RESEARCH ARTICLE

The varying openness of digital open science tools [version 1; peer review: 1 approved with reservations, 1 not approved]

Louise Bezuidenhout ¹, Johanna Havemann ²

¹Institute for Science, Innovation and Society, University of Oxford, Oxford, Oxfordshire, OX2 6PN, UK

²Access2Perspectives, Berlin, Germany

V1 First published: 02 Nov 2020, 9:1292
<https://doi.org/10.12688/f1000research.26615.1>
 Latest published: 17 May 2021, 9:1292
<https://doi.org/10.12688/f1000research.26615.2>

Abstract

Background: Digital tools that support open science practices play a key role in the seamless accumulation, archiving and dissemination of scholarly data, outcomes and conclusions. Despite their integration into open science practices, the providence and design of these digital tools are rarely explicitly scrutinized. This means that influential factors, such as the funding models of the parent organizations, their geographic location, and the dependency on digital infrastructures are rarely considered. Suggestions from literature and anecdotal evidence already draw attention to the impact of these factors, and raise the question of whether the open science ecosystem can realise the aspiration to become a truly “unlimited digital commons” in its current structure.

Methods: In an online research approach, we compiled and analysed the geolocation, terms and conditions as well as funding models of 242 digital tools increasingly being used by researchers in various disciplines.

Results: Our findings indicate that design decisions and restrictions are biased towards researchers in North American and European scholarly communities. In order to make the future open science ecosystem inclusive and operable for researchers in all world regions including Africa, Latin America, Asia and Oceania, those should be actively included in design decision processes.

Conclusions: Digital open science tools carry the promise of enabling collaboration across disciplines, world regions and language groups through responsive design. We therefore encourage long term funding mechanisms and ethnically as well as culturally inclusive approaches serving local prerequisites and conditions to tool design and construction allowing a globally connected digital research infrastructure to evolve in a regionally balanced manner.

Keywords

Open Science, digital, reproducible, low/middle-income countries

Open Peer Review

Approval Status

	1	2
version 2		
(revision)		
17 May 2021		
version 1		
02 Nov 2020		

1. **Björn Brembs** , Universität Regensburg, Regensburg, Germany

2. **Chris H. J. Hartgerink** , Liberate Science GmbH, Berlin, Germany

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the **Research on Research, Policy & Culture gateway**.

Corresponding author: Louise Bezuidenhout (louise.bezuidenhout@insis.ox.ac.uk)

Author roles: **Bezuidenhout L:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Havemann J:** Conceptualization, Formal Analysis, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: No grants were involved in supporting JH in this work and ESRC UK supported LB in this work.

Copyright: © 2020 Bezuidenhout L and Havemann J. This is an open access article distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Bezuidenhout L and Havemann J. **The varying openness of digital open science tools [version 1; peer review: 1 approved with reservations, 1 not approved]** F1000Research 2020, 9:1292 <https://doi.org/10.12688/f1000research.26615.1>

First published: 02 Nov 2020, 9:1292 <https://doi.org/10.12688/f1000research.26615.1>

Introduction

The evolution of the open science ecosystem

Open science encompasses a collection of activities, principles and tools oriented at making scientific research accessible to all levels of society proposed to increase transparency and efficiency in research workflows and scholarly publishing (Rahal & Havemann, 2019). Open science activities are clustered around a number of areas of action, including open data, open access (OA), open educational resources (OER), free and open source software (FOSS), open hardware, open methodologies and open peer review, including the growing citizen science movement and broader societal engagement.

The open science movement has garnered support from both individual researchers as well as high-level policy and funding around the world and given rise to a range of influential regional high-level and grassroots initiatives alike in Africa¹, IberoAmerica², Europe³, North America⁴, Asia⁵ and Oceania⁶ as well as several independent and cross-regional networks and community initiatives⁷. The global scientific community is increasingly recognizing the benefits of learning from each other and aligning technically feasible approaches adopted to regional infrastructure prerequisites. At the same time, research communities are contributing to the development of resources, practice change and activism to establish Open Science and incorporate it into mainstream research workflows; The Open Science MOOC, FOSTER Open Science and the Open Scholarship Knowledge Base (OSKB) are just a few of many examples⁸. The oldest and most visible of these communities are within free and open source software (FOSS) development (Powell, 2012). In recent years, community activities are extending to a wide range of areas, including community-driven and often volunteer-run preprint repository platforms⁹, open peer review services¹⁰, and capacity building programs and training

resources¹¹. These wide-ranging activities are united under core values, such as openness, equitable sharing, access to resources and optimized re-use (Tennant *et al.*, 2019).

The ongoing coronavirus 2019 (COVID-19) pandemic has drawn attention to the key role of openness in research (OECD, 2020). Widespread commitment to openness in COVID-19 research by funders, governments, research institutions and individual researchers has showcased the impact of rapid OA publishing, open data sharing and the open and collective design of hardware (Maia Chagas *et al.*, 2020; Zastrow, 2020)¹². All of these areas have underscored the importance of open research practices as a means to increase efficiency and speed of information sharing. This has been vital not only for medical research, but also for policy makers and practitioners in responding to the impact of COVID-19 on society.

Digital tools supporting open science activities

The increasing support for open research activities provides an opportune moment for critical reflection on the open science movement so far. In particular, it prompts a critical assessment of the evolution of the digital infrastructures, tools and online working practices that underpin open research activities. Of these different areas, the design, deployment and use of the digital tools that support open science activities are the least scrutinized. Indeed, critical evaluations of the evolving landscape of interlinked digital tools supporting open science are scarce (Kramer & Bosman, 2016).

Digital tools are a ubiquitous part of open science. Most steps of the research workflow are nowadays complemented or replaced by online applications. These tools assist researchers to share and collaborate, and thus increase openness and transparency at all stages of the research lifecycle. Many of these tools have changed the way that research is done and how research resources – including datasets, publications, educational resources and software – are circulated globally (Kramer & Bosman, 2016).

In this paper, we collectively term these tools “digital open science tools” (DOSTs). This category encompasses the wide range of digital tools that are involved in facilitating openness during the research lifecycle. In this paper, the term ‘DOST’ includes any digital tool (for-profit, non-profit and community-led entities) used in open research, irrespective of whether they were designed explicitly for open science or have been co-opted into open science practices.

Growing efforts to promote interoperability and open workflows have made interconnection key to the success of any DOST (Wilkinson *et al.*, 2016). The interconnection of tools and the interoperability of their outputs enable users to move

¹ Africa: <http://africanopenscience.org.za/>; <https://info.africanarxiv.org/>; <https://savoirs.cames.online/jsui/>; <http://africaosh.com/>

² Ibero America: <http://amelica.org/index.php/en/home/>; <https://www.redalyc.org/>; <https://scielo.org/>; <http://mutabit.com/grafoscopio/index.en.html>

³ Europe / EU: <https://ec.europa.eu/research/openscience/index.cfm>; Germany: <https://www.osc.uni-muenchen.de/toolbox/index.html>

⁴ North America / USA: <https://www.cos.io/>; <https://our-research.org/>; Canada: <https://pkp.sfu.ca/ops/>

⁵ Asia / Indonesia: <https://rinarxiv.lipi.go.id/lipi/>; India: <https://indiarxiv.in/>; Japan: <https://openscience.jp/>

⁶ Oceania / Australia: <https://www.freeourknowledge.org/>

⁷ Cross-regional: <https://opensciencetools.org/>; <http://openhardware.science/>

⁸ For example, <https://opensciencemooc.eu/>, <https://www.fosteropenscience.eu/>, <https://www.oercommons.org/hubs/OSKB>

⁹ For example, <https://osf.io/preprints/>; <https://www.preprints.org/> - and for an overview of biological-focused pre-print archives see <https://asapbio.org/preprint-servers>

¹⁰ For example, <https://prereview.org/>, <https://peercommunityin.org/>

¹¹ For example, www.fosteropenscience.eu <https://asapbio.org/preprint-servers>

¹² https://www.eurekalert.org/pub_releases/2020-08/hl-crr081220.php and <http://www.oecd.org/coronavirus/policy-responses/why-open-science-is-critical-to-combating-covid-19-cd6ab2f9/>, <https://asapbio.org/preprints-and-covid-19>

between tools at different stages of the research lifecycle to facilitate research, data dissemination and publication. The interconnectedness of the tools, as well as the overlaps in their function within the DOST landscape (“multiplicity”), means that multiple “pathways” exist for data to progress through a research lifecycle (see [Figure 1](#)). How these “pathways” are selected depend on a variety of issues such as user preference, access to specific DOSTs, demands of the research project and preferences of the research community.

New tools are continually being added to the DOST landscape, and new connections between tools are regularly emerging to populate this ecosystem. We use the term “ecosystem” in contrast to the more common “landscape” to designate the dynamism of the online environment as an interconnected system through which resources move. We ground this understanding in biological understandings of ecosystems as biological communities of interacting organisms and non-living components that interact as a system. This DOST ecosystem is dynamic, multiplicitous and subject to internal and external pressures. It includes interconnected/interdependent DOSTs, as well as the

information and communication infrastructures, communities of users and socio-political stakeholders. Internal and external pressures from these actors determine the persistence of the DOSTs and the structure of the ecosystem.

The underlying dynamics and influences within the evolving DOST ecosystem have been extremely influential in driving forward the Open Science movement as a whole. Tools within the ecosystem, such as [GitHub](#), are changing the way collaborations are managed. Publishers like [PLoS](#) and [F1000Research](#) are redefining transparent publishing models, and repositories such as [Zenodo](#), [Open Science Framework](#) (OSF) and [DSpace](#) are offering open platforms for sharing and re-using data.

The constantly growing uptake in usage of DOSTs, and their increasing interconnectivity and interoperability may give the impression that the digital landscape of open science is positively unfolding and developing to support the growing needs of the open science community. Widespread endorsement of many DOSTs, support from socio-political actors and the rapid organization of “user communities” associated with specific

Figure 1. A) Diagram from [Kramer & Bosman \(2016\)](#)¹³ demonstrating diversity of DOSTs, linkages between tools at different stages of workflow. Green line demonstrates a potential research workflow involving DOSTs. Image shared under [CC-BY license](#). **B)** Pictogram of a random digital tool representing the tools displayed in 1A with influencing aspects addressed in this paper: underlying values, financial models, language choices, geographical location, user communities.

¹³ <https://101innovations.wordpress.com/workflows/> (accessed 10 August 2020)

tools has left little time for critical reflection on how the ecosystem is evolving and what power dynamics are shaping its evolution. How, it is increasingly being asked, do the tools developed by multiple scholarly for-profit service providers, non-profit organizations and open source communities contribute to the open science vision and mission to make research workflows and results accessible to all sectors of society across the globe?

In this paper we critically interrogate the DOST ecosystem. We ask how its current structure enables knowledge availability and question whether social, political or economic barriers linked to DOST design and deployment undermine this objective. To do so, we ask three main questions of the ecosystem and its actors:

1. What is the impact of a small number of countries dominating DOST design and deployment?
2. Do heterogeneities in values, funding, and stakeholders that influence tool design and interconnection affect the openness of the DOST ecosystem?
3. How (if at all) are external power dynamics and influences recognized and addressed in the DOST ecosystem?

