A tale of two 'opens': intersections between Free and Open Source Software and Open Scholarship

Jonathan P. Tennant* (0000-0001-7794-0218), Ritwik Agrawal (0000-0002-0778-8916), Ksenija Baždarić (0000-0002-2977-3686), Devin R. Berg (0000-0002-1193-3848), David Brassard (0000-0003-0765-7403), Tom Crick (0000-0001-5196-9389), Daniel J. Dunleavy (0000-0002-3597-7714), Stephen Eglen (0000-0001-8607-8025), Thomas Rhys Evans (0000-0002-6670-0718), Nicholas Gardner (0000-0002-5278-7541), Laurent Gatto (0000-0002-1520-2268), Monica Gonzalez-Marquez (0000-0002-5114-9599), Daniel Graziotin (0000-0002-9107-7681), Bastian Greshake Tzovaras (0000-0002-9925-9623), Daniel Gunnarsson (0000-0002-6058-4285), Johanna Havemann (0000-0002-6157-1494), Mohammad Hosseini (0000-0002-2385-985X), Daniel S. Katz (0000-0001-5934-7525), Marcel Knöchelmann (0000-0003-1050-1303), Leo Lahti (0000-0001-5537-637X), Christopher R. Madan (0000-0003-3228-6501), Paolo Manghi (0000-0001-7291-3210), Alberto Marocchino (0000-0002-5287-8355), Paola Masuzzo (0000-0003-3699-1195), Peter Murray-Rust (0000-0003-3386-3972), Sanjay Narayanaswamy (0000-0002-3069-001X), Gustav Nilsonne (0000-0001-5273-0150), Josmel Pacheco-Mendoza (0000-0002-2251-8092), Bart Penders (0000-0002-2661-9181), Esther Plomp (0000-0003-3625-1357), Olivier Pourret (0000-0001-6181-6079), Michael Rera (0000-0002-6574-6511), John Samuel (0000-0001-8721-7007), Tobias Steiner (0000-0002-3158-3136), Jadranka Stojanovski (0000-0001-7399-522X), Alejandro Uribe Tirado (0000-0002-0381-1269), Rutger Vos (0000-0001-9254-7318), Simon Worthington (0000-0002-8579-9717), Tal Yarkoni (0000-0002-6558-5113).

*Corresponding author: jon.tennant.2@gmail.com.

Abstract

sudo make when finished | grep “the end”

There is no clear-cut boundary between FOSS and Open Scholarship, and the histories, practices, and fundamental principles between the two remain complex.

With this in mind, and based on our thematic comparison here, we conclude that...

1. There is substantial scope for new communities of practice to form within scholarly communities that place sharing and collaboration/open participation at their focus.
2. It is both the principles and practices of FOSS that can be more deeply ingrained in this way. I.e., a balance of pragmatism and social ideology
3. Open Scholarship risks being subverted by commercial players. There is a need for research funders to invest in sustainable open scholarly infrastructure, and the communities that support them, to avoid the capture and enclosure of key research services that would prevent such behaviours.
4. The acceleration towards a system of Open Scholarship will be greatly enhanced by a concurrent shift in recognising a broader range of practices and outputs beyond traditional peer review and research articles.
5. This shift could ultimately lead to a more healthy scientific culture, where competition is replaced by collaboration, resources (including time and people) are shared more efficiently, and it becomes inherently more rigorous and reproducible.

Introduction

"Information wants to be free." – Stewart Brand

It is the best of times, it is the worst of times, it is the age of wisdom, it is the age of foolishness, it is the epoch of belief, it is the epoch of incredulity, it is the season of Light, it is the season of Darkness, it is the spring of hope, it is the winter of despair, we have everything before us, we have nothing before us. Departing from the words of Dickens, in short, when it comes to the state of scholarship, we are at a crossroad where existing and imagined infrastructures overlap, compete and displace one another. At the heart of that struggle lies our evaluation, valuation, use and operationalisation of “openness” in scholarship: what it was, what it is, what it can be. To each of those three, multiple legitimate answers exist. That multiplicity is the topic of the present article.

