



## Policy Brief: Data ethics and structural inequities in science

CODATA data ethics working group

### Summary and recommendations

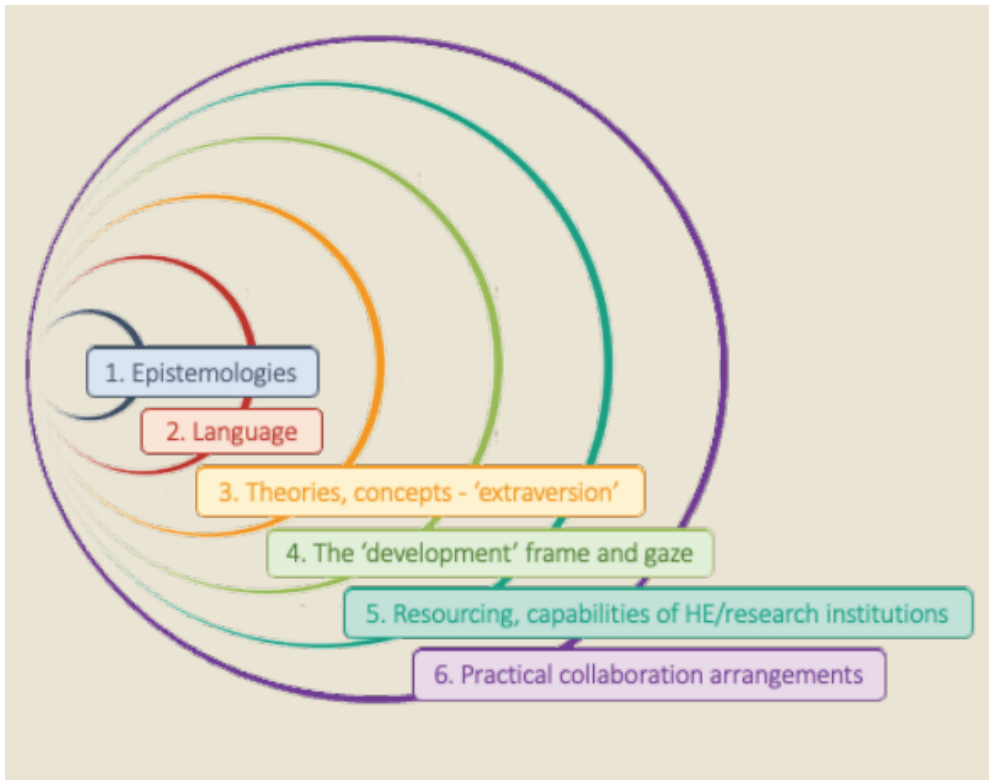
The gap that we identify in *The UNESCO Recommendation on Open Science* is in acknowledging the systematic structural conditions creating inequitable participation in science, and the impact that this has on how a push towards open science might play out in practice.

Science as a global system is riven by inequities. This has five interconnected dimensions:

1. Identity-based inequities shaping participation in science at an individual level
2. International inequities in the strength, visibility and recognition of research systems
3. Inequities in the research infrastructure and access to funding
4. Inequitable access to an increasingly commercialised publishing system
5. Data colonialism

These issues are well known, for example many points about international patterns in inequity are reiterated in previous policy statements, including the recent Africa Charter for Transformative Research Collaborations.

Figure: Schema Uneven playing field in Africa (Africa Charter for Transformative Research Collaborations, 2023).



As well as being unjust these inequities compromise diversity and participation in science and so its strength and richness.

We appreciate that for many open science is seen as a way to address some of these issues, but we argue that there are significant power asymmetries involved and open science as a solution is not necessarily sufficient or even effective.

Given these conditions, the impact of open science and the sharing of data in particular, may play out in ways that are also inequitable. The principle of “intelligent openness”, where the definition of open science is based on context, needs much more development and backing with robust international governance mechanisms and examples of good practice.

Across the different dimensions of inequity, there is a need for:

- Greater explicit acknowledgement of the issues
- More research into the nature and distribution of the barriers to participation in science
- Support to local definitions of the meaning of openness
- Benchmarks for improvement to be established and progress monitored
- All stakeholders to educate themselves about the issues
- More participatory approaches to policy development

## 1. Introduction

The question of how the UNESCO Recommendation is implemented to realise open and equitable OS in practice must account for the structural conditions shaping research at a national and individual level. Scholars in many national contexts face barriers such as lack of

basic infrastructure, unsupportive national policy, problematic incentive structures, the control of the research agenda by funders in high-income countries and the domination of oligopolistic publishers and Big Tech companies. At an individual level, researchers everywhere who do not fit the expected norm of a scholar (white, able bodied, male) face multiple barriers such as conscious and unconscious bias, racism, misogyny, career breaks and societal expectations about caring responsibilities.

