanding, open data discovery, digitalization and preparation, prototyping and evaluating, license management, open data deployment, and publishing, maintenance, IT management), etc. · capability dimensions 	Data quality
Ingwersen & Chavan, 2011 ⁴⁴	Data usage index	Searched records , download frequency, record number, search events, download events, dataset number, search density, download density, usage impact, interest impact, usage balance, usage score, interest score	Data impact
Grace M. Begany <i>et al.</i> , 2017 ⁴⁵	A model for open health data ecosystem	Subject, Object/Objective, Tools, Rules, Division of Labor, Community, and Outcome	Data impact (Health sciences)
Benitez-Paez, <i>et al.</i> , 2018 ⁴⁶	A conceptual framework to improve open geographic data re-usability in cities	<ul style="list-style-type: none"> · community of reuse (efficient feedback resources, comments by data users, notification systems, data quality rating by users) · data users' identification and demand · reuse-focused legal terms · impact (local level, data user communities, geographic approach) 	Data impact (Open government data)
Data. europa.edu ⁴⁷	Open data maturity	<ul style="list-style-type: none"> · policy, impact, portal, quality 	Data quality and impact
Kubler <i>et al.</i> , 2018 ¹⁹	eGovernment Openness Index	<ul style="list-style-type: none"> · Basic data set indicators · Data openness indicators (complete, primary, timely, accessible, machine processable, non-discriminatory, nonproprietary, license-free) · Transparency indicators (government, data reusability, understandability, authenticity) 	Data quality and impact
Abella <i>et al.</i> , 2019 ⁴⁸	Meloda 5	Participation & Collaboration legal licensing, access to information, technical standards, standardization, geolocation content, and updating frequency of data, dissemination, and reputation	Data quality and impact

any data metrics should be shifted accordingly, together with targeted goals adjusted to scenarios. Innovative data sharing models, such as data publishing and others, cutting-edge technology adoption in facilitating the accuracy, reliability, and reproducibility of data, and sound governance strategies smoothly facilitating the sharing of data, as well as others, should all gain proper incentives in the development of metrics methods.

Data are highly aligned with research activities but kept independent for metrics. Although highly aligned with research activities, data are not always affiliated products of research. Rather, data are equally important as other research deliverables,

which may bring competitive intelligence, knowledge, and wisdom to nourish science⁴⁹. Thus, open data should be taken as an independent contribution with fair acknowledgment of data work.

Data may share characteristics similar to research, such as being scientific and deploying rigorous methods and innovative technologies, whereas innovation is not always necessary in open data work. Thus, the principles for evaluating data must be in line with but not limited to regular research work. For instance, unique data, such as those capturing natural phenomena data in climate and Earth science or archaeological sites, are valuable regardless of their creativity. Criteria measuring data should rely mostly

on data quality, such as the clear and reliable method applied, validated technologies addressing transparency and reproducibility, and envisioned data reusability⁵⁰. Diversified data indicate values, such as basic research inputs, final deliverables, or innovative knowledge discovery. Thus, data metrics should be robust enough to cover the broadest scope of data work while making fair acknowledgments.

Comprehensive interpretation of metrics results. Data metrics are used to measure the performance of data work currently while casting light on the capabilities of future potentials. Then, the use of such measuring results must be very careful. Any bias will lead to an imbalance in incentivizing future data work. Considering the sunflower effect in data metrics, adopting and interpreting data metrics should be intelligible in many ways.

① **Unveil the best open data work and get such work incentivized enough.** The top concern on metrics is to tell the good from the bad. So, the interpretation should first highlight data champions.

② **Set up best-practice models to follow.** Besides, interpretation usually follows with advertising and promotion, pinpointing best practices to follow, showcasing what's important, and how to open data successfully.

③ **Mild reminder of data gaps to bridge together with possible ways forward.** Any comparison between different open data works must be prudent, assuming targeted datasets are comparable in referred dimensions. Therefore, we can easily identify gaps and be enlightened by the highlighted practices.

④ **Predict future data potentials.** Any successful adoption and interpretation of data metrics not only tells the current capabilities of data but also unveils future data potentials based on current quality and reusability. Therefore, interpreting data metrics is illuminative and promising, with possible guidance promoting data work.

Learn from measuring and for managing. The ultimate goal of data measurement is to incentivize potential stakeholders for better data sharing and sustained data development. Therefore, lessons learned in data metrics work should nourish future data and data metrics. To fulfill this goal, we need to explore the stakeholder approach to balance their interests. Any sustained data metrics should consider and balance the interests of potential stakeholders. For

example, data producers would probably expect the broadest data sharing with the least effort in curating and sharing data. In contrast, users would welcome data of the highest standards with the least payment. So, the governance should employ effective data market rules to stimulate the vitality of open data game players, ensuring proper rights protections of data owners while enabling better data reusability.

