Research Data Management & publishing of datasets

Global trends with examples from The Netherlands

Kim Ferguson, DANS-KNAW

October 25th, 2023

Wits Open Research Series

CC BY

10.5281/zenodo.10040123

Overview

- *Introduction*
- *Open Science & research data*
- *Open Data in the Netherlands*
- *Where to find & deposit data*
- *Final look at trends*

Data Archiving and Networked Services

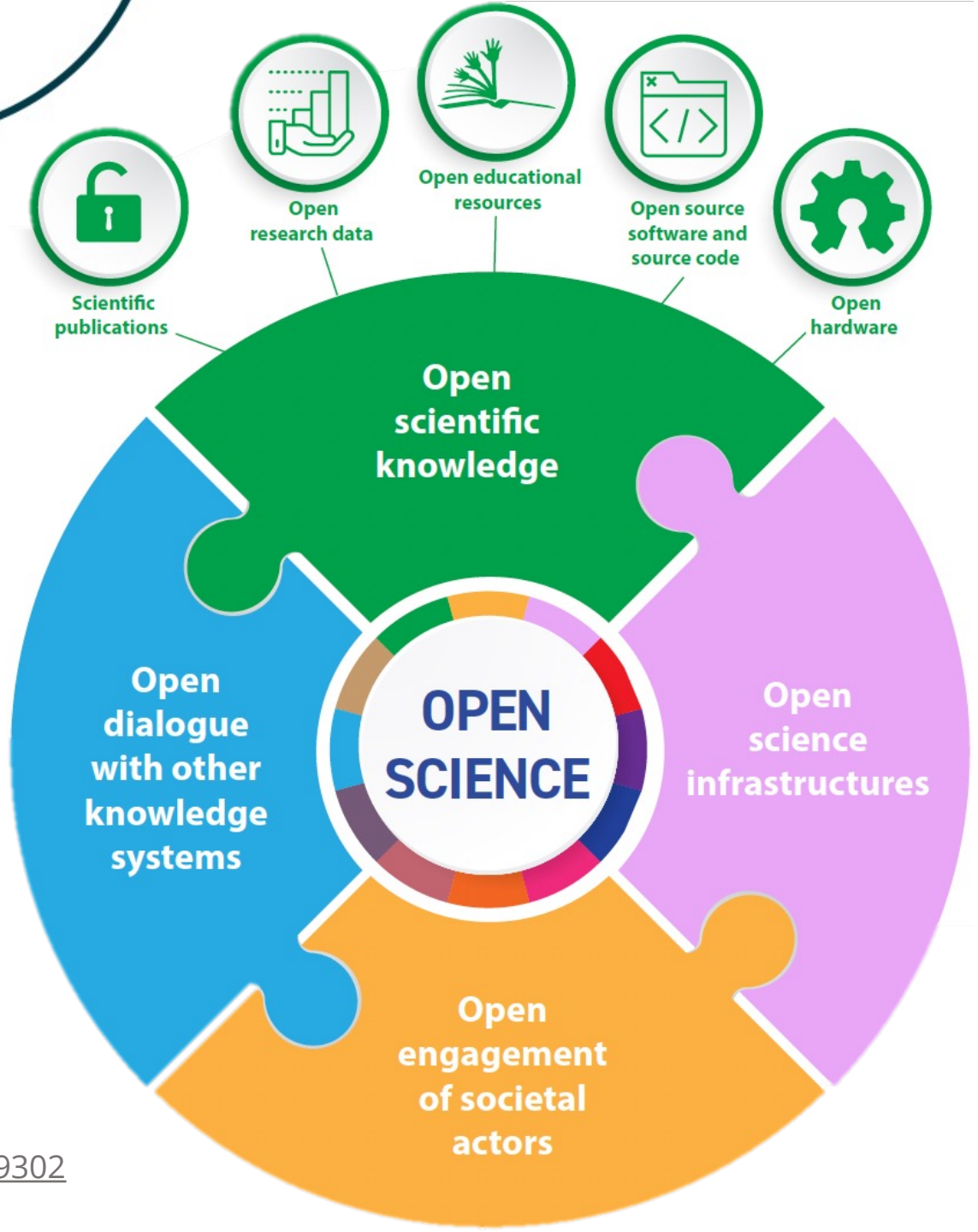
DANS

DANS is the Dutch national centre of expertise and repository for research data, based in the Hague.
We help researchers make their data available for reuse.