The subsections below provide a short background to these three questions, and frame the empirical data presented in the following sections.

Geographic distribution of DOSTs and user communities

The Open Science movement supports the democratization of research resources. Increasing openness in research will make resources available to all individuals in all nations and at all levels of society¹⁴. In this way, Open Science promotes equitable access to resources through the (self-described) model of the “knowledge commons” (Hess & Ostrom, 2007) which promotes a form of direct democracy, where every individual has the right--and ability--to access information, data, and content that is collectively owned and managed by a community of users. While this direct democracy model works well as a model of resource distribution, it complicates the evolution of the DOST ecosystem. The distribution of researchers and resources around the world is unevenly weighted towards a small number of high-income countries (HICs). A 2013 report by UNESCO highlighted that China, the European Union, Japan, the Russian Federation and the USA together accounted for 72% of researchers worldwide. Unsurprisingly, the evolution of DOSTs reflects this distribution, with the majority of tools being developed in countries with a high density of researchers and considerable investment in research and national digital infrastructures. As a result, the design of the tools and the evolution of user communities – as dictated by the majority of users – is weighted in favour of a small number of countries.

The specific geographic location of many DOSTs, as reflected by the location of their development, registration and hosting, contrasts to the approach championed by the FOSS movement. FOSS has long been calling for and implementing a more representative form of democracy for software development and distribution by promoting models that avoid specific geographically-clustered nodes (Tennant *et al.*, 2020; Vermeir *et al.*, 2018). This model of “software mirrors”¹⁵ is commonly used in systems such as GNU as well as Linux distributions like Debian and Fedora. Nonetheless, and likely due in part to economies of scale, this approach of mirroring services to increase access has not been replicated within the DOST ecosystem.

Heterogeneities in purpose and design of DOSTs

The open science movement promotes widely agreed values that also define good scientific practice. These include openness, credibility, reproducibility, and verifiability of any research output (Bartling & Friesike, 2014). Nonetheless, the endorsement of these core values can cause the widespread value/practice-heterogeneity within the open science movement to be overlooked. Indeed, Open science can be thought of both as a practice and as a philosophy (Levin *et al.*, 2016), implying that the motivations for individuals to get involved can vary considerably (Fecher & Friesike, 2014).

Fecher & Fiesike (2014: 17) mention five different open science schools of thought: The *infrastructure school* (which is concerned with the technological architecture), the *public school* (which is concerned with the accessibility of knowledge creation), the *measurement school* (which is concerned with alternative impact measurement), the *democratic school* (which is concerned with access to knowledge) and the *pragmatic school* (which is concerned with collaborative research). It can thus not be assumed that everyone is motivated to a similar degree by the core values. A number of pragmatic reasons also play important roles in the uptake of Open Science practices and tools, including efficiency, career advancement, journal and institutional requirements and community expectations (Ferguson, 2014).

This heterogeneity is further complicated by the number of actors within the DOST ecosystem. The unrestricted development of DOSTs has caused this space to be populated by stakeholders ranging from community projects to commercial companies. These different actors may have highly variable reasons for developing the DOSTs, and rely on highly disparate funding sources to ensure their longevity. While some DOSTs are explicitly designed to further open research practices, some may be a commercial venture responding to a gap in the market. Indeed, the highly variable development of DOSTs has led to the uncoordinated evolution of the DOST ecosystem, meaning that the financial, governmental and infrastructural influences are poorly understood.

¹⁴ <https://en.unesco.org/science-sustainable-future/open-science>

¹⁵ A software mirror is a server that provides an exact copy of data from another server. These mirrors can be held in different geographic locations and are intended to provide fault tolerance, or a means of redundancy in case something goes wrong with the primary or “principal” server.

The ways and reasons through which user-communities are recruited around DOSTs – as with any form of technology – are similarly diverse. These may range from bottom-up community endorsement, advertising, integration with other DOSTs or commercial endorsement. The persistence of a DOST within the ecosystem can thus depend on a range of different reasons, including accessibility, ease of use, visibility through advertising and promotion, or simply that the size of the user community allows it to dominate similar DOSTs (Mody, 2011).

Recognizing the heterogeneity inherent in the motivations for creating DOSTs and recruiting user communities is critical. It negates the assumption that endorsement from members of the open science community means that the tool is designed or deployed to optimally promote the values of the open science movement. To the contrary, the persistence of certain DOSTs over others depends as much on market forces and user preferences as on alignment with open science values.

External power dynamics

Research occurs within highly complex networks of power and influence of financial, governmental and societal actors (Vermeir *et al.*, 2018). As discussed above, the DOST ecosystem, while digital, relies on funders, hosts and infrastructures that are very much located in the physical world. DOSTs are thus subject to national legislation and regulation. Moreover, the ecosystem relies on information and communication infrastructures that are neither open nor designed with openness in mind. Service providers, content delivery networks, and cloud storage facilities, for example, are largely user-agnostic and operated by large international companies, yet are becoming extremely influential in the construction of the DOST ecosystem.

A complex ecosystem of digital heterogeneity

The rapid and diverse evolution of DOSTs has caused an exponential expansion of the ecosystem. In this dynamic space, researchers are continually provided with more options for integrating openness into their daily research workflow. Nonetheless, the rapid expansion of DOSTs and their insertion into the Open Science ecosystem requires careful scrutiny. At present, there is little critical examination of what tools are integrated into—and persist in—this ecosystem, what forces/values/preferences dictate how they are connected, why they are used, and what underlying infrastructures are being endorsed/supported by their presence within the DOST ecosystem.

Recognizing such concerns makes it apparent that the DOST ecosystem cannot be taken as *de facto* open, equitable and transparent. The range of actors and the interconnectivity of the tools makes it likely that there are a range of barriers that hamper certain users from engaging both in the tools and the workflows that they are embedded within. Key considerations include:

- Tools may be uncritically integrated into the ecosystem causing existing power dynamics to be perpetuated, leading to the marginalization of certain user groups

- Governments and commercial companies have undue influence on the landscape due to their hosting, financing, and otherwise influential roles
- The existing DOST ecosystem may become prescriptive of a specific way of “doing”, as one tool becomes hyper-dominant

Table 1 adapts the concept of “data assemblages” developed in Critical Data Studies for use in outlining the DOST ecosystem. Data assemblage refers to the technological, political, social and economic apparatuses and elements that constitute and frame the generation, circulation and deployment of data (Kitchin & Lauriault, 2014). Just as data assemblages map the complex set of stakeholders and pressures that influence the production, dissemination and reuse of data, Table 1 highlights some of the key pressures on the DOST ecosystem and their potential impact on the evolution of these spaces.

To our knowledge, there have not yet been attempts to provide an overview of how the DOST ecosystem shifts and adapts to these pressures. Indeed, the heterogeneity not only of the ecosystem, but also the actors and pressures that influence it, make this a challenging task. This paper presents a methodological attempt to map a selection of the DOST ecosystem including links between the tools. Our intention is to generate an interactive map of the DOST ecosystem so as to be able to test pressure and tipping points that shape ecosystem make-up and functioning.

Methods

DOSTs were identified from a range of different sources. The primary database was developed from two key studies, conducted on open science tools¹⁶. The database was extended by web searches and tools foregrounded in key open science communities such as the Research Data Alliance and the Open Science MOOC. As mentioned in the introduction, we used a very broad definition of DOSTs and included commercial, non-profit and community-driven digital tools that are currently used in open research. We did not make the availability of source code a prerequisite for inclusion. Neither did we limit the tools to those provided free to users. Inclusion criteria for the database were:

- The tool must be currently active and available for use online
- The tool must have a website detailing its function and activity
- There must be evidence of the use of the tool in an open or collaborative research project

Each tool was assessed according to the criteria outlined in section below. We developed the categories based on our analysis of ecosystem pressures presented in Table 1. The information

¹⁶ <https://101innovations.wordpress.com/> and <https://jroost.org/> (accessed 17 June 2020)

Table 1. Key pressures on the DOST ecosystem.

Attributes	Elements	Impact on OS ecosystem
Objectives driving tool creation	Business models, investment, venture capital, grants, philanthropy, profit, community activities	Influence on design decisions Longevity
Recruitment of user community marketing	Word of mouth, advertising, sponsorship, mandated by funder, institution, government, disciplinary community	Evolution of user community Prioritisation of tool over competitors, alternatives
Integration with other tools	Intentional design to connect with specified tools, widespread adoption into other tool designs, community-evolved connections	Interoperability
Host	Host organization, host country	Requirements and expectations of host, political constituency, interruption through economic sanctions
Regulations and legislation	Location-specific legislation, selection of regulation, oversight of activities	Financial legislation Oversight and mandated practice Selection of other codes, regulations, requirements
Materialities and infrastructures	Reliance on underlying digital landscape and information and communication technologies	Integration into digital landscape Financial/technical resources required to effectively use tool
Systems of thought	Social and political values, rationales	Endorsement and influence by capitalism, democracy, egalitarianism, socialism
Specified practice	Data formats, language, software systems	Need for data standards, file formats, user language, etc

used to populate these categories was freely available on the respective websites, each of which was examined by both of authors. Database entries were cross-checked by the authors in duplicate. Discrepancies were discussed until consensus was reached. Using the network mapping criteria above, a database was developed for analysis.

The paper is based on the 3 September 2020 version of this database (Bezuidenhout & Havemann, 2020). It is anticipated that the database will continue to evolve with community input. Contributions to the evolving database are encouraged through communication with authors.

Sorting criteria for DOST database

In this paper we examine the current dataset which includes 242 DOSTs focusing on the information about language, T&Cs, Host institution, and sponsor or funding institutions. The columns in the dataset display the sorting criteria that were applied as follows:

Workflow step: At what point(s) during the research workflow is the tool primarily used? – *Discovery, Analysis, Writing, Publishing, Outreach, Assessment*¹⁷

Open science category: Which subsection of the open science movement is the tool most closely related? – *Open hardware, Open educational resources, Open methodology, Open access, Open data, Open peer review, FOSS (free and open source software), Open lab notebook, Open Science [general category for multi-purpose tools]*.

Host (where applicable): Is the tool hosted by an organization other than itself? – *Named organization, otherwise 'Self', i.e. self-hosted*

Location / host location: In which world region is the tool or host located or registered? – *US (United States of America), UK (United Kingdom), EU (European Union), other|specific country, unspecified*

Language: What interface and description language is offered by the tool? – *Named language*

Funding source: How is the tool funded? – *Commercial, Various commercial, Grant, Various grant, Various mixed [commercial and grant], Institution*

Type of entity: How are the tool activities governed? – *NPO (nonprofit organization), Host affiliated, Commercial, Independent*

User fee: Does the tool require a fee to use all or part of its services? – *Free, Freemium, Membership fee, Services fee, APC (article processing charge)*

¹⁷ Workflow steps as defined by Bianca Kramer and Jeroen Bosman from the University of Utrecht. <http://innoscholcomm.silk.co/page/Workflows>. Accessed 20 March 2020.