The aim of this policy brief is to initiate dialogue on the question of how the UNESCO Recommendation is being implemented to realise open and equitable OS in practice in the context of the structural conditions shaping research at a national and individual level.

## **2. Science: A system riven by inequity**

We see these conditions of inequity as having five interconnected dimensions, set out in the following sections:

### *Identity-based inequities*

Firstly, we point to the inequities in participation in science that occur in almost all research systems at an individual level, based on researcher identity, particularly race, gender, sexuality, dis/ability, class and the intersections between these identities. Researchers everywhere who do not fit the expected norm of a scholar (white, able bodied, male, english speaking) face multiple barriers.

For example, women face particular challenges in building a research career, arising from

- A. Biased assumptions about their capabilities, affecting them throughout the educational system and the pipeline into academia, in selection for academic roles, in promotion, in pay, in research funding, in networking and mentoring, and in peer review and citation, based on conscious and unconscious bias, and at worst on misogyny and other prejudice
- B. Societal expectations and pressures, such as to privilege caring roles over careers, and the failure to create policies that account for career breaks

Black women are often effectively excluded from entering research because their families lack resources to support their education, because they lack knowledge of the educational system and role models in it, and may feel alienated from research institutions, whose history is of privilege and exploitative research (Gilbney, 2022). They may experience isolation and prejudice (Sian, 2017). Even where they are incorporated into scholarly institutions, this may not be accompanied by recognition for their research and citation (Smith and Garrett-Scott, 2021).

While no means universal across all disciplines in all countries, these problems exist globally.

Specifically this creates many inequities around data for scholars with these identities, such as failure to gain funding to create data or manage the deposit process, lack of time and

resources to reuse shared data, potential for data to be appropriated by senior colleagues due to power and status differentials, lack of citation of their data (as well as their outputs; Ross-Hellauer et al 2022).

### **3. International inequities in research systems**

Secondly, we point to the gross inequities in research systems internationally that make participation in science much harder for those in less privileged contexts, where the level of government support in terms of funding and policy is less strong.

Government commitment in policy and funding of research at all levels is much lower in some countries, such as sub-Saharan Africa, than in high-income countries in North America, the European Union etc. Often the government support that there is may be modeled on Western examples and reproduces their assumptions, even though they may not be appropriate to the context. They lag in critical areas such as in supporting DORA.

Supplementing such public funding are funding from charities, non-for-profit organisations such as the Gates Foundation and the Wellcome Trust, as well as commercial and military organisations. All such funding is largely allocated from well-resourced countries and locations, mostly in the high-income countries. The massive power of funding by high-income countries, tends to lead them to dominate the agenda of research, even in low-income countries. Certain types of research (most blatantly in the humanities, but also less visible parts of natural science such as botany or physiology) tends to gain limited funding leading to researchers in these fields producing fewer outputs and much smaller datasets.

Some countries and organisations use science funding as a form of soft influence as part of their policy further undermining locally driven research agendas.

In a related way, some research locations enjoy a great advantage over others by functioning as role models for what best practice is expected to look like within given domains. This is not always the same thing as scientific excellence: in many cases, good research carried out in less reputable locations and domains (not to speak of local and indigenous knowledge of relevance to discovery) gets overlooked in favor of the topics, approaches and contexts that are viewed as most topical and promising (Leonelli 2023).

### **4. Inequities in the research infrastructure**

Many countries, especially low income countries, have limited access to the infrastructure needed to conduct research such as intermittent electricity supplies, poor technical infrastructure (such as internet connectivity and technical support), sub-standard research training, lack of expert data management support, lack of access to the published literature, and to the latest research software (Bezuidenhout et al., 2017). Low level connectivity issues effectively limit access even where material is ostensibly open (Shanahan & Bezuidenhout, 2022). Researchers are also time poor, because of the pressure to teach and limited administrative support. As a result of these issues, researchers may lack confidence in the quality of their work and trust in international systems of communication and research

assessment, which typically do not account for such obstacles and the inequity in resources available to different research groups (Leonelli 2018).

This impacts data sharing in particular, by limiting the funding to collect and share data, the quality of the infrastructure to support data sharing, quality of training in data related skills and places researchers on the back foot in research collaborations in relation to researchers located in more privileged contexts.

## **5. Inequity of access to a commercialized publishing system**

The publishing system is increasingly oligopolistic and has a deeply entrenched bias towards western epistemologies and the English language.

Thus scholars in low-income countries are often under pressure to publish in "prestigious" international journals. That is to say ones that have a high impact or high level of citation within this system, further reinforcing its continuing domination.