Versatile Repositories

Centre of Expertise

Active Collaborator



Research data:
"digital or analogue, raw or processed, metadata, records, images, protocols, analysis codes, workflows, and more"

Source & credit:
https://doi.org/10.54677/UTCD9302
UNESCO, CC BY-SA 3.0 IGO

Why make data open?

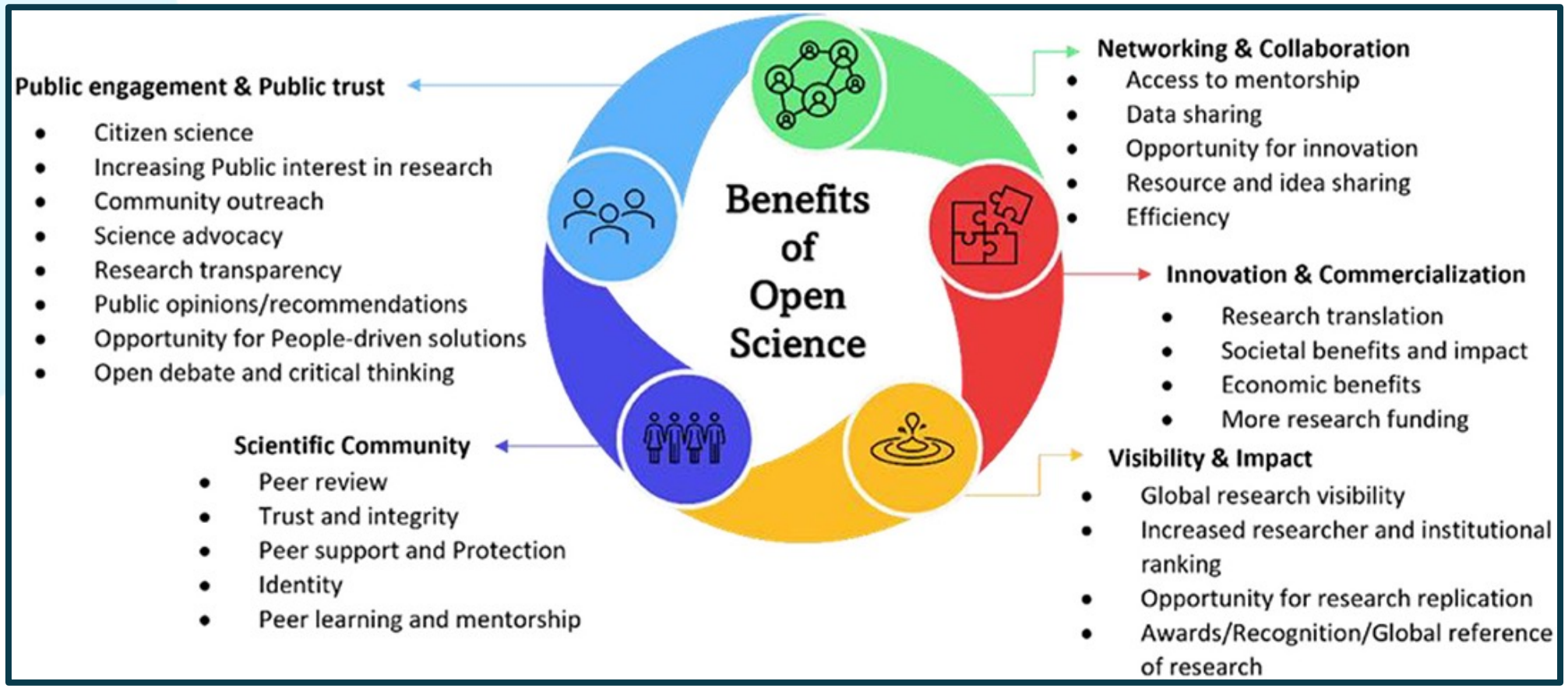


Fig 2, from Okafor *et al.* 2022, *Institutionalizing Open Science in Africa: Limitations and Prospects*

+ One more reason: Funding requirement (In Europe: EU and national funding)

Elements to support open data – researcher level (i)

- Fulfilling commitments to science & society
- Papers with open data tend to receive more citations and re-use (Colavizza *et al.* 2020), especially when it is linked to the repository directly

Elements to support open data – researcher level (ii)

- Papers with published **code** also tend to receive more citations and re-use, though it is a less common practice (Maitner *et al.* 2023)

