



HAL
open science

An empirical assessment of Open Hardware practices in the academic community and their policy implications

André Maia Chagas, Alexandre Hannud Abdo

► To cite this version:

André Maia Chagas, Alexandre Hannud Abdo. An empirical assessment of Open Hardware practices in the academic community and their policy implications. Eu-SPRI Annual Conference, Jun 2023, Brighton, United Kingdom. hal-04488702

HAL Id: hal-04488702

<https://hal.science/hal-04488702>

Submitted on 4 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

An empirical assessment of Open Hardware practices in the academic community and their policy implications

WORK IN PROGRESS

Authors

Andre Maia Chagas (University of Sussex)*

Alexandre Hannud Abdo (LISIS, Univ Gustave Eiffel)

* corresponding author: a.maia-chagas@sussex.ac.uk

Abstract

Open Science Hardware (OSH) is a term for the practice of sharing, as open-source, the material means of science: the reproducible designs, assembly and operation instructions of research hardware, plus any software they may depend on. A practice that spans lab-grade pipettes, rotators, scales, sensors, microscopes, beams, and imaging devices, but may also include cell lines, reagents, genetic parts, and even laboratory automation systems, diagnostic equipment, small satellites and components of particle accelerators (Pearce, 2013; Ravindran, 2020).

Proponents of OSH articulate promises of better science and greater and more positive societal impacts, achieved through facilitated collaboration, higher replicability, reduced efforts, inclusion and more equitable research capacity, freedom to innovate, customise and appropriate technologies, and lower risks of monopolistic and lock-in conditions (Arancio et al., 2022; Chagas, 2018). In the open hardware model, a critical and early condition for such value co-creation are the qualities of the documentation of designs published (Bonvoisin et al., 2017). In the case of OSH, this concerns academic journals and online file repositories that papers refer to: Are they available, complete, and useful to reproduce and build on existing work? Are legal aspects, such as licences, properly dealt with? Is there an infrastructure for collaboration deployed? The answers to these questions are, thus, both informative of actual research practices and needs, and required to reliably connect them to the aforementioned promises, in view of planning, executing, monitoring and evaluating policies to foster OSH.

This paper contains two main contributions:

I) A qualitative assessment of the co-creation enabling characteristics of open scientific hardware publications, and associated features such as publication venue, allowing us to evaluate their progress in time and in relation to the establishment of standards for sharing open hardware designs.

II) A solution to a not uncommon problem with research on bibliographic databases: because OSH is an approach and not a subject of research, usual search strategies in bibliographic databases fail to capture the usage of the indicative term, in our case "open hardware", as it may be absent from title and abstract.

Keywords

Open hardware; Standards; Scientometrics; Socio-technical transitions; Open source;

Introduction

Open Science Hardware (OSH) is a term for the practice of sharing, as open-source, the material means of science: the reproducible designs, assembly and operation instructions of research hardware, plus any software they may depend on. A practice that spans lab-grade pipettes, rotators, scales, sensors, microscopes, beams, and imaging devices, but may also include cell lines, reagents, genetic parts, and even laboratory automation systems, diagnostic equipment, small satellites and components of particle accelerators (Pearce, 2013; Ravindran, 2020).

As policies in support of open scientific practices broadened their scope from publications to research data and software, an organised movement around OSH has taken shape under the umbrella of the Gathering for Open Science Hardware (*Global Open Science Hardware Roadmap - Gathering for Open Science Hardware*, 2017). Based on practice already experimented with and implemented at the level of individual laboratories and businesses (Ravindran, 2020; Stirling et al., 2020), OSH-oriented policy is being worked on at the level of institutions, such as CERN, the Wilson Center, Cambridge University, and UNESCO, as well as funders, such as the Sloan Foundation and UKRI (Parker et al., n.d.; *Policy Briefs - Gathering for Open Science Hardware*, 2021), and can also be linked to the earlier deployment of academic makerspaces.

Proponents of OSH articulate promises of better science and greater and more positive societal impacts, achieved through facilitated collaboration, higher replicability, reduced efforts, inclusion and more equitable research capacity, freedom to innovate, customise and appropriate technologies, and lower risks of monopolistic and lock-in conditions (Arancio et al., 2022; Chagas, 2018). In the open hardware model, a critical and early condition for such value co-creation are the qualities of the documentation of designs published (Bonvoisin et al., 2017). In the

case of OSH, this concerns academic journals and online file repositories that papers refer to: Are they available, complete, and useful to reproduce and build on existing work? Are legal aspects, such as licences, properly dealt with? Is there an infrastructure for collaboration deployed? The answers to these questions are, thus, both informative of actual research practices and needs, and required to reliably connect them to the aforementioned promises, in view of planning, executing, monitoring and evaluating policies to foster OSH.

