



UvA-DARE (Digital Academic Repository)

Infrastructure, Intermediaries, and Artificial Intelligence: A Rejoinder to Commentaries on “From Data Creator to Data Reuser: Distance Matters”

Borgman, Christine L.; Groth, Paul

DOI

[10.1162/99608f92.c17c3adb](https://doi.org/10.1162/99608f92.c17c3adb)

Publication date

2025

Document Version

Other version

Published in

Harvard Data Science Review

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Borgman, C. L., & Groth, P. (2025). Infrastructure, Intermediaries, and Artificial Intelligence: A Rejoinder to Commentaries on “From Data Creator to Data Reuser: Distance Matters”. *Harvard Data Science Review*, 7(2). <https://doi.org/10.1162/99608f92.c17c3adb>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)



Search

Dashboard ▾

Login or Signup

Issue

Co

Published on Apr 30, 2025

DOI 10.1162/99608f92.c17c3adb

SHOW DETAILS ↕

↳ This Pub is a Rejoinder to — 1 of 10 parents ⏪ ⏩



How Platforms Constrain Matters of Distance and Data Reuse

by *Amelia Acker*

Published on Apr 30, 2025 ·

We are deeply honored by the insightful contributions of the distinguished colleagues who reflected upon our article, “From Data Creator to Data Reuser: Distance Matters” **(Borgman & Groth, 2025)**. Rather than attempting to synthesize the full intellectual content of these commentaries, we draw out common themes, offering a roadmap to reading them in conjunction with our article, as a forum on how ‘distance matters’ in creating and reusing research data. The dozen colleagues—all expert in aspects of data sharing and reuse—who wrote these nine thoughtful commentaries hail from three continents and from disciplines as diverse as archives, astronomy, information science, data science, philosophy, physics, communication, classics, computer science, and social studies of science **(Acker, 2025; boyd & Sarathy, 2025; Crane et al., 2025; Goodman, 2025; Leonelli, 2025; Scharnhorst, 2025; Treloar, 2025; Wyatt, 2025; Yakel, 2025)**. Most of the commentaries are in the tone of “yes, and...” (boyd & Sarathy). These articles build upon our theoretical framework to explicate other factors in data reuse, both extending and critiquing our arguments.

The Many Roles of Intermediaries in Reusing Data

Sharing and reusing research data is a vast topic, far broader than can be addressed in a single article, even with our extensive reference list. We constrain the scope of our investigation by focusing on the distance between data creators, taken as a point of origin, and reusers of those data, taken as an endpoint, along six dimensions: domain, methods, collaboration, curation, distance, purposes, and time and temporality. Many intermediaries are involved in knowledge exchange between data creators and data reusers, whether human or machine, and whether to serve practice or policy. Yet more intermediaries play roles in building the knowledge infrastructures necessary to create research data and to steward data in ways they can be reused. To use our commenters' terminology, intermediaries 'knit' (boyd & Sarathy; Treloar) together the 'data communities' (Scharnhorst) and networks of knowledge exchange.

In arguing that 'it takes a village' to share and to reuse data, we focus on four stakeholders in data exchange and relationships between them: data creators, data reusers, data archivists, and funding agencies. Knowledge exchange is an expensive process with an imbalance of labor

and costs. Many questions arise about who should bear those costs and what kinds of intermediaries can accomplish the work most effectively. boyd and Sarathy translate the corporate function of ‘developer relations’ to the ‘data matchmaker’ who creates value through partnering and lubricating human exchanges. In business settings, owners of data products add value by serving as intermediaries between data creators and reusers (**Goedegebuure et al., 2025**). Acker, Crane et al., Leonelli, Scharnhorst, Treloar, Wyatt, and Yakel each identify roles in data exchange where ‘knowledge intermediaries’ (Leonelli’s term) can perform critical functions better than can the data creators or reuses themselves. Crane, Babeu, and Tauber, taking digital translations of Homer as a deep case study, portray classicists—experts in ancient languages and cultures—as intermediaries to contemporary, nonexpert audiences.

Intermediaries facilitate documentation and knowledge exchange at micro levels, such as between individual objects and steps in the research process (Scharnhorst, Treloar), and between research subject and researcher (Yakel). Other mediation roles occur at macro levels across public/private boundaries (boyd & Sarathy, Treloar), local and remote services

(Goodman), ancient and contemporary expertise (Crane et al.), and stakeholders in data governance (Wyatt). Our commenters emphasize the growing need for intermediaries, the invisibility of the work that makes data exchange possible, and call for far greater investment in these roles by funding agencies and institutions.