Terms and Conditions: Are users in specific countries prohibited to use the tool? – *Explicit prohibition, Flags possible problems, No terms of use given, None mentioned*

Disclaimers

While the database produced provides an extensive list of OS tools, it is by no means exhaustive. By making the database an open resource we anticipate that it will be continually discussed and updated, both by ourselves and other practitioners in the field. In this way, the DOST dataset may become a reference resource for the Open Science community for digital tools development and optimization.

Most of the tools included in the database have been developed in Anglophone countries with English interfaces. The authors recognize this linguistic bias and are committed to working with Open Science community members from various linguistic communities to make the future iterations of the database more representative of the global scholarly community and tools available.

Finally, it was not possible to map all the existing institutional repositories due to their high numbers and transitional states and unclear or lacking institutional affiliations. The repositories represented in the database are hosted and maintained by NGOs or small companies. It is anticipated that institutional

repositories can be added over time by community crowdsourcing.

Results

Visualization of dataset

An interactive visual map of the 242 DOSTs was generated in Kumu. The interactive plot can be viewed here: <https://kumu.io/a2p/dost>. The Kumu software allows users to sort the data according to any of the sorting criteria discussed above. Figure 2 below illustrates the distribution of DOSTs according to research workflow steps. As can be seen from this figure, DOSTs actively contribute to all stages of research, but are particularly concentrated around analysis of data and publication.

Geographic distribution of tools and host organizations

The majority of DOSTs included in the database were explicitly connected to specific countries and regions. The geographic location of the DOST was available on the web pages through contact details, named host institutions or details of registration in the terms and conditions (T&Cs). Of these, 18 listed tools did not give a specific geographic location on their websites.

As can be seen from Figure 3 below, a high proportion of DOSTs available to the international research community are registered in (or linked to) the United States. It is therefore likely that the design and deployment of many of these

Figure 2. Visual map using the software Kumu.io. **A)** Clustering overview of all tools sorted by workflow step (url: <https://kumu.io/a2p/dost#dataset/workflow-step>); **B)** Clustering overview by geographical location of the tool or the respective host institution (url: <https://kumu.io/a2p/dost#dataset/workflow-step>); **C)** Clustering overview by host institution for the tool (url: <https://kumu.io/a2p/dost#dataset/host>); **D)** Focus view on hist self-hosted tools – closeup from square in **C**).

Figure 3. Number of tools per host location. Regions displayed are the United States of America (US), the European Union (EU), the United Kingdom (UK) and other parts of the world with concentration on US territory. 'Other' includes Argentina (n=1), Australia (n=2), Brazil (n=1), Canada (n=7), Colombia (n=1), Mexico (n=1), South Africa (n=1), Switzerland (n=5), with a total of n=242.

tools was influenced by the needs and preferences of high income countries (HICs) researchers. Of the tools linked to a specific country, the vast majority were connected to the United States, either as a registered non-profit organization (NPO 501(c)3), a registered commercial company or hosted by a US American institution such as a university or government body. Others were also hosted by parent organizations, such as the [Centre for Open Science](#), [Wikimedia Foundation](#) or [GitHub](#). The numerical distribution of the countries hosting DOSTs is demonstrated in [Figure 3](#).

Financial models

There was considerable heterogeneity in the financial models of the DOSTs within the database. In an attempt to simplify this heterogeneity, the tools were classified into the categories *Commercial*, *Grant*, *mixed* (commercial and grant), and *Institutional*. The distribution of these funding models is visualized in [Figure 4](#). Half of the DOSTs used a mixed model of funding, combining grants, commercial support, membership fees, freemium models, consulting or crowdsourcing.

The geographic location of the DOSTs ([Figure 3](#)) and the variations in funding ([Figure 4](#)) together highlight how the DOST ecosystem is governed by a complex network of financial legislation. NPOs and commercial entities are subject to the respective national legislation governing financial transactions. Similarly, if tools are hosted by an NPO, academic institution or governmental organization they are subject to the legislation governing the host organization.

The complexity of the underlying funding mechanisms has significant implications for the DOST ecosystem. In particular, it complicates efforts to make the DOST ecosystem transparent with regards to funding sources and legislative influence. It also impacts on the financial viability and longevity

of the DOSTs within the ecosystem. Indeed, the reliance of many DOSTs on crowdsourcing and time-limited grants means that many will struggle to achieve financial independence and sustainability.

Highly influential actors

We have documented selected interlinkages between the DOSTs in the database ([Bezuidenhout & Havemann, 2020](#)). From the analysis of the database it became apparent that certain entities are highly interlinked within the DOST ecosystem, such as GitHub, Center for Open Science and Digital Science. [Figure 5](#) below details 8 highly influential organizations within the DOST landscape, demonstrating how these organizations/institutions are linked to DOSTs operating throughout the research workflow.

As shown in [Figure 5](#), 80.9% of the DOSTs in the database are linked to one or more of these 8 entities. These interlinkages were diverse and included direct sponsorship, hosting of the DOST, or the hosting of DOST resources. These interlinkages can also be visualized in the [Kumu plot](#).

Variations in Terms & Conditions

Examination of the Terms and Conditions (T&Cs) of the DOSTs revealed a range of different factors that limited usage/accessibility or imposed liability on users. Strikingly, these T&C limitations were mainly found in DOSTs registered directly in the US, or sponsored by companies/organizations registered in the US, linked to US trade control laws, and thus restrict the services that can be made available to users in countries and territories under US sanctions. Two examples of companies that have explicitly clarified these limitations in their T&Cs are presented in [Table 2](#). It is important to note that the lack of explicit prohibition within the T&Cs of other companies does not necessarily indicate that they are available for

Figure 4. Illustration of the funding models of DOSTs. The funding sources for the respective tools were classified as a) Commercial (n=56, 23.1%); b) Grant (n=19, 7.9%); c) mixed (commercial and grant, n=122, 50.4%), and d) Institutional (n=44, 18.2%). 0.4% of the tools (n=1) had no funding source specified. n=242.

Figure 5. Tool providers across workflow showing the number of tools per workflow step.

access by researchers in US sanctioned countries. More research on the extent of geoblocking is urgently required to clarify these issues.

As many other DOSTs rely on GitHub for infrastructure and hosting of resources, the T&Cs of one commercial company can have far-reaching consequences for the Open Science

ecosystem. In contrast, there has been no systematic study to date examining whether access to the numerous preprint services hosted by the Center for Open Science are blocked in US-sanctioned countries. These T&Cs remain problematic as they place the responsibility on the user of the site to comply with the legislation alluded to. This raises challenges for users as they have to identify and read the relevant legislation – often

Table 2. Example T&Cs of two entities within the OS ecosystem.

DOST	Statement in T&Cs	Notes
GitHub	You may not use GitHub in violation of export control or sanctions laws of the United States or any other applicable jurisdiction. You may not use GitHub if you are or are working on behalf of a Specially Designated National (SDN) or a person subject to similar blocking or denied party prohibitions administered by a U.S. government agency. GitHub may allow persons in certain sanctioned countries or territories to access certain GitHub services pursuant to U.S. government authorizations. [...] To comply with U.S. trade control laws, GitHub recently made some required changes to the way we conduct our services. As U.S. trade controls laws evolve, we will continue to work with U.S. regulators about the extent to which we can offer free code collaboration services to developers in sanctioned markets. We believe that offering those free services supports U.S. foreign policy of encouraging the free flow of information and free speech in those markets ¹⁸ .	The countries affected are Crimea, Cuba, Iran, North Korea, and Syria. There have been reports of access to GitHub being blocked in these countries.
Center for Open Science	The COS is based in the United States. The COS makes no claims that the data or content on its Websites or Services is appropriate or may be downloaded outside of the United States. Access to the Websites and Services may not be legal by certain persons or in certain countries... You may not use the Websites or Services to violate any applicable local, state, national, or international law, including without limitation any applicable laws relating to antitrust or other illegal trade or business practices, federal and state securities laws, regulations promulgated by the U.S. Securities and Exchange Commission, any rules of any national or other securities exchange, and any U.S. laws, rules, and regulations governing the export and re-export of commodities or technical data ¹⁹ .	The T&Cs for the COS are hosted on GitHub, which makes access to the T&Cs from US-sanctioned countries difficult.

¹⁸ <https://help.github.com/en/github/site-policy/github-and-trade-controls> (accessed 16/03/2020)

¹⁹ <https://github.com/CenterForOpenScience/cos.io/pull/1025/files> (accessed 16 March 2020)

in English – and access the T&Cs when they are hosted on GitHub.

US-sanctioned were explicitly mentioned in 79 of the DOSTs in the database. While many of these did not explicitly state that their services were blocked to users in countries under US sanctions, this could still be the case. Indeed, there is considerable anecdotal and documented evidence of research tools and databases being geoblocked to users in countries under sanction from the US. These could include countries such as Sudan (Bezuidenhout *et al.*, 2019), Iran²⁰, Myanmar, North Korea, Venezuela, Cuba, Crimea and Zimbabwe.

Discussion

For the open science movement to progress and the DOST ecosystem to flourish, the evolving digital ecosystem must ensure that “the primary outputs of publicly funded research results – publications and the research data – [are] publicly accessible in digital format with no or minimal restriction” (OECD, 2015, p. 7). It also requires “extending the principles of openness to the whole research cycle, fostering sharing and collaboration as early as possible thus entailing a systemic change to the way science and research is done”.

An effective DOST ecosystem thus has two key roles: 1) to facilitate practices that enhance open and transparent research

as well as 2) to ensure that these practices – and the resultant resources – are available to researchers across the world. The analysis of the current DOST ecosystem presented above suggests that it may struggle to deliver on these roles. The unequal geographic distribution of the tools, the dominance of certain languages, cultures and entities, and the diversity of the funding models supporting the development of new tools all add complexities to the DOST ecosystem. Recognizing these power dynamics, value clashes, and infrastructural bottlenecks is essential for the future of the open science movement. In the section below, we discuss the results and their implications for open science in more detail.

Persistence and preferences

The current structure of the DOST ecosystem means that the persistence of individual tools depends on attracting a community of users and securing stable funding. This might suggest that these features support a meritocracy, whereby the “best” DOSTs persist by common consent and investment. Such a position, however, overlooks key issues such as diversity within user communities and accessibility of funding. Overlooking such issues can undermine the open science values described above – particularly the aspiration that the open science ecosystem be globally accessible and useful.

As illustrated in Figure 2, many of the presented DOSTs are hosted in the United States. It is therefore likely that these tools have been piloted and beta-tested within the immediate research communities and therefore many of the design decisions integrated into the DOSTs dominantly reflect the US research

²⁰ <https://github.com/pi0/github-is-blocked-in-iran> (accessed 16 March 2020)

environment and the preferences of the researchers in this region. Similarly, DOSTs created by commercial companies, or designed with commercialization in mind, will likely reflect the most immediate user community, namely North-American and European researchers.