This paper contains two main contributions:

I) A qualitative assessment of the co-creation enabling characteristics of open scientific hardware publications, and associated features such as publication venue, allowing us to evaluate their progress in time and in relation to the establishment of standards for sharing open hardware designs. For that, we treat documents claiming to publish OSH designs and grade them with the criteria put forward in the Open-o-Meter (Bonvoisin & Mies, 2018).

II) A solution to a not uncommon problem with research on bibliographic databases: because OSH is an approach and not a subject of research, usual search strategies in bibliographic databases fail to capture the usage of the indicative term, in our case "open hardware", as it may be absent from title and abstract. Our solution involves an automated three-level search, starting with full-text searches in Google Scholar, to obtain limited and unstructured information, then searching for each individual result using the Web of Science API, to obtain full and structured metadata. Finally, we download the full text for our qualitative evaluation.

This paper represents an ongoing effort, whose goals are to shed light on the evolving scenario of Open Science Hardware practices, on the process of community learning around them, on the role of standards for adequately sharing hardware designs. We also intend it as the first release of an evolving database for the monitoring of OSH related practices and policy. We expect that representative and detailed data on the essential qualities of OSH publications will assist journals, funding agencies and research institutes to better evaluate the outcomes and impacts of OSH and improve the focus of their efforts so that OSH can fulfil its potential to enhance, and democratise participation in, research and innovation.

Methods

To assess the design documentation of OSH articles in a systematic way, across disciplines and journals, the first step was to identify these articles. We initially attempted to use known bibliographic databases such as Web of Science and Scopus, querying for search terms related to our goal, such as "open hardware", but quickly realised that this missed a significant number of articles. Indeed, such bibliographic databases only contain metadata, searching for information contained in articles titles, abstracts, authors etc. However, many OSH articles only contain

terms indicative of their OSH status in sections of the full-text, like the introduction, methods, etc. One example of such an article can be seen here: [<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0143547>].

While there are databases that collect the full-text of many open-access journals, such as CORE, the open-access status of an article is actually a criteria for openness and so we must account for all open-access status. To our knowledge, the only easily available database that can search the full-text of articles regardless of their open-access status is Google Scholar (GS). However, using GS presents another set of barriers: (a) query results are not downloadable; (b) only the first 1,000 results of each search are available, even if GS knows of more results; (c) GS has systems in place to avoid systematic querying using automated means; (d) GS results only provide partial and limited extracts of a document's metadata, that are not structured in a way that allows for reliable, large-scale comparisons. In spite of issue (a), we were able to put in place an automated and systematic approach for mass downloading query results. Using Scrapy (*Scrapy | A Fast and Powerful Scraping and Web Crawling Framework*, n.d.), an open source library for web scraping, we developed procedures (see section "Code and Data availability") to query GS and store the results in a local database. This further enabled us to use time as a parameter to break down the total body of articles into smaller result packets, overcoming issue (b). In other words, by searching for a query in the form "open hardware" AND "June" restricted to the year 2007, we leverage the fact that all the text in the article is taken in consideration to build queries that are more fine-grained in time than simply restricting the publication year, and that can be put back together to reconstitute the year in question, thus capturing more articles than the limit of 1,000 per year. Finally, to work around issue ©, we proxied and delayed our queries (Scraper API - (*ScraperAPI - The Proxy API For Web Scraping*, n.d.) so that GS wouldn't block us for overquerying.

Having obtained from queries to the full text the partial and limited metadata available from GS, we had to obtain from that properly structured and complete metadata. For that, we employed the Web of Science ('Web of Science Platform', n.d.) API, with queries that used the GS information to target each GS result and get the WOS metadata for all articles. Last, we then used Unpaywall (*Unpaywall: An Open Database of 20 Million Free Scholarly Articles*, n.d.) in order to get information on whether an article was open access (OA) or not, and which kind of OA ("gold", "green", "bronze", "hybrid") did it correspond to. This formed our final database, so that we could proceed to manually evaluate the quality of the OSH documentation shared in the papers from our corpus.