Since the 1980s, some of the key developments in academia have been based on the open revolution, and in particular around Free and Open Source Software (FOSS) and Open Scholarship. FOSS has been a highly successful movement in a number of ways, and there

is much that could be learned from this and applied more broadly to academia, particularly with a resurgence in Open Scholarship in recent years. However, there remains little reciprocal understanding of the intersections between the two concepts, and therefore contrasting FOSS and Open Scholarship together can help to provide insight into these highly dynamic concepts and their relative successes or acceptance to date (Willinsky 2005; 2006; Levin et al. 2016; Levin and Leonelli 2017; Lahti, da Silva, et al. 2017; Grubb and Easterbrook 2011).

Specifically, we will address four key themes in the four sections of the paper below:

- First, we will discuss notions of 'openness' in FOSS and Open Scholarship, and how these constitute and concern one another. This section describes the history of FOSS over the last several decades, and the resurgence in 'open research practices' within modern academia. Supplementary to this, it explores the contested understandings of Open Scholarship or Open Science, and the inherent complexities that underpin these with regard to FOSS.
- Second, we ask which aspects of FOSS and Open Scholarship can be transferred between each other. This section primarily deals with three issues: first, about shaping communities of practice between FOSS and Open Scholarship; second, the fundamental principles and values behind both; and third, questions about the roles of commercial players in each space.
- In the third section, we discuss what we can learn about systems of reward, value, and reputation by comparing the two 'opens'. Many problems within modern scholarship can be traced towards a defective system of reputation accrual and reward. What can Open Scholarship learn from FOSS in this regard?
- In the fourth section we ask how 'openness' intersects with issues around reproducibility and data sharing. Many fields of modern research are undergoing what is often termed a 'reproducibility crisis'. This section will look into how FOSS and open data intersect here with Open Scholarship to address this critical issue.

This article was written in an open and collaborative manner, inspiring the concept of a MOOP, described by Tennant et al. (2019).

Glossary of terms used

- Open Source:
- Free Software:
- Open Science:
- Open Data:
- Open Access:
- Open Scholarship:
- Open Research Practices:
- License:
- Copyright:
- Re-use:

Understanding notions of ‘openness’ in software and scholarship

The large intersection between the principles of FOSS and Open Scholarship, and the partial historical overlap, allows us to investigate the co-production that generated both in its current multitude shapes. We recognise movements of convergence as well as divergence in their shared history. For example, software and data accessibility (e.g., availability, inspection, and re-use) is often now considered a prerequisite for making the full research environment transparent and rigorous enough to reproduce research results (Sandve et al. 2013; Hocquet and Wieber 2018; Millman and Perez 2014). In some modern research environments, workflows, such as data collection and analysis and the preparation and dissemination of publications, rely on a range of essential software or Web-based tools for which powerful FOSS solutions already exist (e.g., R, Python, Julia, GNU PSPP). The combined ability to independently reproduce and refute research is widely recognised as one of the critical elements for most of the scientific enterprise (Popper 1959). Openness in software, data and research also promises to help close the digital divide and enhance the democratic globalisation of research participation (Packer 2009; Fuchs and Horak 2008; May 2006). We can compare such relationships to help elucidate historical and future coevolution of conceptual understanding of openness, and to shed light on the wider system of modern scholarship; especially as the foundations behind FOSS tend to be more thoroughly studied and understood (Bretthauer 2001; Weber 2004).

A brief history of the Free Software and Open Source movements

The origins of *‘free’* software are closely tied to the development of the first mass-produced computers in the 1950s, bundled together as packages by IBM. The adoption of time-sharing mainframes, and the development of standardized operating systems and programming languages for such infrastructures (leading from Multics to Unix and the C language, in the late 1960s and early 1970s), created an environment in which source code for academic computing could be shared: repetitive work was left to machine and know-how could be quickly passed over. However, the Free Software movement itself was launched in 1983 (see the original announcement by Richard Stallman here), followed by the launch of the GNU project based around the free operating system GNU (1984); most importantly this represented the first FOSS organized manifesto. The purpose here was to produce a novel software system by sharing the code know-how generally, which was until then primarily covered by intellectual property. The idea of intellectual property was incompatible with a particular set of values based around what the founders called fundamental freedoms[1],