While these biases could be eliminated by subsequent user-community feedback, this is not always feasible. Limited funds for long-term responsive design, and slow roll-out beyond the US and other High-Income Countries (HICs) along with the unequal distribution of researchers around the world can mean user communities develop around DOSTs before they have had any meaningful engagement from researchers working outside of these “geographical epicentres”. For example, in 2013 the Europe Union (11.4% of the global population) hosted 31% of the world’s researchers²¹.

This can mean that voices from other research communities can easily be overlooked - including non-English speaking countries or low- and middle-income countries (LMICs) with small research communities. These different research communities have to date played a marginal role in the evolution of the DOST ecosystem due to the low level of involvement in elucidation and design of tools and infrastructure. If DOSTs are designed with a specific research community in mind and tested in the same community, it can mean that the design of the DOSTs “closes” before these marginal research communities are able to engage with them (Bijker *et al.*, 2012). This “technological closure” means that the DOSTs available for use by marginal research communities will already have a fixed design and dedicated user community. Fundamental design decisions are unlikely to alter once the DOST has become operational in the original user community. Consequently, certain tools may be integrated into the DOST ecosystem that do not suit use in non-HIC research contexts. As a result, it is possible that certain communities get “locked-in” to the use of these DOSTs without having the opportunity to feed back into design decisions (Arthur, 1989; Leonelli, 2016).

Situations of “locking” research communities into certain DOSTs and digital workflows can cause the ecosystem to unintentionally perpetuate marginalizations. The design and persistence of the DOSTs not only influence the “pathways” that the research follows through the ecosystem, but also the research methods, data collection and curation methods and analysis tools used. The selection of certain tools over others can thus have far-reaching implications. The decisions incorporated into its design reflect a specific geographic context and value system can influence research practices across the globe.

Such concerns relate to the “Juan Valdez problem” discussed by Busch and Juska, (Busch & Juska, 1997) in relation to agricultural systems and technologies. Juan Valdez, a South American coffee farmer, is born into a world in which his choices are limited. Many of these limitations relate to the environment he lives in, and which he accepts as default. On the other hand, certain choices may be deliberately denied to him. “The coffee company may have a local monopoly over purchasing the beans. The state may not have invested in adequate physical infrastructure for the area, thereby making transportation costs high” (Busch & Juska, 1997, p. 696). Thus, what is possible for Juan is dictated by human and non-human relationships alike. Similarly for open science tools, what is possible for marginal research communities may be determined less by their preferences than by decisions made between human partners in geographically remote locations.

It becomes apparent that more research is urgently needed. Qualitative research on the development of DOSTs would shed light on how potential design biases are addressed during the design of these tools. More information on the (lack of) diversity within user communities would highlight issues of “lock in”, while engagement with LIMC researchers about the use of existing DOSTs would provide further information on the usability of these tools in non-HIC research settings.

Power brokers in the DOST ecosystem

From Figure 4 above it is evident that the DOST ecosystem is dominated not only by certain countries, but also by certain companies, organizations and institutions. Such clustering - in light of funding, access to target audiences, permissive legislation and business cultures - is not particularly surprising. Indeed, it may be said to follow other models of technical expansion throughout history. Accepting this expansion as entirely normal from the user perspective, however, does not make it unproblematic.

The DOST ecosystem and the DOSTs themselves are intended to be distributed and multiplicitous to allow the maximal flexibility of research practices. Allowing a small number of entities to dominate the ecosystem and its evolution thus presents challenges to these aims. In particular, two key concerns arise: first, the dominance of certain entities causes centralization and interdependence on individual actors. Second, the dominance of certain entities allows specific approaches to open science, and related values, practices and preferences to be prioritized. This can affect the heterogeneity of the open science movement and foster a perception that there is consensus on how open science “should be done” (Fecher & Friesike, 2014).

In recognizing the former, the DOST ecosystem must confront a paradox. While interconnectedness is vital for fostering open, global research and removing national, disciplinary, and linguistic siloes, the same tools that facilitate this connectedness can lead to a centralism that drives out regional and local expertise and diversity. In particular, having tools such as GitHub dominate various stages of the research lifecycle

²¹ The Big Five (China, European Union, Japan, Russian Federation and USA) still account for 72% of researchers worldwide but the share of China has progressed considerably since 2009, to the detriment of Japan, the Russian Federation and the USA. The share of the European Union (7.1% of the global population) has remained stable, at 22.2% in 2013, compared to 22.5% in 2009. Europe as a whole (11.4% of the global population) hosts 31% of the world’s researchers. <https://en.unesco.org/node/252277> (accessed 17 March 2020)

in a number of tools not only enhances interoperability, but also centralization and dependence, thereby diminishing accessibility to some.

The latter concern relates to an often-overlooked aspect of technology: The intentions, experiences, priorities and cultures of the IT-professionals influence the design and deployment of the technology (Winner, 1986). All DOSTs are created against a backdrop of social values, and designed with specific interpretations of open science in mind. This can lead to considerable heterogeneity in what is foregrounded, prioritized and included in the design of the DOSTs. As a result, DOSTs, like other technologies, are at once both the sites and objects of politics (Jasanoff & Kim, 2009, p. 126), and foreground certain views of openness through their positioning in the DOST ecosystem.

GitHub, for example, is a commercial company based in the US. The design of GitHub, and its operating practices thus align to a specific set of values. As a result of its dominance within the DOST ecosystem, its position on key issues such as inclusion, sharing and transparency are increasingly becoming the “norm” for many users despite its political constraints and accessibility restrictions for many researchers. Recognizing such issues highlights the need for closer scrutiny of the value structures of the tools within the DOST ecosystem. Asking questions such as why tools were created, how users were recruited and why they favour one tool over another will shed light on these issues. In particular, it will highlight the limitations of allowing certain countries, tools and organizations to dominate the DOST ecosystem.

Access and underlying infrastructures

The decisions influencing the design of DOSTs do not only reflect user community preferences and perspectives of open science, but also assumptions about the availability of infrastructures and resources. These include a wide range of different issues, including access to funding and the ability to make online payments, linguistic competence, access to software and hardware, as well as infrastructure availability relating to internet connectivity and bandwidth.

For many DOSTs developed in Europe or the US there is an emphasis on the tools being cloud-based. On the one hand, such an emphasis makes sense in many ways such as the ease of having nothing to install, being able to deliver the latest version of software via the browser and having access to the content from any device anywhere in the world, as long as it is connected to the internet. On the other hand, some institutions especially in the European Union prefer the tools to operate on their own servers to keep them confidential from potential competitors for patenting and to ensure data and content ownership through territorial storage. For research communities in LMICs these same design decisions form a usage barrier because of low bandwidth and intermittent internet connection that make an over-reliance on “online only” tools problematic (Bezuidenhout *et al.*, 2016).

While multiplicity in the DOST landscape can allow marginal researchers to plot alternative pathways through the OS ecosystem, this can mean that they must resort to using less popular tools. As a result, there is a chance that these researchers continue to be excluded from the user communities that are driving research forward. This has obvious implications for collaborations, visibility/engagement with researcher communities and perceptions of worth.

Designing DOSTs for infrastructure present in the dominant geographical regions (such as the US) legitimizes a specific expectation of service access and provision. In this way, the DOST ecosystem fails to address the recognized imbalance between central and marginalized countries and research communities. Indeed, the cost for internet access and [institutional as well as private] connectivity varies drastically across world regions and tends to be extraordinarily high in LMICs²². By perpetuating a set of embedded assumptions like web interfaces or connectivity, open science continues to perpetuate a limited perspective for “inclusion” that often falls short of being inclusive. Ensuring more inclusive design structures and processes will require ethnically and regionally diverse teams of DOST designers to ensure that infrastructural challenges are considered and responses incorporated into design decisions.

Sanctions and political clout

As demonstrated in the results section, the DOST ecosystem has to contend with a range of power dynamics external to research infrastructure. Perhaps the most pernicious of these is the role that financial legislation plays in dictating access to open resources (Bezuidenhout *et al.*, 2019). This is perhaps best demonstrated by the impact that US financial sanctions have on access to DOSTs. As demonstrated by Table 2 a number of DOSTs explicitly prohibit use from individuals located in countries currently under financial sanctions from the US.

The reasons for these prohibitions are complex and often relate to the financial requirements of the funding bodies. DOSTs developed by commercial companies registered in the US, or those funded by commercial companies registered in the US, are subject to US tax law that explicitly prohibits transacting with countries under sanction. As a result, the values and political positions of the US government are integrated into the open science landscape via a range of different tools. From the data available, it was not possible to determine whether US organizations registered as NPO 501(c)3 or receiving fiscal sponsorship would be similarly subject to restrictions. Nonetheless, the limitations elucidated in the T&Cs represented in Table 2 suggest that this issue requires considerable further examination.

In addition to the explicit restrictions noted on T&Cs, users in countries under sanction from the US may be restricted access

²² <https://www.visualcapitalist.com/cost-of-mobile-data-worldwide/>

via three additional pathways. Many of the DOSTs in the database required some form of account or login (see Figure 4). This implied that the location of the users is being monitored via the tool and could provide a means to deny certain users access to the services. Additionally, the DOSTs requiring some kind of payment for services - either freemium or membership fees - could restrict access from countries under sanction, as online financial transactions are largely prohibited from these countries. Third, governance of certain elements of the Open Science landscape by high-level but poorly elucidated legislation - such as cryptography software by the Wassenaar agreement²³ - can mean that providers restrict access as a means of precaution. Expanding analytical services such as AlternativeTo and Terms of Service Didn't Read²⁴ to DOSTs will help researchers make informed decisions as they navigate through the open science ecosystem.

This creates situations of marginalization and lack of access for certain communities of end-users. Even more concerning, however, is that one country's political preferences are able to dictate the evolution of aspects of the DOST ecosystem. While it is important to note that the introduction of these political values is likely done unintentionally or via funding-related necessity, the impact is nonetheless severe. Acknowledging that certain aspects of the DOST ecosystem are unavailable to certain communities of users is vital for further critical reflection on the evolution of open science. In particular, what does this mean for the core values of the open science movement and the notion of a "digital commons" (Bezuidenhout, 2020; Hess & Ostrom, 2007)?

A critical appraisal of the DOST ecosystem

The results and discussion presented in this paper draw attention to problems within the current DOST ecosystem. Without detracting from the importance of the emergence of more and more discipline- and region-specific DOSTs, and the work of dedicated individuals who create them, words of caution are appropriate. The results of this paper demonstrate the heterogeneity of the actors, power dynamics and stakeholders that are currently driving and dominating the evolution of the DOST ecosystem. Even if all DOSTs were created by well-meaning individuals who wish to promote open science, one cannot simply assume that the resultant ecosystem will automatically reflect and perpetuate the core values of open science. Instead, a range of different factors inherent within DOST design create a landscape that continues to perpetuate marginalization and exclusion.

This marginalization is multifaceted. Not only are marginal research communities excluded from design decisions of DOSTs, they are likely also sidelined in the user communities that develop around them. Moreover, DOST (un)availability/accessibility does more than exclude researchers from sharing

communities, it also dictates research practices and digital workflows. In this way, the design of the DOST ecosystem can affect both present and future research. While the DOST ecosystem is dynamic and multiplicitous, the dominance of a few entities is rapidly driving forward a "status quo" of how research should be done. Once such practices reach a "carrying capacity" within the global research community, they are unlikely to be easily adapted. This can mean that the current design of the DOST ecosystem marginalizes future, as well as present, researchers.