For the manual scoring, we created a 18 point system based on the Open-O-Meter (Bonvoisin & Mies, 2018), where the documentation quality of open hardware designs is divided into 8 main areas, then scored by giving 1 or 2 points for each, as well as for whether or not an article was OA. Scorers were instructed to evaluate each area and give one point if the article only fulfilled part of the

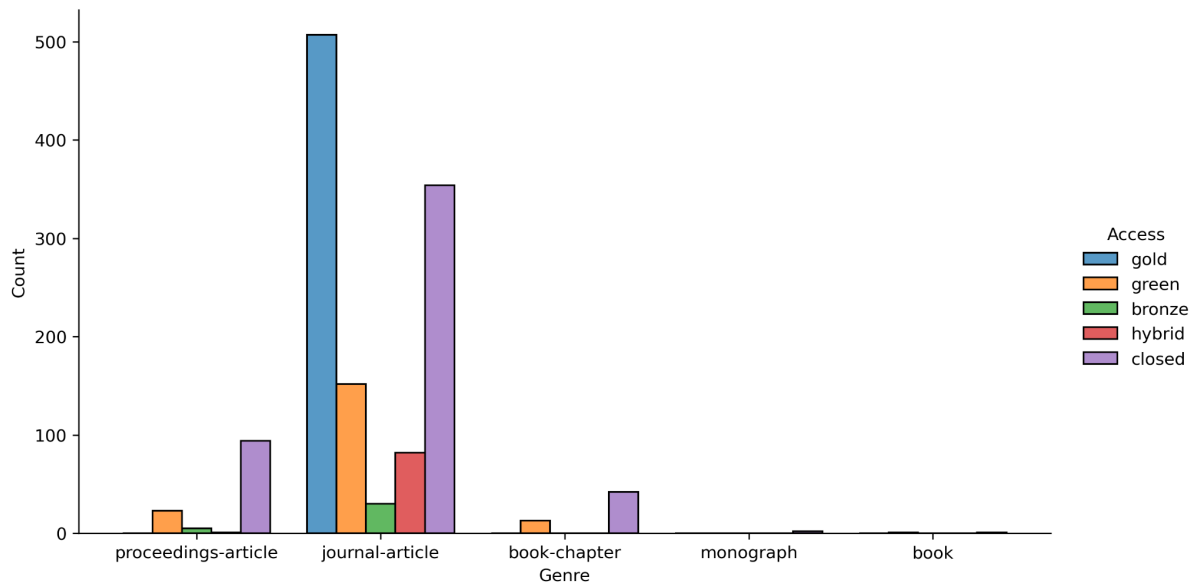
requirements in said area (e.g., if only part of the design files were available in original format, this would result in one point). An 18 score means that a certain OSH design documentation is well structured to be used, modified and updated by others. We added the extra point for OA as it seems contradictory to have OSH designs hidden behind paywalls.

Given the above procedure, circa 100 publications were scored so far, out of 1000 items in our database, enough for us to have a rough estimate of what to expect from concluding our work.