[1] “The freedom to run the program as you wish, for any purpose. The freedom to study how the program works, and change it so it does your computing as you wish. Access to the source code is a precondition for this. The freedom to redistribute copies so you can help your neighbor. The freedom to distribute copies of your modified versions to others. By doing this you can give the whole community a chance to benefit from your changes. Access to the source code is a precondition for this.”

which created a new set of values around encouraging the ability to run, study, change, adapt, and redistribute software. Since 1985, the movement has been overseen by the non-profit Free Software Foundation. The GNU project and a definition of Free Software were created as a result of this, becoming the *de facto* manifesto for the movement. The GNU General Public License (GPL) was first used in 1989 by the GNU developer team to broadly distribute software released as part of the GNU project. GNU is probably one of the most common and known cases of 'copyleft'. The current version (GPLv3), released in 2007, is still widely used today, in 2019.

GNU was just the beginning. In 1998, the Open Source Initiative (OSI) was launched as an offshoot of the Free Software movement. As part of the OSI, Eric S. Raymond outlined his issues with the term Free Software, particularly in its ambiguity regarding either pricing or licensing. He advocated for the replacement of Free Software with Open Source as a label to describe reusable software and the culture surrounding these. Much of the motivation for Open Source was based around the strategic rebranding of technical 'excellence' and pragmatism in software production to make it more compatible with commercial interests around commodification of source code, including through consultation services (Lerner and Triole 2000). However, even now and after 20 years since the inception of the OSI, the definition of Open Source remains contested. We have to remember that technology is undergoing an extreme growth since its birth, and definitions require constant adaptation to new areas that technology is developing; this makes it difficult to establish a global definition that can cover many different fields and interests. Now, FOSS has become an almost ubiquitous part of modern society, although dedicated projects seem to be largely focused in North America, China, and western Europe, with relatively little documented activity across Central America, SE Asia, and Africa (Mombach et al. 2018).