Moreover, take care to break any silos when metrics are delivered. Metrics are always used for comparison; thus, we know what's good and bad. However, it should not rebuild walls to isolate any of these players in the data ecosystem. The best practices to follow cannot be static stereotypes but are all subject to the evolving, changing digital world. Thus, the actions following learning should offset tradeoffs by appropriately using the sunflower effect in measuring data.

Guiding principles under the open-science umbrella

Open data sets itself in the tide of the open science movement and should share open-science core values accordingly, including "quality and integrity, collective benefits, equity and fairness, diversity and inclusiveness."³ Under the umbrella of open science, data metrics take all their endeavors to promote data sharing throughout the full measuring process. Based on the relevant work reviewed, several principles may be mapped into the essential processes of implementing data metrics.

Understandable. Understand your data, your data metrics methods, the results of your metrics, and the lessons you will learn from them for future data and data metrics and the stakeholders behind promoting all these data works. A comprehensive understanding of all these above will guide future data work by pragmatically promoting open data.

Fairness. Metrics should be carefully targeted, designed, and developed to fairly incentivize potential stakeholders to promote data sharing and encourage positive feedback in future data sharing. Metrics should prioritize data sovereignty, intellectual property, and all efforts⁵¹ to facilitate data capture, management, and sharing.

Rigorous. The rigorousness of data metrics is based on the design of methods and development within the full data measuring process. First, avoid reinventing the wheels. Design and development of new methods should build on the milestones so far and make sure the data metrics are not merely countable but are of great importance. Second,

measure with management. As emphasized, data metrics are not simply measurements of data performance but together with highlighted best practices, data champions, and capacity building for better engagement of potential stakeholders in the long term. Thus, the measures should consider best practices and provide guidance for future data management and sharing. Third, involving the full measuring process into a long-term cycle by learning smartly from current practices. No data metrics are static, but should be involved and adjust to the dynamic world and changing data needs. Learning from data metrics is not only for future measuring but also raising awareness of future data exploration.

Prosperous. Successful metrics can tell not only the current reusability but also future potential. Thus, the design of data metrics should not only heavily rely on current quantitative indicators but also compromise quantitative indicators, unveiling the potential of future data and open data work. Adopting data metrics is subject to change and should be adaptive to the involved research environment. There might be tradeoffs in balancing the importance of current data contribution and future possibilities, but it surely will lead to a prosperous future for data development.

The four principles can be neatly mapped with the UNESCO's Recommendation³. First, "understandable" is the prerequisite for the quality and integrity of open data, and "rigorous" provides pragmatic ways to fulfil these objectives. Besides, "fairness" is the basic rule to address collective benefits, highlighting equity. The spirit of "diversity and inclusiveness" is embedded in the stakeholder approach and is foreseen in the future, as indicated as "prosperous." Therefore, data metrics align with the essentials of open science to help tackle data and research challenges.

Future envisions

We may still take several steps further to consolidate the prosperous future data metrics.

Firstly, smart data policies and data ethics adoption call for fair and rigorous data metrics. The policy and ethical framework set up basic rules governing open data and data metrics work. Rules should include guidance on proper management between reserved data rights and enlarged data sharing, toolkits to measure the dynamic boundaries of openness and closeness of data, investments, initiatives to consolidate data production quality, and adequate incentives to encourage all possible data sharing.

Besides, we should take full advantage of cutting-edge technologies for responsible data metrics by a better understanding the measuring work. ICT frontiers, such as AI, can identify the invisible relationships between research objects through huge data computing in a swift manner. Thus, adopting AI might help enhance the reliability and reproducibility of data and data metrics as validation. Block chain highlights trustworthiness by smart contracts that may illuminate precious data and data metrics work. Interpreting data metrics linked by big data analytics may unveil untouched data characteristics that can cast light on future data work. However, adopting information technologies should always be responsible, white-boxed, and subject to monitoring and evaluation to guarantee the integrity and quality of research.

Measurement work should be incorporated into the data lifecycle as a compulsory component. In the traditional data lifecycle, monitoring and evaluation might not be included in the full data management process. On the importance of positive strengths in promoting data work, metrics may also be included to complete the data lifecycle. Therefore, data metrics can help learn from data practices, either good or bad, identify gaps and data vulnerability, thus reducing open data risks and enabling data sharing.