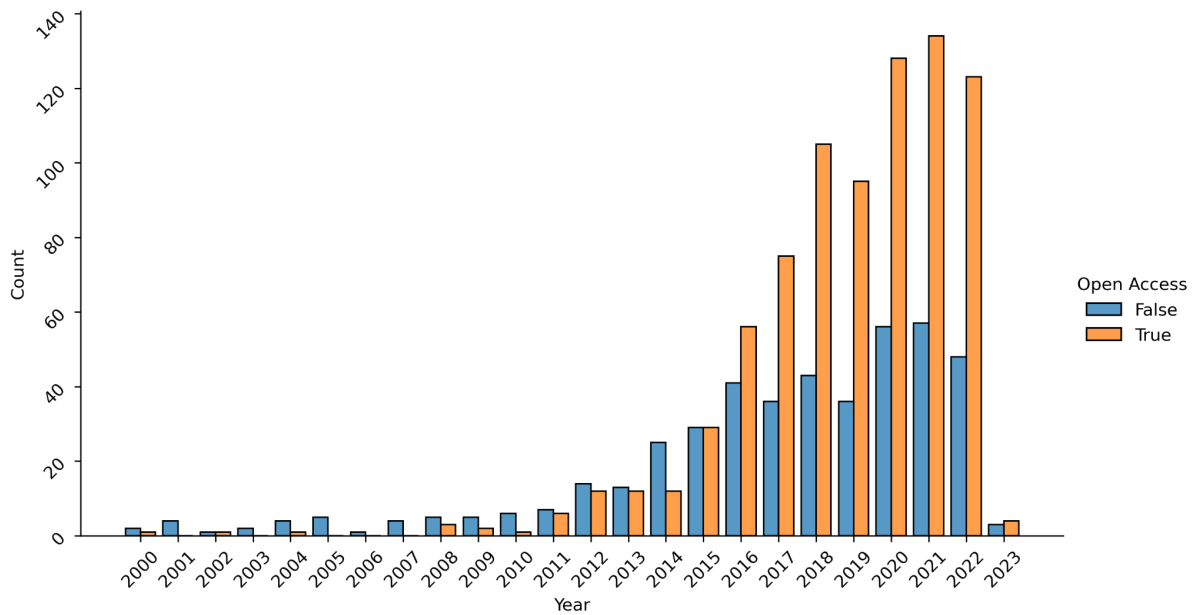
Results

Our systematic search and structuring of GS hits has led to ~1000 items that have “open hardware” as part of their text, with an estimated 500 entries describing OSH designs. The metadata database is publicly available (See section “Code and Data availability”).

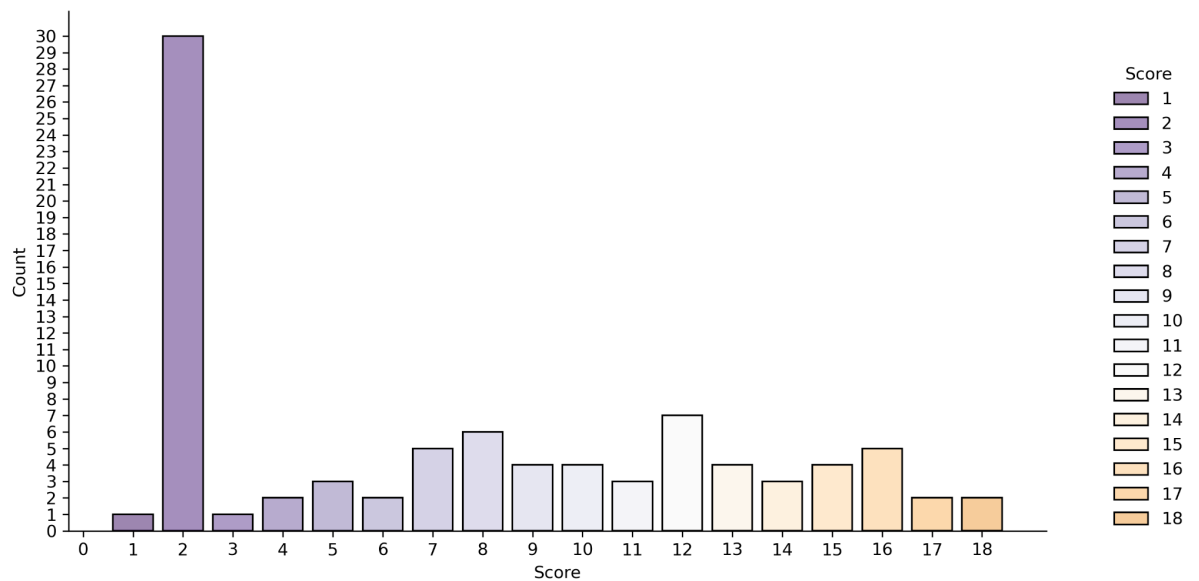
The following figure shows the distribution of work in journal articles, proceeding articles, book chapters, monographs and books. Journal articles make up for the majority of items, with ~900 entries. Out of those, ~600 are OA. The access type for the OA articles is unevenly split, with “gold” composing the majority of the of the database, about 400 articles, “green” making up for about 135 articles, “bronze” about 25 and hybrid “50” articles.