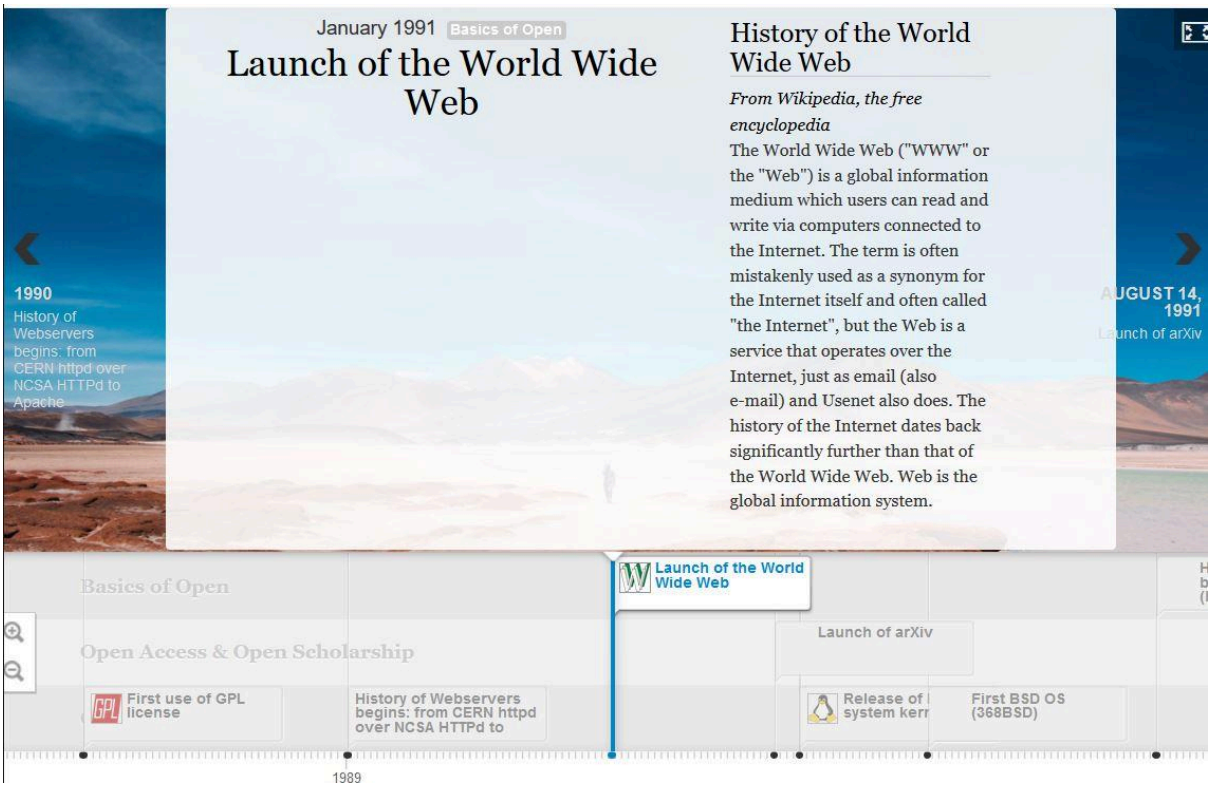

FIGURE X - Timeline: Parallel Evolution of Open Source and Open Scholarship (see https://blog.flavoursofopen.science/timeline/ for interactive version)

A key component uniting both Open Source and Free Software is that people, in principle, have a natural right to share, re-use and modify source code, and should therefore be clearly granted this right. This is different from the right to contribute, and is legally enforced through explicit conditions around licensing and copyright, embedded within the Open Source Definition. Open Source focussed more on the development/production side with a more effective bazaar model (Raymond 2001) for collaborating, which does not explicitly guarantee the Free Software movement's fundamental freedoms. The bazaar is considered an almost utopian, decentralised form of democratic organisation in contrast to the typical cathedral of centralised production. Stallman summarised this divergence as “Open Source is a development methodology; Free Software is a social movement,” also often alluded to in the widely known slogan from Stallman, ‘free’ as in ‘free speech,’ not as in ‘free beer.’ This is equivalent to free as in ‘libre’ rather than free as in ‘gratis’, in many languages; or ‘to be free’ versus ‘to be for free’. The acronyms FOSS or FLOSS (free/libre open source software) are able to encompass all aspects of the Open Source and Free Software movement; nevertheless the shorthand Open Source, without the moral load associated with the word “free” is often the standard term used to refer to this class of software, and is now ubiquitous in the global computing landscape — including within scholarly research.

The interwoven histories of Open Access and Open Scholarship

The early history of the Internet involved a number of academics based at either private or public research institutes (e.g., Tim Berners-Lee, Grace Hopper, Elizabeth Jake Feinler, Sally Floyd, Lawrence Lessig, and Linus Torvalds), suggesting that the ideals that underpin much of FOSS have their origins from within scholarship itself. Others suggest that “Open Science” and the idea of scientific knowledge as a common good has its origins as far back as the late sixteenth and early seventeenth centuries (David 2004; 2008; Martin 2019; Shapin 2011). While it is difficult to identify whether Open Source or Open Scholarship came ‘first’ due to their dynamic histories, both draw from the notions of common goods (and specifically digital common goods that are non-rivalrous) and unrestricted participation and sharing based on the underlying values of equity and sustainability, among many others. Compared to FOSS, Open Scholarship is much more complex; and not just in terms of its convoluted definition (Mirowski 2018). The former only relates to software development (including infrastructure, governance, legal issues and funding), whereas the latter regards the entire process of scholarship, including research articles, grant proposals, data, software, educational materials and methods (Katz et al. 2018; Tennant et al. 2019); sharing of which is now widely implemented and often even mandated within many areas of scholarship.