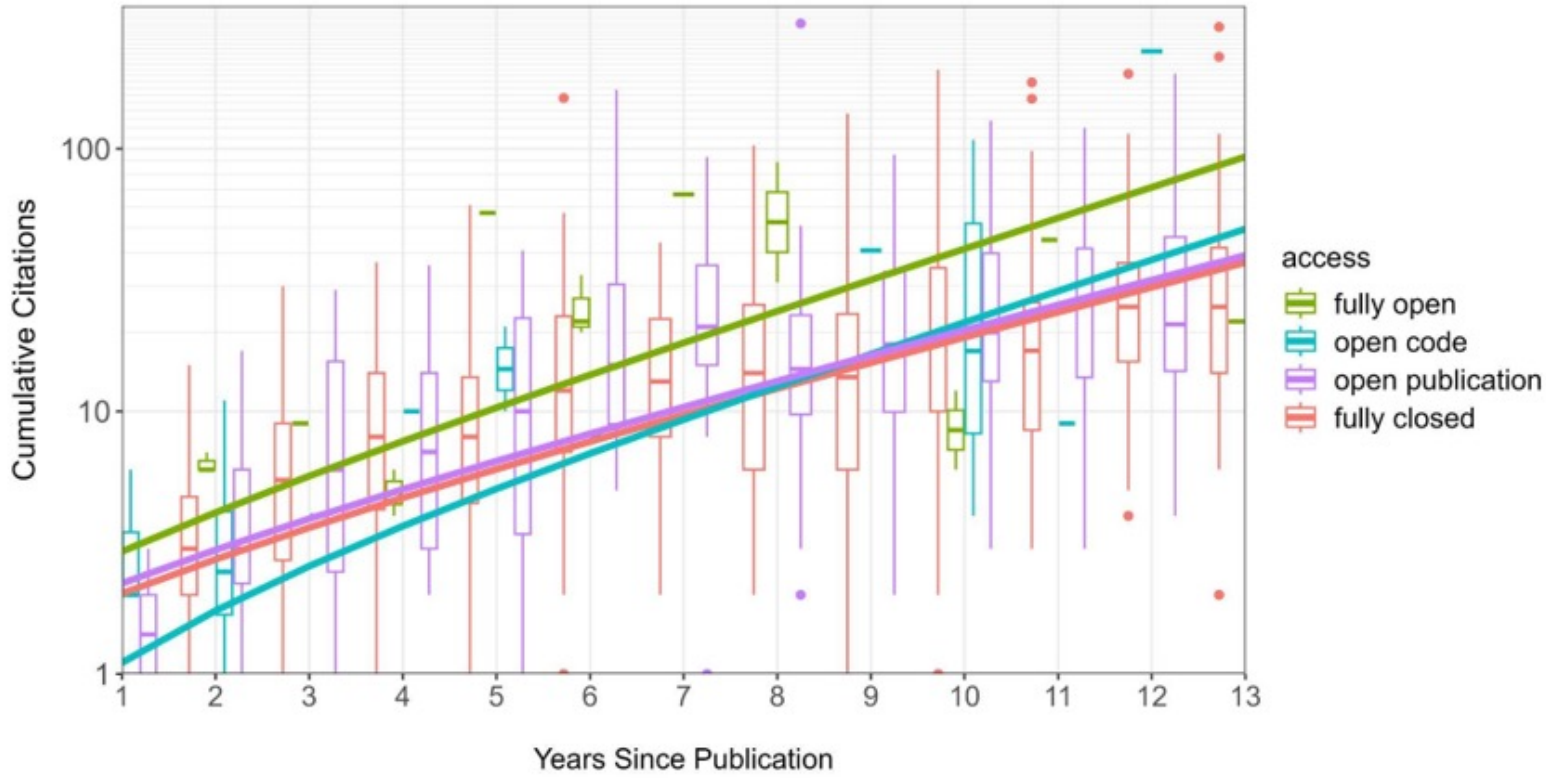


Figure 2 from Maitner et al. 2023, https://doi.org/10.21203/rs.3.rs-3222221/v1 under CC by 4.0

Elements to support open data – researcher level (iii)

Increasing dialogue on rewarding open science and open data practices for researchers (CoARA, OpenScienceNL, etc.)