The results and discussion in this paper point towards the need for a new model to critically evaluate the evolving DOST ecosystem. In particular, it highlights the need for more active inclusion of diverse user communities in all stages of DOST development and deployment²⁵. This will make the embedded politics of the DOSTs ecosystem more transparent. Conversely, there is an imperative to identify examples of DOSTs developed in, for and by researchers in Africa, Asia and Latin America which can serve as examples of alternative design practices. This will provide a better understanding of how diversity can be better supported in the DOST ecosystem. This will allow critical reflection on the politics that are not visible in centrally-located tools that are being made explicit in the non-central ones.

Many of the issues mentioned and concerns raised in this paper will not come as a surprise to open science practitioners. Nor will it be surprising to add that the current model of persistent barriers continues to place certain members of the open science community in uncomfortable and sometimes unethical positions. These include having the choice of open science tool dictated to them through lack of engagement in community consensus or due to feasibility in a local context with digital infrastructure deficiencies. It also includes having to operate in an ecosystem that regularly requires the decision making between non-participation or breaking law by consulting scholarly pirate software.

Allowing such situations to persist undermines the aims of the open science movement. Recognizing this places a responsibility on the global open science community members to make discerning decisions about the tools that they use. This requires that the T&Cs of DOSTs, their funding structures and their infrastructural constituencies are all closely scrutinized before new tools become embedded in the DOST ecosystem. Similarly, funders, research institutions and other stakeholders need to critically assess the impact of introducing DOSTs to the ecosystem, and advocating their use amongst their researchers (ie. through the San Francisco Declaration on Research Assessment).

The section above highlighted how inequalities, marginalization and injustices were perpetrated by the current structure

²³ <https://www.wassenaar.org/> (accessed 17 June 2020)

²⁴ <https://alternativeto.net/>, <https://tosdr.org/>

²⁵ Key resources such as the Open Science Grassroots Community Networks listing by CoS will provide valuable further evidence for inclusion https://twitter.com/Gen_R/status/1146069028546523136?s=20

of the DOST ecosystem. The design of DOSTs, the ways in which they are interlinked, and the dependencies/dominances of certain entities raises the question of whether the DOST ecosystem can realise the aspiration of becoming a truly “unlimited digital commons” in its current structure. From the data presented above, it would seem that things need to change.

Nonetheless, the DOST ecosystem is a complicated landscape, and imposing a specific value set or “way of doing things” will harm the richness and diversity of this rapidly evolving field. Rather than imposing restrictions on what should constitute a DOST, we suggest that those designers and users be supported to critically reflect on the values that they are introducing into the ecosystem. There are many models currently in use on how to balance well-intentioned innovation with pragmatic requirements, and these need to be more strongly developed for DOSTs.

One such model, responsible research and innovation (RRI), has made considerable contributions to discourse around socially responsible innovation. Opening up access to data and support of open science are fundamental components of the RRI model (Stilgoe *et al.*, 2013). To date, little has been done to turn the RRI lens back on the open science movement that it evolved from to ask what an RRI model for Open Science tools could look like. Such a model needs to address questions such as how to foster a free and open “ecosystem” when the OS tools are generated by a diversity of actors - NPO, NGO, governmental, commercial, volunteer) that can hold highly divergent values while supporting open science. Similarly, how a free and open landscape can be created when financial and governmental regulations and requirements influence tool design needs to be looked at as well; a promising assessment is currently underway by the *Invest in Open Infrastructure* initiative.

It is important to note that community-determined standards for what constitutes “Open Science” already exist in a number of different areas. Within open access publishing, for example, both *ROMEO Sherpa* and the Directory of *Open Access Journals* (DOAJ) clearly define what is required of a publication to be open access. Similarly, *re3data* has developed a list of criteria that any open repository needs to demonstrate. Such community standards have been highly influential and are being widely adopted by research communities and provide for cross-regional and cross-disciplinary agreement and functionality. While conversations about open science tool standards have existed for more than a decade, the broader community needs to be engaged for such standards to become a reality.

The design of the DOST ecosystem not only determines how research is conducted today, but also determines the directions and practices of future research. Allowing certain actors, pathways or regions to become too entrenched will allow inequality and marginalizations to persist and become a future norm. Research practices are changing rapidly (ie. AI, big data), international politics are in flux (ie. Brexit, COVID-19 pandemic) and historically marginalized research communities (ie. citizen scientists and LMIC researchers) are increasingly

vocal and influential (Aspesi & Brand, 2020). It is now the right time to critically assess what has already been built, and what the united global research community wants to take forward into the future.

Concluding comments

Much of the OS ecosystem has been developed by volunteers, who donate time and expertise to developing DOSTs, infrastructures and interoperable practices. This community has the history, expertise and perspectives to take up the challenges raised in this paper. How, they need to ask, can they guide and adapt the ecosystem that is rapidly changing research? This requires a reframing of open science responsibilities, from contributing labour and data to discussing the complex power dynamics underpinning the evolving ecosystem. Only then will the UNESCO theme 2019 of “Open Science: leaving no one behind” become a reality.

The OS landscape is ever increasing globally, also in historically underrepresented regions such as Latin America, Africa and Asia. We therefore suggest to tie the digital development and regional adaptation of DOSTs on the Open Science Manifesto, Towards an Inclusive *Open Science for Social and Environmental Well-being*²⁶. In particular for the more dominant digital tools for open research and communities in Europe and North America, there is a dire need for more active consultation and inclusion of research stakeholders from various parts of the world in order to successfully design a truly global open science community, culture and infrastructure (Albornoz *et al.*, 2018)²⁷. Moreover, key expertise from development networks, such as *ICT4Dev* and *Tech4Dev* (Hostettler *et al.*, 2018) can play an important part in developing a more equitable open science ecosystem.

For the moment, however, building a body of evidence detailing DOSTs, their uses and the communities they use them is vital. Only through gathering this evidence can strategic and informed decisions about future ecosystem investments be made inclusively.

Data availability

Underlying data

Zenodo: The Varying Openness of Digital Open Science Tools. <http://doi.org/10.5281/zenodo.4013812> (Bezuidenhout & Havemann, 2020)

This project contains the following underlying data:

- DOST dataset 3 September 2020.xlsx (Full table of DOST information organized according to the categories described in the methods, together with hyperlinks to homepages)

Data are available under the terms of the *Creative Commons Attribution 4.0 International license* (CC-BY 4.0).

²⁶ <https://ocsdnet.org/manifesto/open-science-manifesto/>

²⁷ <https://en.unesco.org/science-sustainable-future/open-science>

References

- Albornoz D, Huan M, Marti I, *et al.*: **Framing Power: Tracing Key Discourses in Open Science Policies.** *Open Science Policies*. ELPUB. 2018.
[Publisher Full Text](#)
- Arthur WB: **Competing technologies, increasing returns, and lock-in by historical events.** *The Economic Journal*. 1989; **99**(394): 116–131.
[Publisher Full Text](#)
- Aspesi C, Brand A: **In pursuit of open science, open access is not enough.** *Science*. 2020; **368**(6491): 574–577.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bartling S, Friesike S: **Opening Science.** Heidelberg. 2014.
[Publisher Full Text](#)
- Bezuidenhout L, Rappert B, Leonelli S, *et al.*: **Beyond the Digital Divide: Sharing Research Data across Developing and Developed Countries.** 2016.
[Publisher Full Text](#)
- Bezuidenhout L, Karrar O, Lezaun J, *et al.*: **Economic sanctions and academia: Overlooked impact and long-term consequences.** *PLoS One*. Public Library of Science (PLoS). 2019; **14**(10): e0222669.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bezuidenhout L: **Being Fair about the Design of FAIR Data Standards.** 2020.
[Reference Source](#)
- Bezuidenhout L, Havemann J: **The Varying Openness of Digital Open Science Tools: dataset.** 2020.
<http://www.doi.org/10.5281/ZENODO.4013812>
- Bijker WE, Hughes TP, Pinch TJ: **The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology.** Cambridge Massachusetts: MIT Press. 2012.
[Reference Source](#)
- Busch L, Juska A: **Beyond Political Economy: Actor Networks and the Globalization of Agriculture.** *Source: Review of International Political Economy*. Winter. 1997; **4**(4): 688–708.
[Reference Source](#)
- Fecher B, Friesike S: **Open Science: One Term, Five Schools of Thought.** In: *Opening Science*. Springer International Publishing: 2014; 17–47.
[Publisher Full Text](#)
- Ferguson L: **How and why researchers share data (and why they don't) Exchanges.** (Accessed: 29 April 2015). 2014.
[Reference Source](#)
- Hess C, Ostrom E: **Introduction: An Overview of the Knowledge Commons.** In: Hess, C. and Ostrom, E. (eds). *Understanding Knowledge as a Commons: From Theory to Practice*. Boston: MIT Press: 2007; 3–26.
[Reference Source](#)
- Hostettler S, Besson SN, Bolay J: **Technologies for Development.** *Technologies for Development*. Dordrecht: Springer International Publishing. 2018.
[Publisher Full Text](#)
- Jasanoff S, Kim SH: **Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea.** *Minerva*. 2009; **47**: 119–146.
[Publisher Full Text](#)
- Kitchin R, Lauriault TP: **Towards critical data studies: Charting and unpacking data assemblages and their work.** In: Eckert, J., Shears, A., and Thatcher, J. (eds). *Geoweb and Big Data*. Lincoln: University of Nebraska. (Accessed: 18 May 2019). 2014.
[Reference Source](#)
- Kramer B, Bosman J: **Innovations in scholarly communication - global survey on research tool usage [version 1; peer review: 2 approved].** *F1000Res*. 2016; **5**: 692.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Leonelli S: **Locating ethics in data science: responsibility and accountability in global and distributed knowledge production systems.** *Philos Trans A Math Phys Eng Sci*. 2016; **374**(2083): 20160122.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Levin N, Leonelli S, Weckowska D, *et al.*: **How Do Scientists Define Openness? Exploring the Relationship Between Open Science Policies and Research Practice.** *Bull Sci Technol Soc*. SAGE PublicationsSage CA: Los Angeles, CA. 2016; **36**(2): 128–141.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chagas AM, Molloy JC, Prieto-Godino LL, *et al.*: **Leveraging open hardware to alleviate the burden of COVID-19 on global health systems.** *PLoS Biol*. Public Library of Science. 2020; **18**(4): e3000730.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mody C: **Instrumental community: probe microscopy and the path to nanotechnology.** Cambridge Massachusetts: MIT Press. 2011.
[Reference Source](#)
- OECD: **Bridging the Digital Divide.** 2015. (Accessed: 13 August 2015).
[Reference Source](#)
- OECD: **Why open science is critical to combatting COVID-19.** Paris. 2020. (Accessed: 17 June 2020).
[Reference Source](#)
- Powell A: **Democratizing production through open source knowledge: from open software to open hardware.** *Media, Culture Society*. SAGE PublicationsSage UK: London, England. 2012; **34**(6): 691–708.
[Publisher Full Text](#)
- Rahal RM, Havemann J: **Science in Crisis. Is Open Science the Solution?** *Forum Wissenschaft*. MetaArXiv. 2019.
[Publisher Full Text](#)
- Stilgoe J, Owen Macnaghten R: **Developing a framework for responsible innovation.** *Res Policy*. 2013; **42**(9): 1568–1580.
[Publisher Full Text](#)
- Tennant J, Beamer J, Bosman J, *et al.*: **Foundations for Open Scholarship Strategy Development.** MetaArXiv. 2019.
[Publisher Full Text](#)
- Tennant J, Agarwal R, Baždarić K, *et al.*: **A tale of two 'opens': intersections between Free and Open Source Software and Open Scholarship.** *SocArXiv Papers*. 2020. (Accessed: 17 June 2020).
[Reference Source](#)
- Vermeir K, Leonelli S, Tariq ASB, *et al.*: **Global Access to Research Software: The Forgotten Pillar of Open Science Implementation.** 2018. (Accessed: 28 September 2018).
[Reference Source](#)
- Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al.*: **The FAIR Guiding Principles for scientific data management and stewardship.** *Sci Data*. Nature Publishing Group. 2016; **3**: 160018.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Winner L: **Do Artifacts Have Politics?** *Daedalus*. 1986; **109**(1): 121–136.
[Reference Source](#)
- Zastrow M: **Open science takes on the coronavirus pandemic.** *Nature*. Springer Science and Business Media LLC. 2020; **581**(7806): 109–110.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ✖