Scholarship, scholarly identities, and the politics and process of scholarly communication have a rich, non-linear, inconsistent, region- and discipline-specific history (Csiszar 2018). The typically institutionalised forms that dominate modern western scholarship (Shapin 2011), shifted towards ‘citizen science’ in the 19th century. This was followed by the explosion of global scholarship in the post Second World War years, leading to major

changes, such as an increase in the growth of the higher education sector, diversification of academic roles and career paths, an increase in international participation in research, and an increase in the publication of academic outputs (Fyfe et al. 2017). All of these factors led to an evolving relationship between 'the public' and scholarship (Geiger 2017; Neylon 2018). Scholarly practices of sharing precede the origins of scholarly journals, based around early forms of scholarly communication, such as witnessing each others' experiments and later, the exchange of letters between scientists, as the first form of so-called 'virtual witnessing' (Shapin 1984). However, these forms of sharing were highly exclusive and catered to a small and local elite only. Over the course of the centuries following the scientific revolution, the 'public' character of science changed primarily due to a shift in who is considered to be included in that 'public'. This shift was informed by an evolution of the 'social contract of science' (Gibbons 1999), that increasingly placed science inside society, as opposed to esoterically hovering above it. This also meant that many more people became legitimate members of the public of science, whose voices were considered worthy of inclusion (Gibbons et al. 1994).

Regarding the dissemination of scientific knowledge, perhaps the biggest innovation in the history of scholarly research was that of the printed journal. Developed in the mid-17th century, journals and research articles became the primary vehicle for communicating research, going beyond personal communication via letters. This extended the physical college into a virtual college (Csiszar 2016; Sztompka 2007), coinciding with and largely based around emerging scientific practices (i.e., working collaboratively to advance knowledge) (David 2004; Lahti, da Silva, et al. 2017). Journal articles then continued to evolve, ideally containing documentation of a research process, with the intention that others could either build upon, or repeat, that research. The journal became, and still to a large extent is, the primary source for peer-to-peer learning within scholarly communities, and also has a role in defining specialized scholarly communities (Potts et al. 2017; Hartley et al. 2019). For many STEM research disciplines, the journal and journal articles within remained largely unchanged as a communication vehicle until the late 20th Century, despite the advent and growth of Web-based technologies that permit more effective methods of sharing and dissemination, greater epistemic diversity in scholarship, and increased democratisation of academic knowledge (Hartgerink and van Zelst 2018; Hartgerink 2019; Tennant 2018b; Cowley 2015). This appearance of new technologies, enabling low-cost and instantaneous sharing of research findings (Martin 2019), led to the rise of the modern Open Access (OA) movement, which, similar to FOSS, focuses on providing unrestricted access to research articles. The motivation here was around helping to increase the global/public circulation of research information, and not just for professional scientists (Adcock and Fottrell 2008; Laakso et al. 2011; Tennant et al. 2016). While OA undoubtedly has its origins as a grassroots, community-led concept as an alternative or complement to typically subscription-based commercial publishing ventures (Willinsky 2005), it has now become widely adopted at a policy level by universities, governments, and funding bodies; often aligned with the impact or 'excellence' agendas of these institutes (McKiernan 2017; Katz et al. 2018; Moore et al. 2017; Finch et al. 2013).

However, these more recent developments remain in direct tension with many traditional methods of scholarly publishing, in which commodification of research has become the *de facto* norm for much of the industry; an industry in which some players have even taken

active stances to subvert, co-opt, or even stop progress towards a more open culture (Posada and Chen 2018; Tennant 2018a). This tension appears to be more focused in North America and western Europe, with some regions such as Latin America seeming to have a much stronger imperative to have scholarly research produced there remain accessible for the betterment of wider society (Packer 2009; Debat and Babini 2019; Alperin and Fischman 2015; Ochoa and Uribe-Tirado 2018). Therefore, OA and Open Scholarship cannot be considered as isolated 'movements', but instead they intersect with a complex array of questions, such as who should 'own' scholarly knowledge, who provides its legitimacy, and whether or not scholarly processes and outputs should be able to be privatised or commodified (e.g., through patents). This plurality helps to frame the complexity of the present scholarly system, especially with respect to synthetic modern understandings of openness, and therefore to present the scope for which an integrated understanding of FOSS culture can be potentially leveraged. This in turn could shape future changes to the definitions and culture of FOSS at an international level.