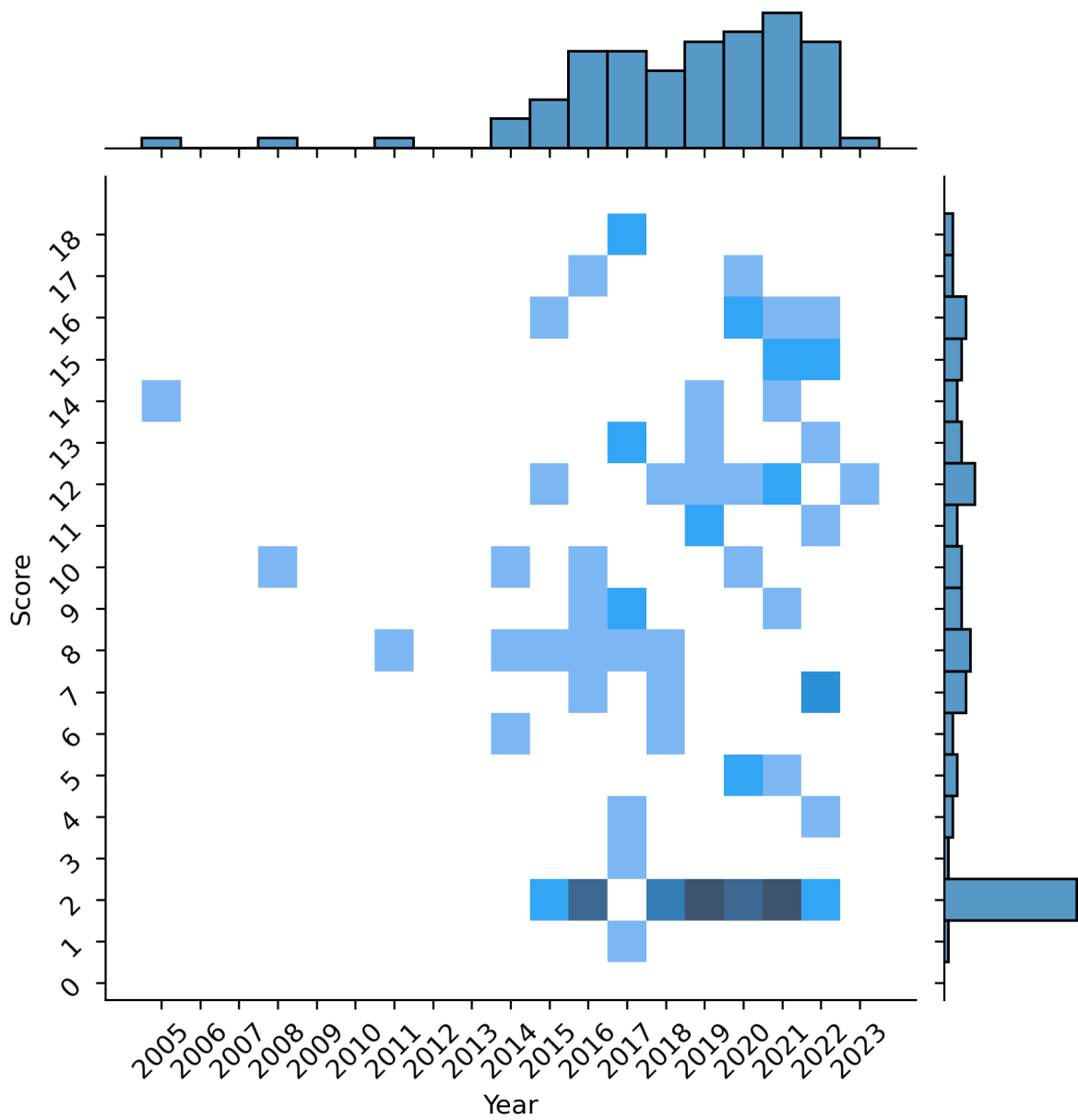
During the period between 2005 and 2021, the number of OA articles has increased year over year at a higher pace than closed source articles, as seen in the figure below.