This is part of a national dialogue, **Recognition & Rewards**, supported by the universities, research institutes, and funders

From 'Room for Everyone's Talent', https://doi.org/10.5281/zenodo.5762094 under CC by 3.0

In NL, there are both ground-up and top-down approaches to increasing open data in practice

Ground-up:

- Offering training to support staff in libraries and digital competency centers – enable them to provide support to researchers
- Open Science Communities at most major university or research centers – researchers, librarians, students, managers, etc.

Top-down:

- Open Science NL – national body that is funded by government and steered by multiple organisations and stakeholders, policy body
- Dutch Data Prize – given out every two years at the Open Science Festival
- National funders requiring open access for publications – but what about open data? Not currently required, “in progress”

How does DANS encourage publishing datasets?

Various archiving services for Dutch researchers (and international projects):

- DANS Data Stations (domains-specific)
  - Archaeology (**140,615** datasets)
  - Social Science & Humanities (**7,196** datasets)
  - Life, Health, & Medical Sciences (**28,706** datasets) → will be launched in November
  - Physical & Technical Sciences (**88** datasets) → will be launched in November
- DataverseNL (managing technical infrastructure of open source repository software for Dutch universities)
  - 125 organisations (multiple per university)
  - **6,697** datasets
- DANS Data Vault (long-term storage, including what is in the Data Stations and Dataverse NL, as well as our previous archive, EASY)

*All counts of datasets are as of October 24, 2023

How else does DANS support publishing datasets

- SSH Data Station: online Open Hours session every Monday
- Supporting projects like the CESSDA Data Management Expert Guide (DMEG) – resources that researchers can use that DANS contributes to, but doesn't need to maintain alone
- Leveraging external projects and initiatives like EOSC and RDA
- Proving training on FAIR data and open science to data stewards and librarians through platforms like Research Data Netherlands (RDNL)

Where to deposit or find data?

DANS Data Stations are great, but what else?

- Re3Data → Registry of Research Data Repositories
  - Can browse by content type, subject, or country
  - Filter according to access policies and licence type
  - Filter according to language (not the widest range)
  - Filter according to certification of the repository

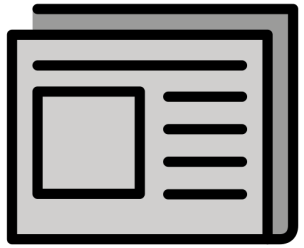
However, registries have gaps and bias
→ librarian support is still key

Multiple possibilities to share data throughout research trajectory

Sole focus on publication (the “final product” of research) overlooks opportunities for researchers

→ presentations, posters, datasets – these are all outputs to disseminate

My own open data adventure: 2015-2020

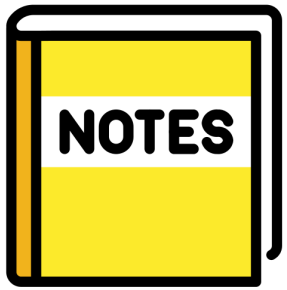


Publications:
5 in peer-reviewed journals, open access
(4 were preprints first)
1 preprint, open access
(2019-2020)

Additional Datasets

14 in figshare,
2 in DANS EASY
(2018-2020)

Code

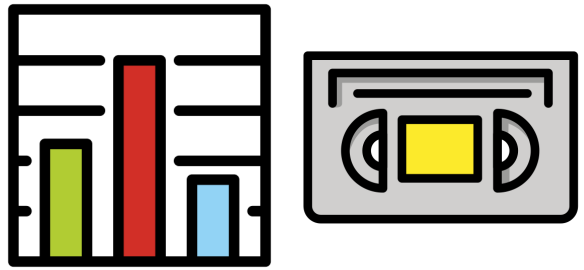


Supplementary materials in articles

Posters

1 in figshare
(2018)

Presentations & recordings



8 in figshare
(2017-2020)

Extensive work on both uptake of and barriers to Open Science/Data in the African region

- *African researchers do not think differently about Open Data*, Skelly & Chiware 2022
- *Rethinking the A in FAIR data: issues of data access and accessibility in research*, Shanahan & Bezuidenhout 2022
- *Institutionalizing Open Science in Africa: Limitations and Prospects*, Okafor *et al.* 2022

- There are also hopeful policy initiatives, but time will tell if the uptake is not a burden on researchers and support staff