A resurgence in 'open research practices'

Particularly during the last three decades, notions of 'openness' and 'open' have seen a surge in attention among a variety of stakeholder groups. Over the years, a wealth of research in the movement's sub-fields as well as cross-disciplinary studies have proliferated. While undoubtedly having a much longer and richer history, the onset of what can be considered a modern understanding, or new wave, of Open Scholarship can probably find its origins within OA. Taking the *Budapest Open Access Initiative* as a point of genesis (BOAI 2002), a wide array of stakeholders ranging from technologists and librarians to researchers and educators have established their individual and peer group's perspectives on the multifaceted notions of 'openness'. Policymakers and research funding agencies on local, regional, national and international levels took note of the emerging trend and began to issue reports, guidelines, and policy documents that further shaped the conditions of openness (Moedas 2015; Union 2019; Participants of African Open Science Platform Strategy Workshop et al. 2018). As others have noted, the agenda of 'open' is "extending [...] into all areas and activities of the academy, impacting its core missions of teaching and research, as well as the systems and processes that are critical to individual and institutional success." (Corrall and Pinfield 2014, p.293).

In principle, OA is highly distinct from traditional, subscription-based publication. Subscription-based research articles have a strange duality in which they are published, but are largely inaccessible to the public due to copyright and licensing constraints, as well as paywalls. Although the ideas behind OA certainly pre-date the advent of the Web, the potential of digital technologies and their associated workflows catalysed the widespread development of OA. The widespread development is facilitated by the relatively low cost and increased efficiency of copying, transmitting, and storing documents for Internet-based distribution relative to the print era, overcoming resource-based constraints around production scarcity. As well as unconditional free public availability for reading, one fundamental aspect of Open Access has been re-use with attribution; often equated with the Creative Commons Attribution (CC-BY) license (Tennant et al. 2016). As citation or attribution is a norm in scholarly research, this requirement is not typically considered to be a

restriction and is fully compatible with standard practices. This basis for OA highlights unrestricted re-use as a critical factor underpinning OA (although licensing specifics between both domains of OA and FOSS do vary).

Accompanying this, a range of 'open research practices' are now commonly called upon and heralded as potential methodological improvements to issues around reproducibility, publication bias, questionable research practices, and more efficient research workflows (Levin et al. 2016; Bowman and Keene 2018; Crick, Hall, and Ishtiaq 2017; Fraser et al. 2018; McKiernan 2017; Fanelli 2018; Masuzzo and Martens 2017; McKiernan et al. 2016), including practices such as pre-registration and registered reports (Nosek et al. 2018; Nosek and Lakens 2014), sharing of code and data (Masuzzo and Martens 2017; Levin 2015; Mons 2018), and opening up the peer review process in different ways (Tennant et al. 2017; Ross-Hellauer 2017). However, we must also be careful not to attribute too much power to openness. Sociologist of science Harry Collins, already demonstrated that sharing everything that can be shared is still insufficient to produce, for instance, replicable research. There is a lot to science that cannot be made explicit and thereby cannot be shared on paper or digitally; known as tacit knowledge and tacit expertise (Collins 1992; Derksen and Rietzschel 2013; Penders, Holbrook, and de Rijcke 2019).

Regional variations in uptake of open research practices are noteworthy. In Latin America, "Open Access" (aka Ciencia Abierta/Acceso Abierto) seems to have become more widely adopted with the advent of SciELO in 1997 (Packer 2009) and Redlayc in 2002 (López et al. 2008), while other regions, including Europe, have a more diverse and fragmented history of making widespread OA in academia the norm, with correspondingly diverse economic and social tensions (Barić et al. 2017). For example, while many western European countries and much of North America seem to be struggling with the political and economic dimensions of OA (Johnson 2019), other countries, such as Croatia, have strong national and governmental support for a 'diamond' OA model, and are very much demonstrating leadership in this space more akin to progressive initiatives across Latin America (e.g., AmeliCA) (Debat and Babini 2019). To overcome such heterogeneity, the European Commission (EC) issued an OA pilot in 2007, which became a mandate in 2013 when the Horizon 2020 Open Research Data pilot (mandate with opt-out option) started. In parallel, to advocate, support and monitor such actions, the EC has established and invested in the OpenAIRE infrastructure (a legal entity as of 2018) as a means to implement OA and Open Science principles in Europe and beyond. In 2018, the Technical Advisory Board for the African Open Science Platform released a strategy paper on how "to put African scientists at the cutting edge of contemporary, data-intensive science as a fundamental resource for a modern society" (Participants of African Open Science Platform Strategy Workshop et al. 2018). Similar to these research and technology-focused developments, Open Education programs, for example, has also seen a substantial rise during the last 10 to 15 years; see the OERworldmap. These examples are not exhaustive, but serve to illustrate the fact that Open Scholarship, in its various forms, has now become a global phenomenon, though it remains strongly heterogeneous at a regional level (Debat and Babini 2019; Tennant, Beamer, et al. 2019).