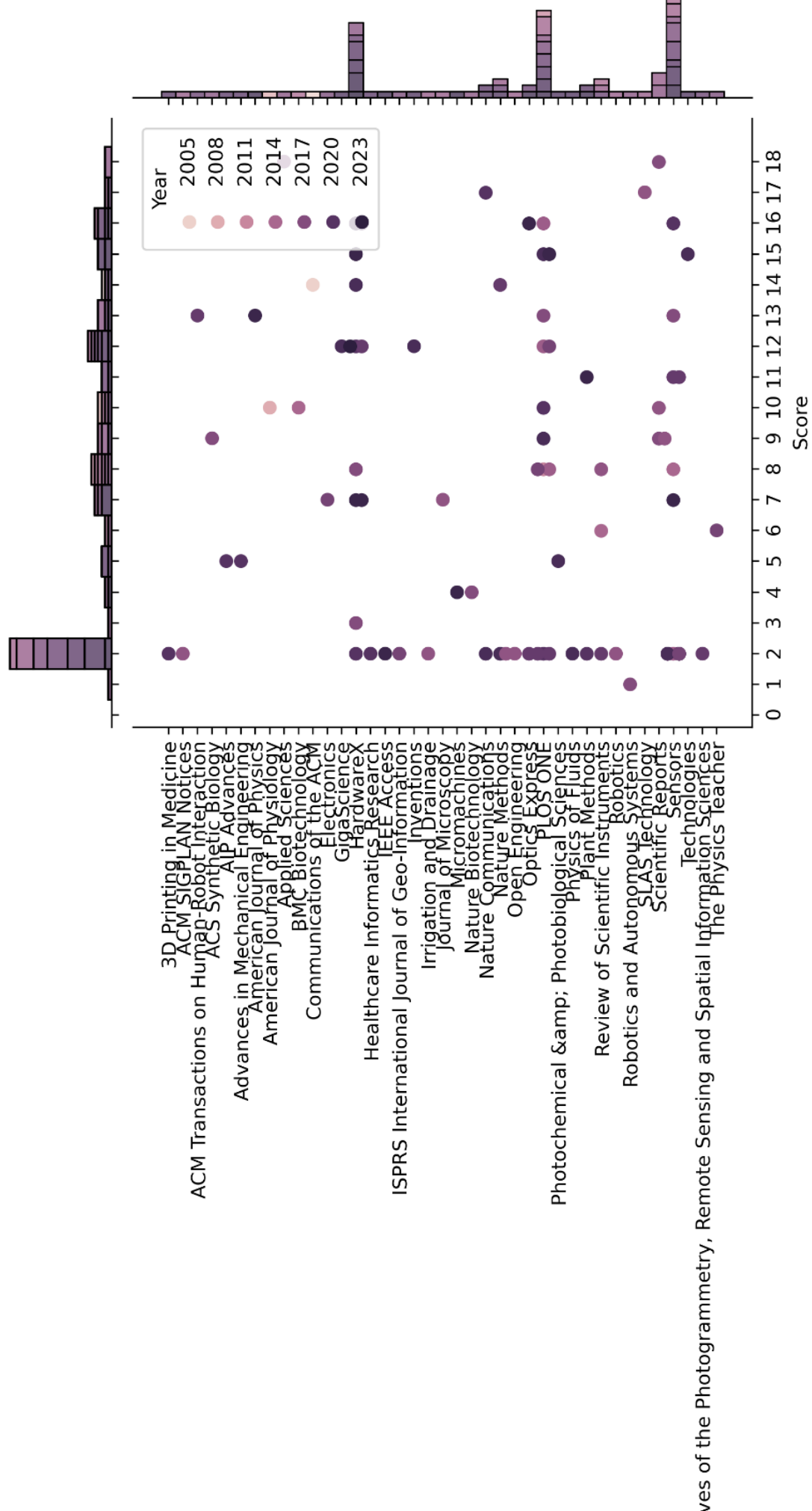
So far, we have scored ~100 journal articles across 39 journals. The next figure shows the distribution of scores, which range between 1 and 18. We note a generally very wide distribution, with a large number of entries that score only 2 points, despite these articles claiming to publish “open hardware”.



To better understand the evolution of quality in time, we can plot a bidimensional histogram, disentangling scores from year of publication. There, we actually see the large number of score 2 articles belong to recent years and not to old practices that could have waned.



Our final figure shows in addition the venue of publication, where we can see that bad open hardware practices are not tied to a particular journal or discipline.



Discussion

This paper shows two main findings: First, a method that enables keywords to be found in the entire text of an item in a bibliographic database (as opposed to just search title and abstract). Because this solution can be used for virtually any search terms, it opens up the possibility to investigate topics that are interdisciplinary and not amenable to data collection using traditional search engines in bibliographic databases.