Being successful in publishing research within this system involves researchers being forced to do one or all of the following:

- A. Conform to western epistemologies (including research themes, reference points in the literature and preferred research methods),
- B. Gain approval from editors and peer reviewers who are biased and/or ignorant of the author's context, and disconnected from the peer networks most likely to be familiar with the author's approach and methods
- C. Read and write in a language, English, a language that is not their native tongue, and
- D. Pay article publishing costs (APC) which are very high - indeed, often unfeasible - in the context of limited financial resources. APC waivers help where they are offered but are available in a limited number of contexts.

Even where some of these inequities do not exist the inhospitable nature of the publishing system pushes researchers to work outside it, yet, publishing outside this system can lead to bibliographic invisibility, although it is true that preprint publishing may offer some benefit.

Merely placing open scientific practices over the top of this system does little to dismantle the systematic inequities embedded within it. Open science, as currently implemented chiefly through infrastructures and training resources in English, is reinforcing the bias towards English language publishing.

Ironically, attempts in high-income countries to reach transformative agreements with publishers, often strengthen the hand of the scholars from those countries in controlling publication - since many countries in low-income countries do not have capacity to enter such agreements in the first place.

## **6. Data colonialism**

Increasingly research data is also being drawn into this oligopolistic system as a commodity. This process underlies increasingly extractive data practices. It is part of a wider system of

digital colonialism where big Tech companies based in high-income countries gain control over digital infrastructure to dominate nations politically, economically and socially.

Data and algorithmic colonialism are one key dimension of digital colonialism (Brihane, 2020; Abebe et al., 2021; Coleman, 2019). Data is extracted by large corporations who seek to control the digital infrastructure in regions such as Africa, use it without contextual understanding, blocking the development of local solutions. In this context the beneficiaries of data sharing are rarely data subjects or local communities.

Other powers may also have strategies of digital colonialism (Gravett, 2020). Moreover, the well-meaning donation of research infrastructures, technologies or training materials from high-income countries to low-and-middle-income countries may itself be problematic, whenever it happens without adequate consultation with recipients around their own needs and concerns.

## **7. Interconnected inequities**

We do not see these layers of inequity as separate, rather they are all inter-related and ultimately derived from a world order rooted in capitalism, patriarchy and coloniality. As such this state of affairs cannot be “fixed” but UNESCO should be supporting a long-term process by which these inequities are challenged and addressed. This will help to address historic epistemic violence and promote diversity in science as a source of strength and richness.

Thus, for example, many indigenous communities are under-represented in multiple ways, such as by the number of researchers from this background in the system, lack of role models, lack of funding, lack of acknowledgement of their epistemologies and corresponding methodologies for knowledge building. The history of science and research institutions with their links to colonialism alienates people from indigenous communities from participation in research careers. The references to the CARE principles and the value of indigenous knowledge in the UNESCO recommendation is to be welcomed, but does not go far enough in acknowledging the nature of the underlying problems. The emphasis on indigenous knowledge is all too restrictive, since many forms of local knowledge (including from farmers, patient groups and others with relevant expertise to research domains) tend to be systematically excluded from open science initiatives.

This is not a simple binary division where scholars in the “global north” are included and other scholars excluded. Rather we see a spectrum of issues, from very privileged contexts such as in some disciplines in the US, through to gross marginalisation in some countries.

While for some open science has always been about challenging these issues and while the Recommendation contains many progressive elements such as the emphasis on “dialogue with other knowledge systems”, inclusion of knowledge from traditionally marginalised scholars, the CARE principles, and the stress on “equity and fairness” and “diversity and inclusiveness”, we argue that the entrenched power structures that underlie current inequities are not sufficiently acknowledged. The first step is to acknowledge this wider context for open science initiatives.

We also draw attention to the vicious cycles undermining trust in open science. The inequities in support, visibility and capacities generate a vicious cycle whereby researchers who are less able to participate in open science have less and less incentives to do so – for instance, because they deem their data not to be of a good enough quality to be shared, because they do not gain benefit from using others' data, because they fear their data that could be misinterpreted by those who do not understand their context, and because they are troubled by the commercial interests who benefit from shared data.

The meaning of intelligent openness, where it is defined locally, must be fully developed and realised. While we identify many inequities, we do not want to promote a view of research in low-income countries through a deficit perspective (Abbott and Cox, 2020), rather developing local understandings of OS should be seen as a priority.

Lauded principles such as FAIR place too much emphasis on technical aspects of data sharing, while appropriating the word FAIR without due acknowledgements of ethical, social and justice issues.