Data Reuse as Gift vs. Transaction

Throughout our article, we emphasize the ways in which research data exist in a web of complex social and technical relationships. Two of the comments (Wyatt, boyd & Sarathy) set our framing in the larger anthropological concept of 'gift exchange,' both citing the canonical work of Mauss, first published in 1925 (**Mauss, 2006, 2016**). Science and academe are rooted deeply in a gifting economy. Scholars give gifts such as reviewing each other's papers, writing reference letters, organizing conferences—and sharing data. Such labor is volunteered with the expectation that these gifts create indebtedness, encourage reciprocity, and enhance reputations.

The gift exchange culture is the trust fabric that underlies the knowledge infrastructure in which data are transferred between parties. As explained by Wyatt and by boyd and Sarathy, the

gifting economy of academe is breaking down due to scale, structural inequities, economic dynamics, and other factors. Wyatt compares the growing alarm about breakdowns in gifting, trust, and reciprocity of Mertonian norms (**Merton, 1942**) arising in World War II, responding to the democratic social order necessary for science, to similar concerns of today about disinvestment in science, misinformation, attacks on academic freedom, and dystopian scientific futures. boyd and Sarathy view data sharing as a 'Maussian bargain' of open science, expressing similar concerns about equity, and identifying ways in which funding agencies can contribute to knitting the social fabric back together.

Acker builds upon Mirowski's (**2012, 2018, 2023**) critiques of open science to assess the consequences of these breakdowns in Mertonian norms. Profit-driven platforms arose to lubricate the flow of data from lab to analysis to publication to deposit to reuse, intervening in many components of the knowledge infrastructure necessary for data exchange. Acker explains how these platforms can amplify all six of the distance dimensions we identify. Platforms impose their own proprietary metadata formats and provenance chains, leading to enclosure that runs counter to open

science principles and disciplinary needs. Acker, Yakel, Wyatt, Scharnhorst, and boyd and Sarathy all highlight the commodification of data that results from the dominance of private platforms. Platforms induce a transactional relationship between parties, focusing on profit and efficiency. Breakdowns in the gift economy lead to breakdowns in actual knowledge exchange, which may undermine the conduct of science.

Credit is among the gifts exchanged in science, whether as data, citations, acknowledgments, tenure letters, peer reviews, or private debts. Exchanges of credit are subject to scaling factors. Within small communities, individuals know and call upon each other. At the vast scale of today's scientific communities, comprised of knowledge infrastructures that span disciplines, journals, and continents, credit itself is becoming commodified. As platforms enclose the means of knowledge exchange, they amplify distances between players (Acker). Some would replace social credit with transactional credit for sharing, reusing, citing, reviewing, or other scholarly contributions. However, these credit systems risk making data exchange even more transactional, breaking down collaborative relationships rather than enhancing science. Crane, Babeu, and Tauber discuss how

intellectual property licensing and aggregation of digital resources can both strengthen and undermine named credit to data creators.

Transparency is a key tenet of open science, on the grounds that access to evidence, methods, and process promotes trust. Transparency of scientific process does not necessitate inspection of sensitive content, such as medical or human subjects material, points pursued by Leonelli and Yakel. Leonelli in particular argues for methods such as data stories that do not necessitate data sharing directly but instead embrace the interpretative nature of data reuse. Crane, Babeu, and Tauber encourage better provenance tracking to promote transparency.

Technology, Scale, and Artificial Intelligence in Data Reuse

Our article frames data reuse as a sociotechnological process, exploring ways in which technology can lubricate—and can cause friction—in knowledge exchange. The commentaries identify numerous cases of technological lubrication and friction, many of which are related to the platform issues discussed above.

By focusing on relationships between those who create and those who reuse data, we largely sidestepped issues of how characteristics of data influence knowledge exchange. We also placed data mining and other forms of 'distant reuse' outside the scope of our article. Several commenters took the opportunity to extend our theory of knowledge exchange into matters of technology, scale, and artificial intelligence. Goodman, whose research straddles astronomy and computer science, suggests the metaphor of 'elevation' as a 'topographic contour' to our landscape. Large data sets, such as those common in physics, genomics, and many other fields, can only be exchanged via platform and cloud infrastructures. The means to acquire and reuse data will vary along all the distance dimensions we identified. We mentioned the Alzheimer's Disease Neuroimaging Initiative as an illustration of how data reuse can be improved through specialized curation of large imaging and biomarker data sets. Several other commenters, including Acker, Leonelli, Scharnhorst, Treloar, and boyd and Sarathy offered examples of the technical infrastructures necessary for exchanging data at scale. In many research areas, the era of moving compute capacity to the data, rather than vice versa, has arrived (**Bourne et al., 2022; Szalay & Barish, 2019**). Leonelli argues for the

importance of data mining algorithms that visit data to deal with issues of data protection.