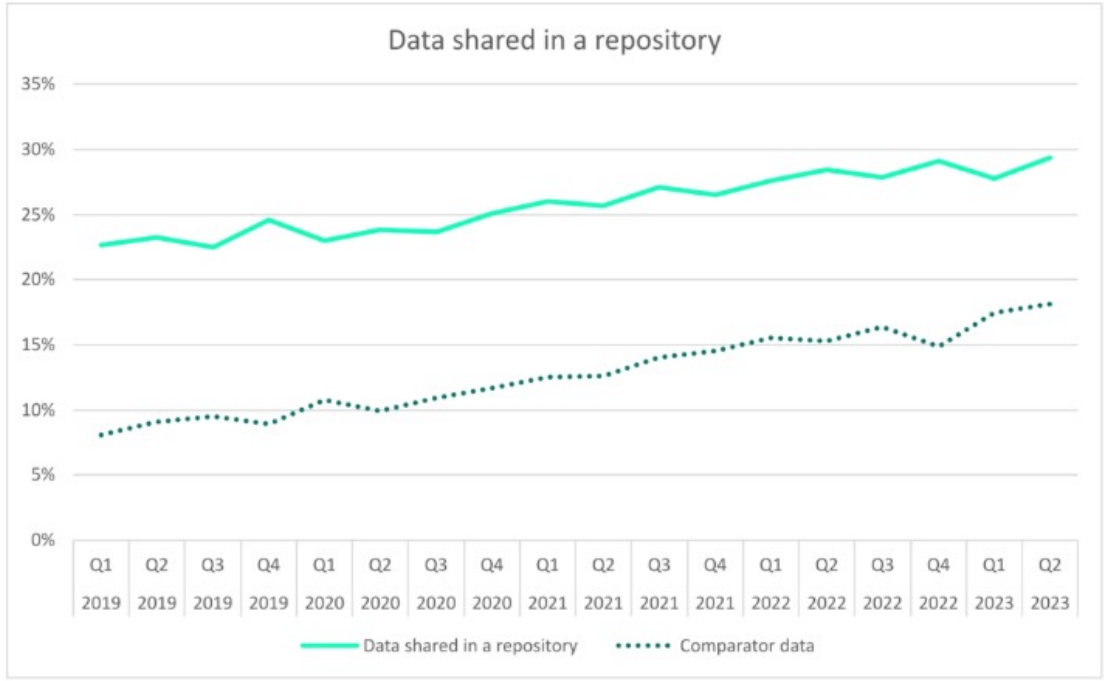
Are there parallels with the Netherlands when it comes to Open Data practices?

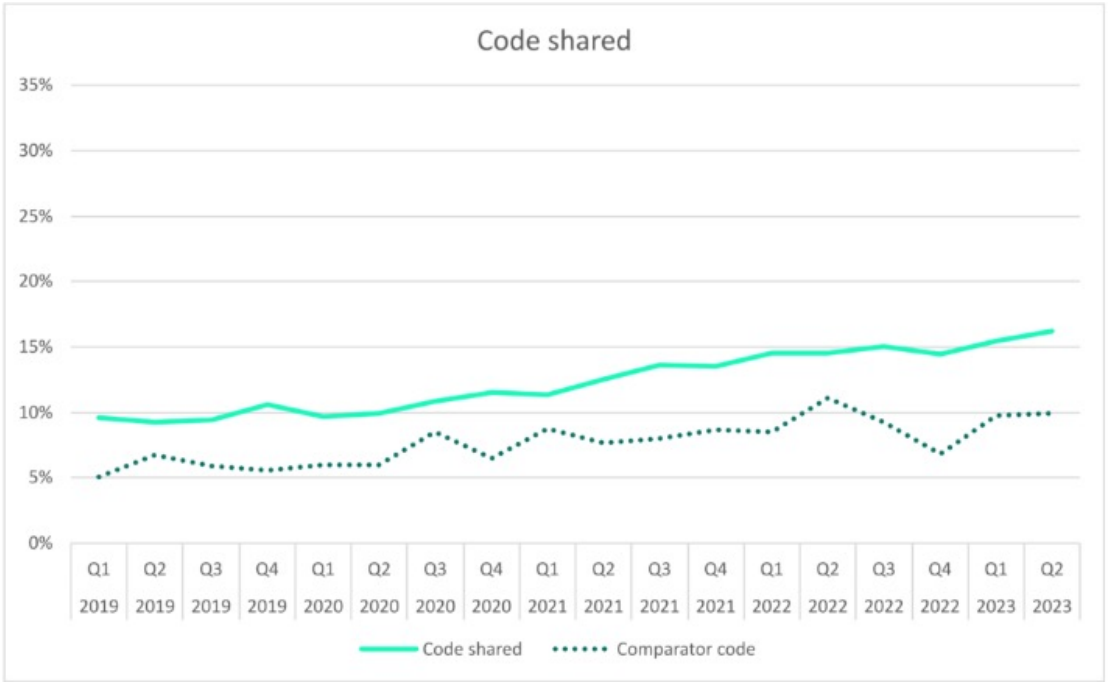
- Dutch Research Council, main funder in the Netherlands, polled its researchers in 2021 about Open Science (over 1000 respondents).
- Key areas of concern:
  - Cost (both material and time)
  - Need more infrastructure
  - Guidelines are unclear or changing
  - No recognition for Open Science efforts
  - Larger divide between disciplines
    - Humanities & social sciences respondents (45%) making their data open less often than natural and life sciences respondents (65%)