Development of comprehensive and interoperable systems to stimulate multifaceted data value and ensure these could be reflected fairly in metrics. As mentioned above, data sharing may benefit scientific, social, and economic domains. Quality-based and knowledge-based metrics mostly reflect scientific value. Social and economic benefits are unveiled either by interest-based or value-based metrics. Emphasizing different dimensions and characteristics of data requires tailored metrics to accompany them. Thus, comprehensive metrics should consider all valuable considerations and weigh the importance accordingly.

Open dialogues break silos, narrow down data divide, and proliferate further data work. The open dialogues may be those between metrics reviewers and reviewers, different data metrics designers, open-data collaborative teams across borders and regions, or multiple stakeholders' voices within the data ecosystems. Through such open dialogues, the echo chamber will gain mutual trust and co-develop the open data ecosystem. Interoperable metrics among different research systems can extend the understanding of current data work and ensure their long-term prosperity.

Last but not least, nourish the culture of open science in pursuing prosperous data metrics and, ultimately, a better open-data world. Open science is prevailing, revolutionizing research fundamentals and thus boosting open data. Based on the spirit of open science, contributions to better data metrics may include the development of open-science infrastructures to facilitate metrics implementation, open metrics knowledge and methods to avoid re-inventing the wheels, open engagement of societal actors to address the needs of potential stakeholders fairly and open dialogue with other metrics system for an understandable, fair, rigorous and prosperous data future.

Conclusion

This paper identifies several dilemmas in measuring data work, including open data deficit, data credit debt, one-size-fits-all solutions in changing scenarios, embarrassment of being trapped in metrics, and shadow data sharing out of current monitoring systems. In the face of these challenges, relevant works revealed that current data metrics are categorized as quality- and impact-based metrics. Quality-based metrics focus more on the quality of data and affiliated data facilities. Impact metrics mostly focus on the reusability counts received by sharing data, such as interest-based data metrics, knowledge reproduction-based metrics, and value-based metrics. In reality, data metrics sometimes merge different data indicators for a comprehensive interpretation of data. Moreover, this paper identified the key process in implementing data metrics, including fixed-up metrics goals, design of metrics methods, adoption and interpretation, and lessons learned for further data development. Guiding principles that drive such data metrics work include being understandable, fair, rigorous, and prosperous. Mapping with the *UNESCO Recommendation on Open Science (2011)*³ affirmed its guiding strategies, while a summary of future work also includes suggestions on the adoption of data policies and ethics, cutting-edge technologies, enriched data lifecycle, interoperable data systems, open dialogues and data culture towards co-developing an open-data ecosystem. Future studies will address quantitative methods to validate the theories and discussions mentioned above.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (No.72104229). I would like to thank my colleagues, Ms. Xueting Li and

Yandi Li, for collecting preliminary materials for the scoping research and reviewing the paper.

Reference

- 1 Gates C and Matthews P, Data is the new currency, In *Proceedings of the 2014 New Security Paradigms Workshop*, (2014) 105-116.
- 2 Cannon M, Chandramouliswaran I, Day L, *et al.*, The State of Open Data 2022. Available at <https://www.digital-science.com/resource/the-state-of-open-data-2022/> (Accessed on 18 June 2024).
- 3 United Nations Educational, Scientific and Cultural Organization (UNESCO). The UNESCO Recommendation on Open Science. Paris, France: UNESCO (2021). Available at <https://unesdoc.unesco.org/ark:/48223/pf0000379949> [Accessed on 15 June 2024].
- 4 CODATA, CODATA IDPC, CODATA China High-level International Meeting on Open Research Data Policy and Practice, Hodson S, Mons B, Uhlir P and Zhang L, The Beijing Declaration on Research Data, *Zenodo*, (2019). <https://doi.org/10.5281/zenodo.3552330>.
- 5 NSF Directorate for Engineering. Data Management Plans Guidance for Principal Investigators. (2018). Available at https://www.nsf.gov/eng/general/ENG_DMP_Policy.pdf [Accessed on 13 June 2023].
- 6 European Commission. Creating Value through Open Data: Study on the impact of re-use of public data resources (2015) 1-112. Luxembourg, Publications Office of the European Union. 10.2759/328101.
- 7 Petersen A M, Pan R K, Pammolli F, and Fortunato S, Methods to account for citation inflation in research evaluation, *Research Policy*, 48 (7) (2019) 1855-1865.
- 8 European Commission, Directorate-General for Research and Innovation, Peters I, Frodeman R, Wilsdon J. *et al.*, Next-generation metrics – Responsible metrics and evaluation for open science, Publications Office (2017). <https://data.europa.eu/doi/10.2777/337729>.
- 9 Barbero M, Bett K, McLaren J, and Slotin J, Reimagining Data and Power: A roadmap for putting values at the heart of data. Available at <https://www.data4sdgs.org/reimagining-data-and-power-roadmap-putting-values-heart-data> (Accessed on 17 June 2024).
- 10 Government of Canada White Paper: Data Sovereignty and Public Cloud. Available at <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/cloud-services/gc-white-paper-data-sovereignty-public-cloud.html#toc1> (Accessed on 17 June 2024).
- 11 European Commission, European Data Portal, Huyer E, Knippenberg LV, and Arriëns EL. The Economic Impact of Open Data Opportunities for value creation in Europe. The Publications Office of the European Union, 2020. Available at <https://data.europa.eu/sites/default/files/the-economic-impact-of-open-data.pdf> (Accessed on 17 June 2024).
- 12 Do4Africa. Open data in Africa. Available at <https://www.do4africa.org/en/open-data-in-africa/> (accessed on 17 June 2024).
- 13 Data.gov. Mission. Available at <https://data.gov/> (Accessed on 17 June 2024).
- 14 The Statem Council, P.R.C. Measures of managing scientific data. 2018. Available at https://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm (Accessed on 17 June 2024). (*In Chinese*).