Version 1

Reviewer Report 07 December 2020

<https://doi.org/10.5256/f1000research.29380.r74156>

© 2020 Hartgerink C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chris H. J. Hartgerink

Liberate Science GmbH, Berlin, Germany

The authors present “a methodological attempt to map a selection of the DOST [digital open science tools] ecosystem including links between the tools” (p. 6) to answer three research questions revolving the (1) geographic impact of DOSTs, (2) whether funding/values/stakeholders impact the openness of DOSTs, and (3) whether external power dynamics are recognized and addressed in DOSTs (p. 5).

I think the authors present a valuable open dataset and research questions, although the manuscript does not answer the questions set out by the authors themselves.

As a research paper, I noticed the research paper is heavily weighted towards a perspective piece. As the core function of the manuscript is described to answer the three aforementioned research questions, 60 paragraphs for the introduction and discussion, in contrast to 26 paragraphs for the methods and results sections, seems unbalanced. It seems like the authors are trying to make a lot of points in this piece above and beyond the research questions, and I wonder whether those parts of the analysis make sense in this research paper in its current form. To me personally, it also distracted from understanding the manuscript's core points.

Starting out with the self-described goal of the paper on p.6 and mentioned at the start of this review, I do not see the description of how links between the various DOSTs are determined, or what the results of these links were. Would the authors be able to shed more light on this? As I inspected the open data I was able to find some links (column "Reliant on other DOSTs") but these are not presented in the paper.

Regarding RQ1 (impact of geographical origin of DOSTs), I think the authors make a valuable point and that such an investigation is worthwhile. The results show there is inequity in where the DOSTs originate from. The authors do not clarify how they operationalize and assess impact, which makes it difficult to assess the results in that light. Implicitly, it assumes usage numbers or other indicators are irrelevant to assess the impact. Impact is a non-trivial thing to measure, and without a definition, I would argue it is increasingly difficult to even answer RQ1. Similarly, the

impact on what is left undefined in RQ1.

For RQ2, the impact of funding/values/stakeholders on openness is left undefined. In the results, descriptive statistics are given for the funding and providers, without dependent measures where differences could be measured. What is the openness that the funding/values/stakeholders would have an impact on? Again, an operationalization is not included. Additionally, the mentioned values are not included in the methods or results, hence, it might've been omitted in the reporting or it is superfluous in RQ2.

For RQ3, whether external power dynamics are recognized and addressed in DOSTs, external power dynamics seem to be operationalized as geoblocking due to trade sanctions. This was not clear from the introduction. If the authors want to focus on trade sanctions, which is a valuable perspective, I would suggest limiting the RQ to this instance and clarify that external power dynamics are operationalized as such. Speculative remarks such "While many of these did not explicitly state that their services were blocked to users in countries under US sanctions, this could still be the case."

Whether the DOSTs address these external power dynamics (second part of RQ3) is left unanswered and comes across as a difficult question to begin with. How can DOSTs address (eg) trade sanctions? Maybe if the DOST is run by a multibillion dollar corporation, they could lobby for this, but I fail to understand how any smaller organization could address such external power dynamics? Maybe I am missing the point, and it refers more to other ideas the authors have but are not included at the moment. I would be interested to read those ideas.

All in all, I think the authors investigate an incredibly valuable set of research questions and that their methods limit their findings. I would recommend the authors to focus on the research questions and add operationalizations to provide a clearer indication of what their results indicate. The answers I currently get from the manuscript for their original three research questions are:

1. Many DOSTs originate from a few countries/regions. May lead to neglecting needs of researchers outside those countries/regions (no evidence provided).
2. Not answered.
3. Provides insight regarding superstructures of how the origin country of a DOST is affected by trade sanctions, and data on how many DOSTs recognize this in their Terms of Use (72/242).

I hope that the authors can address these remarks, because in its current for the manuscript does not seem to achieve the goals it sets out for itself. I think there is a lot of value in their work.

Tangential remarks:

- Funding models are categorized in this study, and for mixed funding it is all seen as one and the same, despite potential variation in the degree of how mixed the funding is (e.g., 9:1 grant funding to own revenue is denoted as mixed, just like 1:9). This is now an implicit assumption and an implicit limitation that might be worth mentioning.
- Would it be possible to include some information on how much disagreement there was in the coding? That would be informative to understand how easy it was to code all of these

data

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: I know the second author personally and as a co-author on a preprint: <https://digitalcommons.du.edu/collaborativelibrarianship/vol11/iss2/2/>, <https://doi.org/10.31222/osf.io/et8ak>. I think I was still able to be impartial, but ultimately I cannot decide this for the reader. As you may see in my review, I am critical of this work and I am not sparing any criticism that is reasonable.

Reviewer Expertise: meta-research, statistics, methodology, library and information sciences, psychology,

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 09 Mar 2021

Louise Bezuidenhout

Reviewer 2

The authors present "a methodological attempt to map a selection of the DOST [digital open science tools] ecosystem including links between the tools" (p. 6) to answer three research questions revolving the (1) geographic impact of DOSTs, (2) whether funding/values/stakeholders impact the openness of DOSTs, and (3) whether external power dynamics are recognized and addressed in DOSTs (p. 5).

With respect to the reviewer, these research aims are slightly mis-stated. The object of the paper - and the study - was not to investigate the openness of specific DOSTs, but rather to speculate as to whether issues such as geographic location, funding source or legal restrictions might have an

impact on the overall openness of the DOST ecosystem, and thus on the Open Science landscape. We do not try to assess or evaluate the openness of individual DOSTs. In order to make this more apparent we have included additional text in the introduction and methods sections.

I think the authors present a valuable open dataset and research questions, although the manuscript does not answer the questions set out by the authors themselves.

As a research paper, I noticed the research paper is heavily weighted towards a perspective piece. As the core function of the manuscript is described to answer the three aforementioned research questions, 60 paragraphs for the introduction and discussion, in contrast to 26 paragraphs for the methods and results sections, seems unbalanced. It seems like the authors are trying to make a lot of points in this piece above and beyond the research questions, and I wonder whether those parts of the analysis make sense in this research paper in its current form. To me personally, it also distracted from understanding the manuscript's core points.

As stated above, the objective of this paper is to highlight characteristics of the DOST ecosystem that require further discussion and investigation. In no way does this paper attempt to provide an exhaustive analysis of the limitations of each individual DOST. Rather, this paper attempts - unlike many other papers that look at individual digital tools applied in the research workflow and across disciplines - to look at DOSTs as an interlinked ecosystem.

The categories in the dataset were populated from the publicly available information on the DOST websites. Thus, due to the variation in websites and documentation, some of the categories are not complete. All information about funding, geographic location and links to other DOSTs were obtained in this manner.

As discussed above, the methods section has been extended.

Starting out with the self-described goal of the paper on p.6 and mentioned at the start of this review, I do not see the description of how links between the various DOSTs are determined, or what the results of these links were. Would the authors be able to shed more light on this? As I inspected the open data I was able to find some links (column "Reliant on other DOSTs") but these are not presented in the paper.

As above, and similarly relating to the comments below, assessing individual impact of DOSTs is beyond the scope of this paper.

Regarding RQ1 (impact of geographical origin of DOSTs), I think the authors make a valuable point and that such an investigation is worthwhile. The results show there is inequity in where the DOSTs originate from. The authors do not clarify how they operationalize and assess impact, which makes it difficult to assess the results in that light. Implicitly, it assumes usage numbers or other indicators are irrelevant to assess the impact. Impact is a non-trivial thing to measure, and without a definition, I would argue it is increasingly difficult to even answer RQ1. Similarly, the impact on what is left undefined in RQ1.

Please see text on p28:

It is important to recognise that the power dynamics shaping the DOST ecosystem are likely influenced by multifarious pressures. These include community decisions regarding the quality of the DOSTs, the unequal distribution of DOST users and researchers/developers around the globe, economic issues including the opportunities for alternative careers in DOST development in developed countries. The limits of this study make it impossible to engage with the influence of these different pressures in detail. Rather, the study seeks to highlight not only the complexities of the DOST ecosystem, but also the critical need for more scholarship on its evolution. This paper highlights how these pressures cause the dominance of certain DOSTs to be a “bug” in the current structuring of the ecosystem, rather than as an inevitable outcome of the status quo of scientific research.

For RQ2, the impact of funding/values/stakeholders on openness is left undefined. In the results, descriptive statistics are given for the funding and providers, without dependent measures where differences could be measured. What is the openness that the funding/values/stakeholders would have an impact on? Again, an operationalization is not included. Additionally, the mentioned values are not included in the methods or results, hence, it might've been omitted in the reporting or it is superfluous in RQ2.

While assessing the openness and limitations of funders would be a useful activity, it was out of the scope of this paper. What this category aimed to highlight was the diversity of the funding within the DOST landscape, rather than attempting to quantify it in any way. The following text has been added to P16:

Because this study was a desktop review, it was not possible to gather specific financial details about each DOST. The information included was publicly-available funding reporting on the DOST websites.

We do mention that financial backing of a tool has a direct effect on its sustainability/longevity in our analysis.

For RQ3, whether external power dynamics are recognized and addressed in DOSTs, external power dynamics seem to be operationalized as geoblocking due to trade sanctions. This was not clear from the introduction. If the authors want to focus on trade sanctions, which is a valuable perspective, I would suggest limiting the RQ to this instance and clarify that external power dynamics are operationalized as such. Speculative remarks such “While many of these did not explicitly state that their services were blocked to users in countries under US sanctions, this could still be the case.”