Source: https://www.nwo.nl/en/only-way-forward

Final look at trends

- Open Science Indicators from PLOS
- Current dataset: PLOS articles from Jan 1 2019 to June 30 2023, **over 78,000 articles**
  - plus a comparator set from PubMedCentral, **over 16,000 articles**
- Not a domain-wide metric, but something to think on
- Generally, data sharing via repository and code sharing is on the rise, but not equally across domains





Source: *The Official PLOS Blog*. 3 October 2023. https://theplosblog.plos.org/2023/10/open-science-indicators-q2-2023/.

Thinking back to publishing of code,

When the distinction between publications and datasets becomes more fluid, this would also include code sharing and other elements of research data

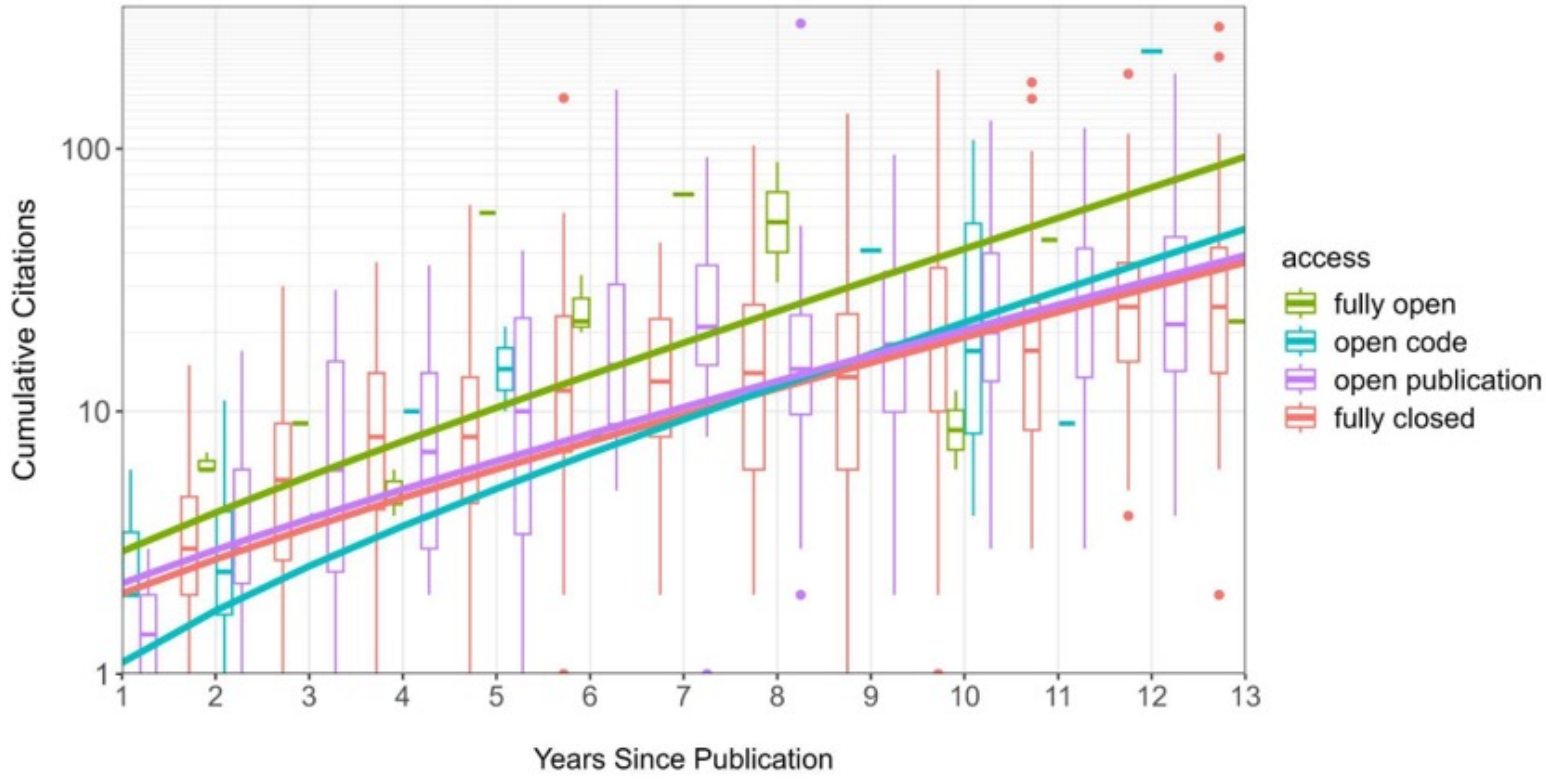


Figure 2 from Maitner et al. 2023, https://doi.org/10.21203/rs.3.rs-3222221/v1 under CC by 4.0

More information

Visit our website **www.dans.knaw.nl**

And follow us online

Mastodon
@DANS_knaw_nwo

LinkedIn
@DANS

X
@DANS_knaw_nwo

DANS is an institute of KNAW and NWO
Dutch national centre of expertise and repository for research data

References

Cadwallader, Lauren, Lindsay Morton, and Iain Hrynaszkiewicz. 2023. 'A First Look at Open Science Indicators for Articles Published in 2023'. *The Official PLOS Blog* (blog). 26 June 2023. https://theplosblog.plos.org/2023/06/open-science-indicators-update-q1-2023/.

Cadwallader, Lauren, Lindsay Morton, and Iain Hrynaszkiewicz. 2023. 'A New Open Science Indicators Dataset Is Here!' *The Official PLOS Blog* (blog). 3 October 2023. https://theplosblog.plos.org/2023/10/open-science-indicators-q2-2023/.

Colavizza, Giovanni, Iain Hrynaszkiewicz, Isla Staden, Kirstie Whitaker, and Barbara McGillivray. 2020. 'The Citation Advantage of Linking Publications to Research Data'. *PLOS ONE* 15 (4): e0230416. https://doi.org/10.1371/journal.pone.0230416.