## 8. Recommendations

- ❖ While open science has many potential benefits it has to be re-assessed as a goal in the context of the gross structural inequities that we identify. Explicit acknowledgement of these issues should be made to frame the long term goals of policy in science.
- ❖ The principle of “intelligent openness”, where the definition of Open Science is based on context, needs much more development and backing with robust international governance mechanisms and examples of good practice. Simply calling for open data is not a solution. Similarly, as the roll-out of Open Science increases, attention must be paid to how “meaningful connectivity”, in place of a “digital divide”, is understood so as to ensure that all users are not only connected but able to transact online as they see fit.
- ❖ Key stakeholders (funders, journal publishers, editors, peer reviewers, all researchers) need to educate themselves about the issues.
- ❖ Actions to combat the five dimensions of inequity should be captured and shared as good practice.
- ❖ Actions to address individual level inequity
  - Research institutions and funders to take proactive action to narrow inequities, including identifying role models, monitoring outcomes more closely and publishing evidence openly
  - Training in conscious and unconscious bias for stakeholders across the system of science, from employers to peer reviewers
- ❖ Actions to address inequitable research systems
  - Open science policy should acknowledge issues discussed here, creating an environment in which those with less resources can gain equal benefit from open science and become meaningful participants in the open science movement.
  - International funders should weigh up the impact of their funding decisions

- ❖ Actions to address inequities in research infrastructure
  - Further research into the nature of the impact of inequities of infrastructures
  - Actions by policy makers to create an equal playing field
- ❖ Actions to address the growing power of the commercialized publishing system
  - Support regulation of commercial pricing over publishing
  - Support Diamond Open Access solutions, where neither authors nor readers pay
  - Encourage commercial publishers to work together with academic institutions and scholarly societies towards a fair, sustainable publishing system
- ❖ Actions to prevent data colonialism
  - Stronger participatory international data governance mechanisms
  - Closer scrutiny of extractive data practices of publishers and other commercial companies
  - Foster bottom-up, community based data sovereignty
- ❖ Across the different dimensions of inequity, there is a need for:
  - Greater explicit acknowledgement of the issues with a fundamental shift of the agenda
  - More research into the nature and distribution of the barriers to participation in science
  - Benchmarks for improvement to be established and progress monitored
  - Support to local definitions of the meaning of openness
  - All stakeholders to educate themselves about the issues
  - More participatory approaches to policy development

## Acknowledgements

The development and coordination of this thematic policy brief was led by Suchith Anand, Louise Bezuidenhout, Andrew Cox, Johannes John-Langba and Sabina Leonelli ( CODATA Data Ethics Working Group). We acknowledge the invaluable feedback on the document by the other members of the CODATA Data Ethics Working Group.

## References

- Abebe, R., Aruleba, K., Birhane, A., Kingsley, S., Obaido, G., Remy, S. L., & Sadagopan, S. (2021). Narratives and counternarratives on data sharing in Africa. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 329-341).
- Abbott, P., & Cox, A. (2020). Perceptions of Rwanda's Research Environment in the Context of Digitalization: Reflections on Deficit Discourses. In R. K. Bandi, R. C. R., S. Klein, S. Madon, & E. Monteiro (Eds.), *The Future of Digital Work: The Challenge of Inequality* (pp. 50–64). Springer International Publishing. [https://doi.org/10.1007/978-3-030-64697-4\\_6](https://doi.org/10.1007/978-3-030-64697-4_6)
- Bezuidenhout, L. M., Leonelli, S., Kelly, A. H., & Rappert, B. (2017). Beyond the digital divide: Towards a situated approach to open data. *Science and Public Policy*, 44(4), 464-475.



Birhane, A. (2020). Algorithmic colonization of Africa. *SCRIPTed*, 17, 389.

Coleman, D. (2018). Digital colonialism: The 21st century scramble for Africa through the extraction and control of user data and the limitations of data protection laws. *Michigan Journal of Race & Law*, 24, 417.

Gilbney, E. (2022). How UK science is failing on diversity. *Nature*, 390-5.

Gravett, W. H. (2020). Digital coloniser? China and artificial intelligence in Africa. *Survival*, 62(6), 153-178.

Leonelli, S. (2017) Global Data Quality Assessment and the Situated Nature of “Best” Research Practices in Biology. *Data Science Journal* 16(32): 1-11. DOI: 10.5334/dsj-2017-032

Leonelli, S. (2023) *Philosophy of Open Science*. Cambridge University Press.

Ross-Hellauer, T., Reichmann, S., Cole, N. L., Fessler, A., Klebel, T., & Pontika, N. (2022). Dynamics of cumulative advantage and threats to equity in open science: a scoping review. *Royal Society Open Science*, 9(1), 211032.

Shanahan, H., & Bezuidenhout, L. (2022). Rethinking the a in FAIR Data: issues of data access and accessibility in research. *Frontiers in Research Metrics and Analytics*, 7, 912456.

Sian, K. (2017). Being black in a white world: Understanding racism in British universities. *Papeles de Cheic*, 2.

Smith, C. A., & Garrett-Scott, D. (2021). “We are not named”: Black women and the politics of citation in anthropology. *Feminist Anthropology*, 2(1), 18-37.