Whether the DOSTs address these external power dynamics (second part of RQ3) is left unanswered and comes across as a difficult question to begin with. How can DOSTs address (eg) trade sanctions? Maybe if the DOST is run by a multibillion dollar corporation, they could lobby for this, but I fail to understand how any smaller organization could address such external power dynamics? Maybe I am missing the point, and it refers more to other ideas the authors have but are not included at the moment. I would be interested to read those ideas.

All in all, I think the authors investigate an incredibly valuable set of research questions and

that their methods limit their findings. I would recommend the authors to focus on the research questions and add operationalizations to provide a clearer indication of what their results indicate. The answers I currently get from the manuscript for their original three research questions are:

1. Many DOSTs originate from a few countries/regions. May lead to neglecting needs of researchers outside those countries/regions (no evidence provided).
2. Not answered.
3. Provides insight regarding superstructures of how the origin country of a DOST is affected by trade sanctions, and data on how many DOSTs recognize this in their Terms of Use (72/242).

As highlighted above, we feel that the reviewer has perhaps mistaken the objectives of the paper and the source of the data used for this analysis. We present this paper as a novel perspective on DOSTs that takes into consideration their geographic, online and financial interlinkages. We present this analysis and speculative discussion based on our personal experience from working in global south contexts and being in close collaboration with researchers living and working in those regions as well as assessment of publicly-available information on DOSTs, not on in-depth interviews or data mining. We aim for this paper to highlight areas within the Open Science landscape that require further scrutiny and research, rather than per se offering a definitive analysis of the pitfalls of the highly complex and diverse DOST ecosystem.

I hope that the authors can address these remarks, because in its current form the manuscript does not seem to achieve the goals it sets out for itself. I think there is a lot of value in their work.

Tangential remarks:

- Funding models are categorized in this study, and for mixed funding it is all seen as one and the same, despite potential variation in the degree of how mixed the funding is (e.g., 9:1 grant funding to own revenue is denoted as mixed, just like 1:9). This is now an implicit assumption and an implicit limitation that might be worth mentioning.

Further studies on the funding models of DOSTs are undoubtedly needed. However, as this paper was based on publicly available knowledge from DOST websites, being able to ratio the contributions of various donors was not possible and in such detail also not the focus of our work. We deem any mixed funding approach as preferable to reliance on a single funding /revenue source.

- Would it be possible to include some information on how much disagreement there was in the coding? That would be informative to understand how easy it was to code all of these data

The methods section has been expanded and includes the explanation on p13: Discrepancies were discussed until consensus was reached.

Competing Interests: No competing interests were disclosed.

Reviewer Report 27 November 2020

<https://doi.org/10.5256/f1000research.29380.r74157>

© 2020 Brembs B. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Björn Brembs

Institut für Zoologie-Neurogenetik, Universität Regensburg, Regensburg, Germany

This is an amazing piece of work, one of its kind (hopefully the first of its kind) in breadth and scope. It is highly timely and dearly needed. I also fully support the social agenda the authors are promoting in this work and my suggestions below are aimed towards strengthening their important arguments. To achieve change, arguments for change need to be convincing. The goals and direction of the authors' arguments are clear and I support them. However, the way they are articulated here, weakens their power and makes them too easy to dismiss. Below my suggestions of making the arguments more powerful.

One of the main conclusions from the work here is that DOSTs are heavily biased towards scholars in rich countries. While this is an expected outcome, a more critical reader could counter that this may be purely a result of the bias of the authors: selection bias confirming the perspective the authors had when they started to work. To counter this argument, the authors need to explain in great detail, how the database they analyzed was derived. For instance, the Kramer and Bosman database consists of currently 696 DOSTs, but the one presented here only contains 242 DOSTs. The authors need to explicitly state, in sufficient detail, how the 242 were selected from the 696 in order to ensure that "selection bias" cannot be used against them. For instance, a critical reader may claim that the remaining, un-analyzed 454 DOSTs may all have been from LMICs and have been left out in order to score political points. More perniciously, a more critical reader may claim that LMIC are not represented with many DOSTs because they either lack the ability to design DOSTs or that the designed DOSTs from LMICs are so bad, nobody would ever want to use them, so they never make it into any database. Such alternative (and analogous) explanations for the findings must be invalidated (e.g. in the methods section), or the authors' valid arguments will be ignored. It must be absolutely clear and reproducible how the 243 DOSTs were selected from the 696 (and the limitations even of the 696 be made explicit). Ideally, there should be a graph showing which fraction of the 454 DOSTs were rejected for which reason, if the selection was not done at random. Reference to 101innovations and JROST is not nearly sufficient. At the very least, inasmuch as they cannot be addressed methodologically, these alternative explanations (see also below) must be mentioned and discussed. Some more examples:

Statements like: "The results of this paper demonstrate the heterogeneity of the actors, power dynamics and stakeholders that are currently driving and dominating the evolution of the DOST ecosystem.", are not valid in this form as long as it is not clear that these results cannot be explained by selection bias. Without ruling out selection bias, it would be admissible to write "indicate" or "suggest", but not "demonstrate".

Also drawing from the skewed distribution of countries and organizations in their sample (about

35% of the total 696 DOSTs in the Kramer & Bosman database), the authors make the very valid case of "lock-in", "centralization" and "dominance" that should be avoided or at least mitigated. One of several good arguments the authors are using to make their case is that the decision about "how open science is done" should not be put in the hand of the dominant players. However, the authors argue as if politics, power, history and other such unrelated factors were the only explanation for the observed dominance in their sample. They fail to consider and provide arguments against alternative explanations for the dominance, such as selection bias in their sample; superior quality of the DOSTs; larger proportion of researchers/developers being able to produce more DOSTs; more opportunities for alternative careers in DOST production for ex-scholars in developed countries; etc. If the reader is supposed to understand why the dominance ought to be mitigated or balanced, the reader needs to understand why the dominance is a bug and not an inevitable outcome of the status quo beyond science or even a feature. Otherwise, readers will just stop reading.

The authors also only briefly mention in their penultimate paragraph of their discussion (instead of prominently featuring) one of the most powerful mitigation strategies both against a dominant "way to do open science" and lock-in: open, community-derived standards. It may be counter-productive to admonish readers to behave better and then only mention in passing what such better behavior would actually be, towards the end when most readers have already 'tired'. To keep people reading, in my humble opinion, this aspect should even be mentioned in the abstract as a potential recommendation or outcome prompted by the data.

It would also strengthen the authors' arguments if they emphasized the area between arbitrarily setting a standard of "how open science should be done" and the demarcation towards pseudo-science and non-science. Ultimately, there is a distinction between science and other forms of knowledge (or pseudo-science) and there is a distinction between open science and other forms of science. There is a difference between aiming for broad participation and "anything goes". This distinction is missing from this discussion and this omission, again, makes it too easy to attack the authors' valid and important argument simply by making an "anything goes" strawman out of it, to be dismissed and ignored. Especially in an age of fake news and alternative facts, this distinction cannot be left out of this discussion.

Along the same lines, the authors write: "Nonetheless, the DOST ecosystem is a complicated landscape, and imposing a specific value set or "way of doing things" will harm the richness and diversity of this rapidly evolving field." In science, diversity cannot be a value in itself, it must be a means to an end; to improve science. After all, not all results are equally valid, we often design experiments with the explicit goal to find out which result is the valid one. The diversity of valid results is important, not diversity in and of itself (An aside: analogous arguments can be made for openness. Openness in science is also a means to an end, and not a value in and of itself). The kind of diversity that does not improve science, science can do without. This is why there is education and training - to reduce the diversity that does not improve science and increase the diversity that moves science forward. The authors need to make clear what should happen to ensure the right kind of diversity and prevent the wrong kind of diversity, instead of emphasizing diversity as if it would be an unmitigated good under all circumstances.

A similarly easy to dismiss argument is when the authors point out that internet connectivity is still a problem in many countries. At least for this reader, this passage read as if DOST designers ought to keep this in mind when designing their DOSTs. However, the internet is what allowed open

science to be invented. No internet, no open science. I would consider such critique misdirected.

I hope I have provided a few examples of how some of the authors' arguments may be all too easily dismissed by a more critical reader than me. I also hope these examples were sufficient to allow the authors to go over all of their arguments in the discussion and think of ways to dismiss them, such that the authors are empowered to revise their discussion section to make it more forceful, convincing and less easy to dismiss.

I see no issues with the introduction and results section.

Minor comments:

The authors briefly mention "institutional repositories" as missing from their database. This statement is somewhat too narrow. In my field, biology, alone, there are about 1500 databases for raw data that serve not only to archive and make the data accessible (often sequences of biological molecules, but also structures, measurements etc.), but also to provide the respective communities with means to navigate and evaluate the gigantic amounts of data being collected in such 'big science' fields. Physics and climatology spring to mind as other fields with likely similarly large and diverse databases. All of these ought to be classified as DOSTs and together with institutional repositories may even number in the low tens of thousands in total. In the light of such numbers, it becomes even more important to emphasize the way the 242 DOSTs were selected.

"IT-professionals influence the design and deployment of the technology" sounds a bit like it were better if amateurs would have that influence, or at least that they ought to have the same influence. Experts and professionals may be biased, but they are nonetheless the best we have to offer. Using amateurs only because of their different biases will not improve piloting a plane, nor medical surgery, nor science.

"The section above highlighted how inequalities, marginalization and injustices were perpetrated by the current structure of the DOST ecosystem." - 'perpetrated' or 'perpetuated'?

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Biology, Neuroscience, infrastructure modernization, open science, open access, open data, FOSS.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 09 Mar 2021

Louise Bezuidenhout

This is an amazing piece of work, one of its kind (hopefully the first of its kind) in breadth and scope. It is highly timely and dearly needed. I also fully support the social agenda the authors are promoting in this work and my suggestions below are aimed towards strengthening their important arguments. To achieve change, arguments for change need to be convincing. The goals and direction of the authors' arguments are clear and I support them. However, the way they are articulated here, weakens their power and makes them too easy to dismiss. Below my suggestions of making the arguments more powerful.

One of the main conclusions from the work here is that DOSTs are heavily biased towards scholars in rich countries. While this is an expected outcome, a more critical reader could counter that this may be purely a result of the bias of the authors: selection bias confirming the perspective the authors had when they started to work. To counter this argument, the authors need to explain in great detail, how the database they analyzed was derived. For instance, the Kramer and Bosman database consists of currently 696 DOSTs, but the one presented here only contains 242 DOSTs. The authors need to explicitly state, in sufficient detail, how the 242 were selected from the 696 in order to ensure that "selection bias" cannot be used against them.