Maitner, Brian, Paul Santos-Andrade, Luna Lei, George Barbosa, Brad Boyle, Matiss Castorena, Brian Enquist, et al. 2023. 'Code Sharing Increases Citations, but Remains Uncommon'. Research Square. https://doi.org/10.21203/rs.3.rs-3222221/v1. [preprint]

Okafor, Izuchukwu Azuka, Smart Ikechukwu Mbagwu, Terkuma Chia, Zuwati Hasim, Echezona Ejike Udokanma, and Karthik Chandran. 2022. 'Institutionalizing Open Science in Africa: Limitations and Prospects'. *Frontiers in Research Metrics and Analytics* 7. https://www.frontiersin.org/articles/10.3389/frma.2022.855198.

'Open Science 2030 in the Netherlands: NPOS2030 Ambition Document and Rolling Agenda'. n.d. Accessed 24 October 2023. https://doi.org/10.5281/zenodo.7433767.

Shanahan, Hugh, and Louise Bezuidenhout. 2022. 'Rethinking the A in FAIR Data: Issues of Data Access and Accessibility in Research'. *Frontiers in Research Metrics and Analytics* 7. https://www.frontiersin.org/articles/10.3389/frma.2022.912456.

Skelly, Lara, and Elisha R. T. Chiware. 2022. 'African Researchers Do Not Think Differently about Open Data'. *Frontiers in Research Metrics and Analytics* 7. https://www.frontiersin.org/articles/10.3389/frma.2022.950212.

'The Only Way Forward | NWO'. n.d. Accessed 24 October 2023. https://www.nwo.nl/en/only-way-forward.

UNESCO. 2022. 'Understanding Open Science'. https://doi.org/10.54677/UTCD9302.

VSNU, NFU. 2019. 'Room for Everyone's Talent: Towards a New Balance in the Recognition and Rewards for Academics', November. https://doi.org/10.5281/zenodo.5762094.

All emojis on slide 14 are designed by OpenMoji – the open-source emoji and icon project. License: CC BY-SA 4.0