Data mining, a vast topic largely outside the scope of our article, increasingly will involve artificial intelligence tools. We offered the Common Crawl data set as an example of how generative AI presumes access to data at massive scale. Our commenters touched on implications of AI for data mining (Leonelli), benchmarking (boyd & Sarathy), repurposing (Treloar), and open science (Acker). Generative AI may lubricate data exchange by producing some forms of metadata and by tailoring documentation to particular users. Conversely, as platforms commodify data, they amplify distances between data creators and reusers. Concerns arose about how AI may destroy context and provenance (Acker, Treloar). Scharnhorst argues for better traces of 'data history.' Crane, Babeu, and Tauber argue for provenance records that capture credit more discretely, as incentives for participating in the gift exchange culture of scholarship.

Conclusions

We developed our theory of how distance matters in creating and reusing data in the spirit of the 'long now,' the physical and metaphorical 10,000-year clock created by Stewart Brand and

Danny Hillis as an 'invitation to think and engineer at the timescale of civilization' (**The Long Now Foundation, 2025**). Knowledge infrastructures should be constructed with similar attention to long-term sustainability (**Ribes & Finholt, 2009**). Of the six dimensions of distance we developed, time and temporality is the one least examined in these commentaries, and the one least well explored in current literature on data reuse. Wyatt extends our historical framing back to the mid-20th century; Crane, Babeu, and Tauber take us back to the fifth century B.C.E. Most of the commenters take us forward into uncertain technical and political futures, not all of which are optimistic about sustainability.

Our treatment of the sociotechnical landscape of data sharing and reuse is extensive but not exhaustive. These impressive commentaries suggest the breadth and richness of the scholarship on data practices and data reuse, extending our original reference list of about 230 items by at least 100 more. We offer 'distance' as a theoretical construct to advance research and as a practical metaphor for data creators, data reusers, archivists and other intermediaries, and research funders. Crane, Babeu, and Tauber employ all six of our dimensions in a case study of data sharing and

reuse of digital resources for Homer's *Illiad* and *Odyssey*, framing the situation in Greco-Roman studies. We strongly encourage the use of our model for case studies and hypothesis testing in other domains.

This forum on how 'distance matters' in data sharing and reuse, kindly invited by Xiao-Li Meng as Editor-in-Chief of *Harvard Data Science Review*, is a rare opportunity to highlight the intellectual and disciplinary range of concerns about the stewardship of research data. Data are never 'raw,' nor are they tangible objects to be bought, sold, exchanged, or stored. However, data are expensive, whether as gifts or commodities. Their value lies in context, in the knowledge infrastructures in which they exist, and in people and organizations whose contributions to stewardship are too often invisible.

Disclosure Statement

Christine L. Borgman and Paul Groth have no financial or nonfinancial disclosures to share for this article.

References

Acker, A. (2025). How platforms constrain matters of distance and data reuse. *Harvard Data Science Review*, 7(2).

<https://doi.org/10.1162/99608f92.c0d03e91>

Borgman, C. L., & Groth, P. T. (2025). From data creator to data reuser: Distance matters. *Harvard Data Science Review*, 7(2).

<https://doi.org/10.1162/99608f92.35d32cfc>

Bourne, P. E., Bonazzi, V., Brand, A., Carroll, B., Foster, I., Guha, R. V., Hanisch, R., Keller, S. A., Kennedy, M. L., Kirkpatrick, C., Mons, B., Nusser, S. M., Stebbins, M., Strawn, G., & Szalay, A. (2022). Playing catch-up in building an open research commons. *Science*, 377(6603), 256–258.

<https://doi.org/10.1126/science.abo5947>

boyd, d., & Sarathy, J. (2025). Strengthening human infrastructure through data matchmakers. *Harvard Data Science Review*, 7(2).

<https://doi.org/10.1162/99608f92.c35ca0f3>

Crane, G., Babeu, A., & Tauber, J. (2025). Humanities data reuse: Humanity first. *Harvard Data Science Review*, 7(2).

<https://doi.org/10.1162/99608f92.598a17b4>

Goedegebuure, A., Kumara, I., Driessen, S., Van Den Heuvel, W.-J., Monsieur, G., Tamburri, D. A., & Nucci, D. D. (2025). Data mesh: A systematic

gray literature review. *ACM Computing Surveys*, 57(1), 1–36. <https://doi.org/10.1145/3687301>

Goodman, A. A. (2025). Beyond distance: Why elevation matters too. *Harvard Data Science Review*, 7(2).
<https://doi.org/10.1162/99608f92.e488fbd9>

Leonelli, S. (2025). What to do about data distance? Responsible alternatives to data sharing. *Harvard Data Science Review*, 7(2).
<https://doi.org/10.1162/99608f92.e95b5c26>

Mauss, M. (2006). *The gift: The form and reason for exchange in archaic societies*. Psychology Press.