In the text it is noted that the list of DOSTs we were working from was by no means exhaustive. Rather, by basing our list primarily on the JROST and 101innovation lists we were starting our analysis with a selection of - to our knowledge - the most prominent and widely-used DOSTs in current circulation. It is true that the 101innovations list has since its publication 2016 has expanded widely, and offers many more DOSTs to potentially include in our future analyses. In order to address the concerns raised by reviewer 1, the following additions have been made:

- 1. Footnote 16 has been expanded to include the above paragraph. This footnote highlights the expansion of the 101innovations list that identifies a range of additional DOSTs that require analysis. It also emphasises the intention of this project, and our dataset, as a dynamic and ongoing venture that will require continual and collective updating and revision.*
- 2. A more comprehensive description of inclusion and exclusion criteria has been added to the methods section. This includes the text on page 12 that states: The inclusion criteria inevitably was subject to selection criteria, as DOSTs with smaller user groups might not have been included. Similarly, DOSTs designed to operate in languages outside of English, French, Spanish and Portuguese were less likely to be included. While these selection biases*

are important to note, it is important to recognize that the current method of DOST selection ensured the final list represented tools that were in common use by international researcher communities (as verified by 101Innovations and JROST) and covered a broad range of disciplines.

For instance, a critical reader may claim that the remaining, un-analyzed 454 DOSTs may all have been from LMICs and have been left out in order to score political points. More perniciously, a more critical reader may claim that LMIC are not represented with many DOSTs because they either lack the ability to design DOSTs or that the designed DOSTs from LMICs are so bad, nobody would ever want to use them, so they never make it into any database. Such alternative (and analogous) explanations for the findings must be invalidated (e.g. in the methods section), or the authors' valid arguments will be ignored.

While the reviewer's concerns are both extremely pertinent and appreciated, in our view they do not necessarily detract from the strength of our argument or analysis. In our methods we recognise limitations relating to our sampling, namely that we were focused predominantly on prominent DOSTs within the research ecosystem. Because DOST development often requires considerable time and financial investment, it is not particularly surprising that many of these DOSTs have originated in countries with well-developed and well-funded research ecosystems. This, however, in no way casts any negative light on DOSTs developed in LMICs. Indeed, the evolution of this dataset - and the work of 101innovations and JROST - might represent a very different geographical distribution of DOSTs in the coming years. This observation has been included in the text on page 12:

The final list of DOSTs did not contain a high proportion developed in low/middle-income countries. This was not due to any selective bias during the list construction, but rather is likely a reflection of the current DOST ecosystem. Because DOST development often requires considerable time and financial investment, it is not particularly surprising that many of these DOSTs have originated in countries with well-developed and well-funded research ecosystems. This, however, in no way casts any negative light on DOSTs developed in LMICs. Indeed, the evolution of this dataset - and the work of 101innovations and JROST - might represent a very different geographical distribution of DOSTs in the coming years.

It must be absolutely clear and reproducible how the 243 DOSTs were selected from the 696 (and the limitations even of the 696 be made explicit). Ideally, there should be a graph showing which fraction of the 454 DOSTs were rejected for which reason, if the selection was not done at random. Reference to 101innovations and JROST is not nearly sufficient. At the very least, inasmuch as they cannot be addressed methodologically, these alternative explanations (see also below) must be mentioned and discussed.

As discussed above, the inclusion and exclusion criteria in the methods section has been expanded to justify the DOSTs included in our analysis. The DOST ecosystem is rapidly evolving, so any accuracy here and years after both initiatives have been actively curated would be redundant/misleading/irrelevant in our view. We have included the following text on page 11: As the DOST ecosystem is rapidly expanding, the use of these two independently verified lists provided an important starting basis for our analysis.

Statements like: "The results of this paper demonstrate the heterogeneity of the actors, power dynamics and stakeholders that are currently driving and dominating the evolution of the DOST ecosystem.", are not valid in this form as long as it is not clear that these results cannot be explained by selection bias. Without ruling out selection bias, it would be admissible to write "indicate" or "suggest", but not "demonstrate".

The text has been amended by replacing "demonstrate" with "indicate"

Also drawing from the skewed distribution of countries and organizations in their sample (about 35% of the total 696 DOSTs in the Kramer & Bosman database), the authors make the very valid case of "lock-in", "centralization" and "dominance" that should be avoided or at least mitigated. One of several good arguments the authors are using to make their case is that the decision about "how open science is done" should not be put in the hand of the dominant players. However, the authors argue as if politics, power, history and other such unrelated factors were the only explanation for the observed dominance in their sample. They fail to consider and provide arguments against alternative explanations for the dominance, such as selection bias in their sample; superior quality of the DOSTs; larger proportion of researchers/developers being able to produce more DOSTs; more opportunities for alternative careers in DOST production for ex-scholars in developed countries; etc. If the reader is supposed to understand why the dominance ought to be mitigated or balanced, the reader needs to understand why the dominance is a bug and not an inevitable outcome of the status quo beyond science or even a feature. Otherwise, readers will just stop reading.

While the reviewer makes a valuable counter-argument, we believe that our analysis stands. We do not speculate as to the cause of this unequal distribution of DOSTs - and indeed some of the suggestions offered by the reviewer are highly likely. Rather, we suggest that regardless of the multiple causes underpinning this distribution, that unequal distribution nonetheless occurs. This is both highly influential on the future of the DOST ecosystem, but unlikely to be rectified by any one cause. More text has been added to the paper on page 28 to clarify this issue:

It is important to recognise that the power dynamics shaping the DOST ecosystem are likely influenced by multifarious pressures. These include community decisions regarding the quality of the DOSTs, the unequal distribution of DOST users and researchers/developers around the globe, economic issues including the opportunities for alternative careers in DOST development in developed countries. The limits of this study make it impossible to engage with the influence of these different pressures in detail. Rather, the study seeks to highlight not only the complexities of the DOST ecosystem, but also the critical need for more scholarship on its evolution. This paper highlights how these pressures cause the dominance of certain DOSTs to be a "bug" in the current structuring of the ecosystem, rather than as an inevitable outcome of the status quo of scientific research.

The authors also only briefly mention in their penultimate paragraph of their discussion (instead of prominently featuring) one of the most powerful mitigation strategies both against a dominant "way to do open science" and lock-in: open, community-derived standards. It may be counter-productive to admonish readers to behave better and then only mention in passing what such better behavior would actually be, towards the end when most readers have already 'tired'. To keep people reading, in my humble opinion, this aspect should even be mentioned in the abstract as a potential recommendation or

outcome prompted by the data.

We thank the reviewer for this suggestion, and have incorporated it into the abstract.

It would also strengthen the authors' arguments if they emphasized the area between arbitrarily setting a standard of "how open science should be done" and the demarcation towards pseudo-science and non-science. Ultimately, there is a distinction between science and other forms of knowledge (or pseudo-science) and there is a distinction between open science and other forms of science. There is a difference between aiming for broad participation and "anything goes". This distinction is missing from this discussion and this omission, again, makes it too easy to attack the authors' valid and important argument simply by making an "anything goes" strawman out of it, to be dismissed and ignored. Especially in an age of fake news and alternative facts, this distinction cannot be left out of this discussion.

Footnote 15 has been added: While there are many interpretations of how openness can be enacted in research, there is a growing community of researchers who recognize open science practices such as open publishing, data sharing and reproducible research. There is also a growing consensus on the values underpinning the open research movement and its commitment to equity and access. Our argument attempts to problematize the extent of this value and practice consensus within research communities, rather than to engage in the boundaries between science and pseudo-science. To address these boundaries we regard as beyond the scope of this paper.

Along the same lines, the authors write: "Nonetheless, the DOST ecosystem is a complicated landscape, and imposing a specific value set or "way of doing things" will harm the richness and diversity of this rapidly evolving field." In science, diversity cannot be a value in itself, it must be a means to an end; to improve science. After all, not all results are equally valid, we often design experiments with the explicit goal to find out which result is the valid one. The diversity of valid results is important, not diversity in and of itself (An aside: analogous arguments can be made for openness. Openness in science is also a means to an end, and not a value in and of itself). The kind of diversity that does not improve science, science can do without. This is why there is education and training - to reduce the diversity that does not improve science and increase the diversity that moves science forward. The authors need to make clear what should happen to ensure the right kind of diversity and prevent the wrong kind of diversity, instead of emphasizing diversity as if it would be an unmitigated good under all circumstances.

This point is well-taken, and we have qualified our support for diversity with clarification that we support diversity of data, methods and perspectives. The following text has been added to page 30:

While methodological diversity is not inherently a value of research, it remains a core element of scientific research. Finding a balance between centralized efficiency and freedom to innovate is a central element of flourishing research systems. Similarly, finding that balance within the DOST ecosystem will be extremely influential in its future successes. Rather than imposing restrictions on what should constitute a DOST, we suggest that those designers and users be supported to critically reflect on the values that they are introducing into the ecosystem. There are many

models currently in use on how to balance well-intentioned innovation with pragmatic requirements, and these need to be more strongly developed for DOSTs.

A similarly easy to dismiss argument is when the authors point out that internet connectivity is still a problem in many countries. At least for this reader, this passage read as if DOST designers ought to keep this in mind when designing their DOSTs. However, the internet is what allowed open science to be invented. No internet, no open science. I would consider such critique misdirected.

We have added footnote 20: While some scholars argue that access to the internet allowed open science to be invented, we take a more nuanced view. Both openness and internet access are not "binary" positions in which users are either open/closed or online/offline. Having poor internet connections can allow some participation in open science, but severely restrict the amount of activity - or the types of DOSTs - available to the individual working in these settings. Openness, like internet access, is a continuum.

I hope I have provided a few examples of how some of the authors' arguments may be all too easily dismissed by a more critical reader than me. I also hope these examples were sufficient to allow the authors to go over all of their arguments in the discussion and think of ways to dismiss them, such that the authors are empowered to revise their discussion section to make it more forceful, convincing and less easy to dismiss.

I see no issues with the introduction and results section.

Minor comments:

The authors briefly mention "institutional repositories" as missing from their database. This statement is somewhat too narrow. In my field, biology, alone, there are about 1500 databases for raw data that serve not only to archive and make the data accessible (often sequences of biological molecules, but also structures, measurements etc.), but also to provide the respective communities with means to navigate and evaluate the gigantic amounts of data being collected in such 'big science' fields. Physics and climatology spring to mind as other fields with likely similarly large and diverse databases. All of these ought to be classified as DOSTs and together with institutional repositories may even number in the low tens of thousands in total. In the light of such numbers, it becomes even more important to emphasize the way the 242 DOSTs were selected.

Institutional repositories were not included because of the capacity of the research team. We hope to include their analysis in future iterations of this project, of which we hope that this paper is the first proof of concept.

"IT-professionals influence the design and deployment of the technology" sounds a bit like it were better if amateurs would have that influence, or at least that they ought to have the same influence. Experts and professionals may be biased, but they are nonetheless the best we have to offer. Using amateurs only because of their different biases will not improve piloting a plane, nor medical surgery, nor science.

We feel that the reviewer misinterpreted the sentence. What this sentence does is highlight how the intentions, experiences, priorities and cultures of the IT-professionals are introduced

into the design of DOSTs. *We believe that their inclusion in design decisions can influence the overall design of DOSTs, in keeping with scholarship in the field of Science and Technology Studies. The statement has nothing to do with amateurs - or the reviewer's rather trite suggestion of amateurs piloting planes or contributing to surgery.*

"The section above highlighted how inequalities, marginalization and injustices were perpetrated by the current structure of the DOST ecosystem." - 'perpetrated' or 'perpetuated'?

Correction made.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research