Response to Editor and Referee

## To Editor

*I’d also like to suggest that a few things be cleaned up on the next revision, notably the references to OKD on pages 4 and 7 (you probably meant to say OKF)*

Correct, thanks for pointing that out. Defined, and then replaced.

*and the incomplete citation and references on page 7, where you have CENSUS, YEAR and xxx as stand-ins for information that was never added later.*

Thanks. I have added (improved) references, and more complete information for US, Canada, and France.

I have also had to remove a statement I had included, based on preliminary analysis by CIDR @ Berkeley. In the earlier edition, I stated

*In fact, in a recent survey (CIDR), respondents were asked to name the top two factors that could support early-career African scholars’ publication success. 46\% of respondents chose `Providing access to datasets and data management tools' as one of the two factors. University staff and students (across multiple disciplines) were also asked which aspect they most needed funding for, and 38\% mentioned `Data analysis software'.*

However, when following up with CIDR to obtain the final citation, they noted that the final analysis no longer supports this statement. The final analysis only supports the following statement:

In fact, in a recent survey conducted by CIDR, respondents were asked to name the top two factors that could support early-career African scholars’ publication success. 55\% of African scholars noted that providing access to research resources, such as journals, datasets, and software would improve their journal acceptance rate.

I have instead, as you suggested, looked into worldwide patterns of software downloads, based on data I had recently accessed for Stata packages, as well as regularly updated data on downloads of R software. I do note in the text that this is indicative and not definitive because of various definitional weaknesses (and add a footnote to that extent). The paragraph and figure appear on or around Page 11.

## To Referee

Thank you for the kind assessment of my article, and for the excellent comments, which have allowed me to (I hope) improve the manuscript. Let me address each point individually.

The discussion on page 3 about the size of networks required to be considered

“open” lacks a clear theoretical grounding. The author might consider

clarifying why a large number of agents is assumed to be a necessary (or

sufficient) condition for openness. For instance, even a two-agent network

could, in principle, be “open” if the access conditions are met. In Section

2, the author writes that “the number of people who have some probability

of accessing the resource (the size of the network) can be taken as an

approximate measure of openness.” If this definition is indeed adopted, it

should be stated more explicitly.

Thank you for pointing out that lack of clarity. I do note that in my attempt at a actionable definition of openness, it is the **possibility** of access, not the actual realized access in the real world that matters. A two-agent network, or rather, the fact that only two agents have accessed the network, is not per-se informative. Rather, it is the set of possible who could potentially, now or in the future, access that network, that characterizes the openness of the network.

I have removed the somewhat flippant “Clearly, n=2…” sentence, which detracts from the main idea. I have added a few sentences that clarify that I have potential entrants in mind, with the realized size (relative to some potential pool of entrants) a proxy for that.

2. While the paper offers a rich and well-informed narrative, some terms like

“openness,” “access,” and “proprietary” are used before being precisely

defined. A more systematic early clarification would help readers better

navigate the distinctions made throughout the article.

I attempted to do so with the “Concepts” section, but clearly did not fully implement this. I have gone through the entire text, and moved concepts that needed a definition to the “Concepts” section, for greater clarity. I hope I did not miss any. I left any open definitions in the introduction “dangling”, as they will then be resolved in the subsequent “Concepts” section.

3. My main concern about the paper is that the section entitled “The Benefits

of Open Science in Economics” is too elliptical. In my view, this is the

central issue, and the discussion focused on the two papers by Roth (2022)

and Chaisemartin and Ramirez-Cuellar (2024) is too limited. A substantial

literature has been devoted to this topic, both in the sciences more

broadly and in economics in particular. The author could make a greater

effort to convince the reader of the actual benefits of the accessibility she

promotes. While I am personally convinced of the value of open science,

it should not be treated merely as a self-evident paradigm. Rather, the

paper should more explicitly demonstrate its advantages for the scientific

community.

I agree that the grandly titled section under-delivers. I have added 3 paragraphs to this section, referencing a few of the works that make a much more comprehensive job discussing benefits, including some figures from Ferguson et al (2023), before then referencing the concrete example I bring to this discussion, which I do believe is a unique contribution; I am not aware of others that have tried to even anecdotally, as here, enumerate publications that explicitly rely on prior articles’ open data and code. I have added a third example.

4. In the paragraph devoted to organic data (page 7), the author mentions

that “4% of US-based empirical authors have had some access to the US

system for providing access (FSRDC).” It would be helpful to include

similar figures for other data infrastructures, such as the Centre d’acc`es

s´ecuris´e aux donn´ees (CASD) in France, if available. More generally, while

the paper primarily draws on U.S.-based infrastructures and institutions

(e.g., AEA, NBER, FSRDC), and the author has made some effort to

include examples from other contexts, this comparative perspective could

be further developed to broaden the article’s relevance and appeal.

Unfortunately, I am not aware of other studies that come to similar quantitative answers, though I believe it would be feasible to do so for CASD and IAB data (both keep track of projects and publications). I have added some contextual numbers for CASD and the Canadian network in terms of absolute numbers.

## Response to Minor comments

1. The sentence ”For non-local authors, this meant traveling for extended

time periods as a Ph.D. student (Margolis) or spending a sabbatical in

Paris (Abowd), neither of which is a cheap endeavor.” should be placed

in a footnote.

Done.

2. Section 2: Concepts. The juxtaposition of the UNESCO and OKF definitions

could benefit from further commentary. For instance, the author

might clarify whether these definitions are complementary or whether

they reflect fundamentally different paradigms (institutional vs. grassroots

openness).

My contribution in this article is to describe “open science” in economics, not to define “open science”. To clearly identify that I don’t want to disambiguate multiple definitions, I have added “, and multiple overlapping definition are commonly referenced.”, which, in combination with my subsequent sentence “In this article, I will focus…” will hopefully clarify that I chose one, without loss of generality.

3. Figure 1 presents an original and thought-provoking representation of access

trade-offs. However, the theoretical justification for using time and

network size as the two key summary metrics (second paragraph, page 4)

remains insufficiently developed. For instance, if data from a company

such as Meta are publicly accessible but require eight hours to use them

through the CASD infrastructure, would this disqualify them from being

considered “open”? Conversely, if a dataset is accessible instantly and

without cost, but is only of interest to two researchers due to its niche

nature (e.g., “How many angels can dance on the head of a pin?”), would

it also fail to qualify as open data under this framework? These examples

suggest that the conceptual link between access time, audience size, and

openness could benefit from further clarification and theoretical grounding.

I am personally convinced by the relevance of this figure, but a more

detailed justification would enhance the persuasiveness of the framework

for the readers.

The figure is meant to convey the trade-off to how many \*potential\* users there are, relative to the costs of doing so. It is not meant to convey anything about value of such data, which the referee’s question (“interest”) implies. I have adjusted the text to read “, regardless of how interesting or valuable they might consider the data to be.” in order to make it clear that this incorporates no value judgement.

4. The Section 4 (Access to Software) gives an excellent empirical illustration

of cost asymmetries in software access (Table 3). However, it could be enriched

by connecting these observations to existing literature on inequality

in research capacity across countries or institutions. This would reinforce

the policy relevance of the argument and open the door to actionable

recommendations.

I have described the importance of open data in the Benefits section by referencing a World Bank report on the issue. I am not aware of any discussion about software, despite asking many experts in the field. I have added to the cost asymmetries by adding for context laptop prices from each respective country/area.