- 15 Matsumura H, Hosoda Y, and Furukawa S, Amorphous to Polysilicon growth and “Sunflower Effect” Observed in Catalytic-CVD, *MRS Online Proceedings Library (OPL)*, 336 (1994) 37.
- 16 Moreira A and Savov A, The macroeconomics of shadow banking, *The Journal of Finance*, 72, 6 (2017) 2381-2432.
- 17 Quinn, Thomas P, Stephan Jacobs, Manisha Senadeera, Vuong Le and Coghlan S, The three ghosts of medical AI: Can the black-box present deliver? *Artificial intelligence in medicine*, 124 (2022) 102158.
- 18 CoreTrustSeal Standards and Certification Board, CoreTrustSeal Requirements 2023-2025 (V01.00), *Zenodo*, (2022). <https://doi.org/10.5281/zenodo.7051012>.
- 19 Kubler S, Robert J, Neumaier S, Umbrich J and Le Traon Y, Comparison of metadata quality in open data portals using the Analytic Hierarchy Process, *Government Information Quarterly*, 35 (1) (2018) 13-29.
- 20 Nicolas R G, Jiménez-Contreras E and Daniel Torres-Salinas, Analyzing data citation practices using the data citation index, *Journal of the Association for Information Science and Technology*, 67 (12) (2016) 2964-2975.
- 21 Tang L, FAIR your data, *Nat Methods* 17, 127 (2020). <https://doi.org/10.1038/s41592-020-0742-y>.
- 22 Welter D, Juty N, Rocca-Serra, P *et al*, FAIR in action - a flexible framework to guide FAIRification, *Sci Data* 10, 291 (2023). <https://doi.org/10.1038/s41597-023-02167-2>.
- 23 Krans N A, Ammar A, Nymark P, Willighagen EL, Bakker MI, and Quik JTK, FAIR assessment tools: evaluating use and performance, *NanoImpact*, 27 (2022): 100402. <https://doi.org/10.1016/j.impact.2022.100402>.
- 24 Capgemini Consulting parameters. The Open Data Economy Unlocking Economic Value by Opening Government and Public Data. Available at https://www.capgemini.com/wp-content/uploads/2017/07/the_open_data_economy_unlocking_economic_value_by_opening_government_and_public_data.pdf (Accessed on 17 June 2024).
- 25 Huntington P, Nicholas D and Jamali H R, Website usage metrics: A re-assessment of session data, *Information Processing & Management*, 44 (1) (2008) 358-372.
- 26 Jana S and Chatterjee S, Quantifying Web-site visits using Web statistics: an extended cybermetrics study, *Online Information Review*, 28 (3) (2004) 191-199.
- 27 Obrien P, Arlitsch K, Sterman L, Mixter J, Wheeler J, and Borda S, Undercounting file downloads from institutional repositories, *Journal of Library Administration*, 56 (7) (2016) 854-874.
- 28 Kratz J E and Strasser C, Making data count, *Scientific data*, 2 (1) (2015) 1-5.
- 29 Konkiel S, Assessing the impact and quality of research data using altmetrics and other Indicators, scholarly assessment reports, 2 (1) (2020).
- 30 Zahedi Z and Costas R, General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators, *PloS one*, 13 (5) (2018) e0197326.
- 31 Erikson M G and Erlandson P, A taxonomy of motives to cite, *Social studies of science*, 44 (4) (2014) 625-637.
- 32 Zhang G, Ding Y and Milojević S, Citation content analysis (CCA): A framework for syntactic and semantic analysis of citation content, *Journal of the American Society for Information Science and Technology*, 64 (7) (2013) 1490-1503.
- 33 Buneman, Peter, *et al*. "Data citation and the citation graph." *Quantitative Science Studies* 2 (4) (2021), 1399-1422.
- 34 Small H, Visualizing science by citation mapping, *Journal of the American Society for Information Science* 50 (9) (1999) 799-813.