Mauss, M. (2016). *The gift: Expanded edition* (J. I. Guyer, Trans.; Expanded). Hau Press.

Merton, R. K. (1942). Science and technology in a democratic order. *Journal of Legal and Political Sociology*, 1, 115–126.

Mirowski, P. (2012). The modern commercialization of science is a passel of Ponzi schemes. *Social Epistemology*, 26(3–4), 285–310.
<https://doi.org/10.1080/02691728.2012.697210>

Mirowski, P. (2018). The future(s) of open science. *Social Studies of Science*, 48(2), 171–203.
<https://doi.org/10.1177/0306312718772086>

Mirowski, P. (2023). The evolution of platform science. *Social Research: An International Quarterly*, 90(4), 725–755.
<https://muse.jhu.edu/pub/1/article/916352>

Ribes, D., & Finholt, T. (2009). The long now of technology infrastructure: Articulating tensions in development. *Journal of the Association for Information Systems*, 10(5).
<https://doi.org/10.17705/1jais.00199>

Scharnhorst, A. (2025). Distance matters and data in the evolution of knowledge. *Harvard Data Science Review*, 7(2).
<https://doi.org/10.1162/99608f92.5cflad26>

Szalay, A. S., & Barish, B. C. (2019). The emergence of long-lived, high-value data collections. *Bulletin of the AAS*, 51(7).
<https://baas.aas.org/pub/2020n7i016/release/1>

The Long Now Foundation. (2025). Long Now.
<https://longnow.org/>

Treloar, A. (2025). Dimensions of distance: Descriptive or distracting? *Harvard Data Science Review*, 7(2).
<https://doi.org/10.1162/99608f92.71c3e176>

Wyatt, S. (2025). Distance and other metaphors about reuse of research data. *Harvard Data*

Science Review, 7(2).

<https://doi.org/10.1162/99608f92.1d109ab1>

Yakel, E. (2025). Going the distance achieving sufficiency in data reuse. *Harvard Data Science Review*, 7(2).

<https://doi.org/10.1162/99608f92.3a5ebf2b>

©2025 Christine L. Borgman and Paul Groth.

This article is licensed under a [Creative Commons Attribution \(CC BY 4.0\) International license](#), except where otherwise indicated with respect to particular material included in the article.

Connections

1 of 9 children and siblings ▼

Another **Commentary** on *From Data Creator to Data Reuser: Distance Matters*



Going the Distance: Achieving Sufficiency in Data Reuse

by Elizabeth Yakel

Published on Apr 30, 2025 ·
hdsr.mitpress.mit.edu

Another **Commentary** on *From Data Creator to Data Reuser: Distance Matters*



Distance and Other Metaphors About Reuse of Research Data

by Sally Wyatt

Published on Apr 30, 2025 ·
hdsr.mitpress.mit.edu

Another **Commentary** on *From Data Creator to Data Reuser:*

Distance Matters



Dimensions of Distance: Descriptive or Distracting?

by *Andrew Treloar*

Published on Apr 30, 2025 · hdsr.mitpress.mit.edu

Another **Commentary** on *From Data Creator to Data Reuser: Distance Matters*



Distance Matters and Data in the Evolution of Knowledge

by *Andrea Scharnhorst*

Published on Apr 30, 2025 · hdsr.mitpress.mit.edu

Another **Commentary** on *From Data Creator to Data Reuser: Distance Matters*



What to Do About Data Distance? Responsible Alternatives to Data Sharing

by *Sabina Leonelli*

Published on Apr 30, 2025 · hdsr.mitpress.mit.edu

Another **Commentary** on *From Data Creator to Data Reuser: Distance Matters*



Beyond Distance: Why Elevation Matters Too

by *Alyssa Goodman*

Published on Apr 30, 2025 · hdsr.mitpress.mit.edu

Another **Commentary** on *From Data Creator to Data Reuser: Distance Matters*



Humanities Data Reuse: Humanity First

by *Gregory Crane, Alison Babeu, and James Tauber*

Published on Apr 30, 2025 · hdsr.mitpress.mit.edu

Another **Commentary** on *From Data Creator to Data Reuser: Distance Matters*

Distance Matters



How Platforms Constrain Matters of Distance and Data Reuse

by *Amelia Acker*

Published on Apr 30, 2025 · hdsr.mitpress.mit.edu

Another Commentary on From Data Creator to Data Reuser: Distance Matters



Strengthening Human Infrastructure Through Data Matchmakers

by *danah boyd and Jayshree Sarathy*

Published on Apr 30, 2025 · hdsr.mitpress.mit.edu

REFERENCES • 21

LICENSE • Creative Commons Attribution 4.0 International License (CC-BY 4.0)

COMMENTS • 0

?



No comments here

Why not start the discussion?