Our analysis shows that the number of papers describing OSH designs has significantly grown over the last two decades, with the vast majority being published as open access resources. We attribute this increase to a couple of factors: I) The improvement in storage capabilities of hosting websites and platforms, allowing anyone to make large datasets publicly available at low costs. This enables the sharing of complex designs with multiple files, such as code, computer aided designs (CAD) files as well as images, and videos needed for assembly instructions. II) the development of “maker platforms and ecosystems” in which electronic components and machines capable of complex functions are now readily available at affordable costs and come together with detailed “how-tos” (eg [openMV](#) is an OSH camera system capable of complex machine vision tasks, which can be bought for under 100 dollars). III) The open science movement has brought attention to the importance of the tools used for research, including ways to incentivise their open development and sharing. Moreover, granting agencies and governments are starting to focus on open science in their granting policies as a way to improve scientific output and correct invest in open science approaches (the United States federal government has announced 2023 as the [year of open science](#)).

Despite the growth in the number of OSH articles, our research indicates that not all “OSH is created equal” and the completeness level of their documentation varies greatly, with no correlation to year, type of access, and which journal published the article. This indicates that even though there are several specifications ([OSHA definition](#), [DIN spec](#)) defining what is OSH and how it should be documented, the academic community still does not leverage their use, and journals do not have guidelines in place to make use of these systems.

Code and Data availability

Code and data are available at the following web locations:

- <https://github.com/amchagas/open-hardware-supply>
- https://www.zotero.org/groups/4871493/open_hardware_database

References

- Arancio, J., Tirado, M., & Pearce, J. M. (2022). *Equitable Research Capacity Towards the Sustainable Development Goals: The Case for Open Science Hardware*. *Journal of Science Policy & Governance*.
https://www.sciencepolicyjournal.org/article_1038126_jspg210202.html
- Bonvoisin, J., & Mies, R. (2018). Measuring Openness in Open Source Hardware with the Open-o-Meter. *Procedia CIRP*, 78, 388–393.
<https://doi.org/10.1016/j.procir.2018.08.306>
- Bonvoisin, J., Mies, R., Boujut, J.-F., & Stark, R. (2017). What is the “Source” of Open Source Hardware? *Journal of Open Hardware*, 1(1).
<https://doi.org/10.5334/joh.7>
- Chagas, A. M. (2018). Haves and have nots must find a better way: The case for open scientific hardware. *PLOS Biology*, 16(9), e3000014.
<https://doi.org/10.1371/journal.pbio.3000014>
- Global Open Science Hardware Roadmap—Gathering for Open Science Hardware*. (2017, December 20).
<https://openhardware.science/global-open-science-hardware-roadmap/>
- Parker, A., Dosemagen, S., & Molloy, J. (n.d.). *Open Hardware: An Opportunity to Build Better Science | Wilson Center*. Retrieved 14 May 2023, from <https://www.wilsoncenter.org/publication/open-hardware-opportunity-build-better-science>
- Pearce, J. (2013). *Open-Source Lab—1st Edition* (1st ed.).
<https://www.elsevier.com/books/open-source-lab/pearce/978-0-12-410462-4>
- Policy Briefs—Gathering for Open Science Hardware*. (2021, September 29).

<https://openhardware.science/policy-briefs/>

Ravindran, S. (2020). How DIY technologies are democratizing science. *Nature*, 587(7834), 509–511. <https://doi.org/10.1038/d41586-020-03193-5>

ScraperAPI - The Proxy API For Web Scraping. (n.d.). ScraperAPI. Retrieved 14 May 2023, from <https://www.scraperapi.com/>

Scrapy | A Fast and Powerful Scraping and Web Crawling Framework. (n.d.). Retrieved 14 May 2023, from <https://scrapy.org/>

Stirling, J., Sanga, V. L., Nyakyi, P. T., Mwakajinga, G. A., Collins, J. T., Bumke, K., Knapper, J., Meng, Q., McDermott, S., & Bowman, R. (2020). The OpenFlexure Project. The technical challenges of Co-Developing a microscope in the UK and Tanzania. *2020 IEEE Global Humanitarian Technology Conference (GHTC)*, 1–4. <https://doi.org/10.1109/GHTC46280.2020.9342860>

Unpaywall: An open database of 20 million free scholarly articles. (n.d.). Retrieved 14 May 2023, from <https://unpaywall.org/>

Web of Science platform. (n.d.). *Clarivate*. Retrieved 14 May 2023, from <https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/webofscience-platform/>