- 35 Hood A S C and Sutherland W J, The data-index: An author-level metric that values impactful data and incentivizes data sharing, *Ecology and Evolution*, 11 (21) (2021) 14344-14350. DOI: <https://doi.org/10.1002/ece3.8126>.
- 36 Park H, You S and Wolfram D, Informal data citation for data sharing and reuse is more common than formal data citation in biomedical fields, *Journal of the Association of Information Science and Technology*, 69 (11) (2018) 1346-1354. DOI:10.1002/asi.24049.
- 37 Sadowski J, When data is capital: Datafication, accumulation, and extraction, *Big Data & Society*, 6 (1) (2019). <https://doi.org/10.1177/2053951718820549>.
- 38 Yang J, Zhao C and Xing C, Big data market optimization pricing model based on data quality, *Complexity*, 1 (2019) 5964068.
- 39 Johnson J A, From open data to information justice, *Ethics and Information Technology*, 16 (2014), 263-274.
- 40 FAIR Data Maturity Model Working Group, FAIR Data Maturity Model. Specification and Guidelines (1.0) (2020). <https://doi.org/10.15497/rda00050>.
- 41 Devaraju A, Huber R, Mokrane M, Herterich P, Cepinskas L, de Vries J , L'Hours, H, Davidson J and White A, FAIRsFAIR Data Object Assessment Metrics (0.4), *Zenodo* (2020). <https://doi.org/10.5281/zenodo.4081213>.
- 42 Moraga C, Moraga M Á, Calero C and Caro A, S QuARE-aligned data quality model for web portals. In *2009 Ninth International Conference on Quality Software*, (2009) 117-122. IEEE.
- 43 Çaldağ M T and Gökalp E, Developing a reference model for open data capability maturity assessment, In *evolving software processes: trends and future directions* (2022) 135-150. <https://doi.org/10.1002/9781119821779.ch6>.
- 44 Ingwersen P and Chavan V, Indicators for the Data Usage Index (DUI): an incentive for publishing primary biodiversity data through global information infrastructure, *BMC Bioinformatics*, 12 Suppl 15 (2011) S3. 10.1186/1471-2105-12-S15-S3.
- 45 Begany G M and Martin E G, An open health data engagement ecosystem model: Are facilitators the key to open data success? In S. Erdelez & N.K. Agarwal (Eds.), *Proceedings of the Association for Information Science and Technology*, Hoboken, NJ: Wiley, (2017) 621–623. <https://doi.org/10.1002/prae.2017.14505401091>.
- 46 Benitez-Paez F, Comber A, Trilles S and Huerta J, Creating a conceptual framework to improve the re-usability of open geographic data in cities, *Transactions in GIS*, 22 (2018) 806–822. <https://doi.org/10.1111/tgis.12449>.
- 47 data.europa.eu. Measuring open data maturity (ninth edition). 2023. Available at https://data.europa.eu/sites/default/files/odm2023_method_paper.pdf (Accessed on 17 June 2024).
- 48 Abella A, Ortiz-de-Urbina-Criado M and De-Pablos-Heredero C, Meloda 5: A metric to assess open data reusability, *Information Professional*, 28 (6) (2019).
- 49 Hackman L, Mack P and Ménard H, Behind every good research there are data. What are they and their importance to forensic science, *Forensic Science International: Synergy*, (2024) 100456. <https://doi.org/10.1016/j.fsism.2024.100456>.
- 50 Nature scientific data. Submission Guidelines. Available at <https://www.nature.com/sdata/publish/submission-guidelines> (Accessed on 18 June 2024).
- 51 Rafiq K, Jordan N R, McNutt J W, Neelo J, Attias N, Boersma D and Abrahms B, The value of field research in academia, *Science*, 384 (6698) (2024) 855-856.