Contested understandings of Open Scholarship (and 'Open Science')

Openness in scholarly research is now an evolving spectrum of interrelated principles, values, and practices, rooted in the idea of unrestrained participation and the Commons (cf. e.g., Suber 2007). Now, Open Scholarship is often used to refer to a variety of processes and practices, such as sharing of research outputs, different modes of communication, the research process itself, or increasing reproducibility (Tennant, Beamer, et al. 2019; Nosek et al. 2015; Moore 2017; Packer 2009; Debat and Babini 2019). Some researchers now even self-identify as 'open scholars' or 'open scientists' as a result of this. It is noteworthy that these diverging extensions of the web of meanings and corresponding applications neither happen in congruous, simultaneous and/or uniform manner, nor do they encompass the same object of identification. It also is unclear what specifically identifies one as an 'open scientist' or 'open scholar', what the differences or similarities between the two are, who or what defines 'inclusion' (or exclusion) within the wider community, and who has the moral authority to define what the values underpinning this (and Open Scholarship in general) are.

The variety of existing "open movements" have each undergone an evolution in terminology. One case in point is that of Open Science. During the 2000s and early 2010s, Open Science, as Titus Brown states, could be understood as "...the philosophical perspective that sharing is good and that barriers to sharing should be lowered as much as possible. The practice of Open Science is concerned with the details of how to lower or erase the technical, social, and cultural barriers to sharing" (Brown 2016). Such a notion can also be discerned in earlier readings such as Michael A. Peters' timely review of existing Open Science approaches (Peters 2010), or the now often-cited and foundational "Five Schools of Thought" of Open Science by Fecher and Friesike (2014). A sixth school was recently proposed as an addition here, community and inclusion, noting that the principles of diversity and inclusivity are key and should be embedded in notions of Open Scholarship (Tennant, Beamer, et al. 2019). There is no widely known 'Free Science' social movement, akin to Free Software.

During the last few years, though, Open Science has come to be often understood in a much narrower sense; although see Lahti, Silva, et al. (2017). While earlier notions of Open Science emerged as an "umbrella term that encompasses almost any dispute about the future of knowledge creation and dissemination" (Fecher and Friesike 2014), the use of "Open Science" in most recent literature usually remains limited to the domain of STEM-based research-related practices, technologies and tools, and its underlying ecology of the production, dissemination and reception of knowledge from a research-based point-of-view. Such a focus on STEM-based research, implicitly excludes, for example, Arts, Humanities and Social Sciences (HSS) (Eve 2014), as well as the different roles that researchers have as educators and communicators, and an underlying open ideology of openly sharing knowledge beyond research practices. This fragmentation becomes imminently visible in one of the most commonly-used graphical representations of Open Science from the EU-funded FOSTER project (FIGURE X). Here, Open Source is significantly demoted to being a 'third level' element within Open Science, rather than of equivalent importance to OA or Open Data. Any public element is also almost entirely

absent, for example via Citizen Science, which other organisations such as the OECD include within their classifications of Open Science. Furthermore, the linking of the various 'open concepts' together in a hierarchy exposes (not just pictographically) that there is very little horizontal connection between them. Instead, this represents a fairly fragmented landscape within the 'umbrella' of Open Science. Also, it seems to be a linguistic particularity which is mainly confined to English and French speaking academe, since in many other languages, terminology employed to denote "science" (e.g., "Wissenschaft" in German or "Wetenschap" in Dutch) is more general than in English (Tennant et al. 2019, sec. 4.2.2).

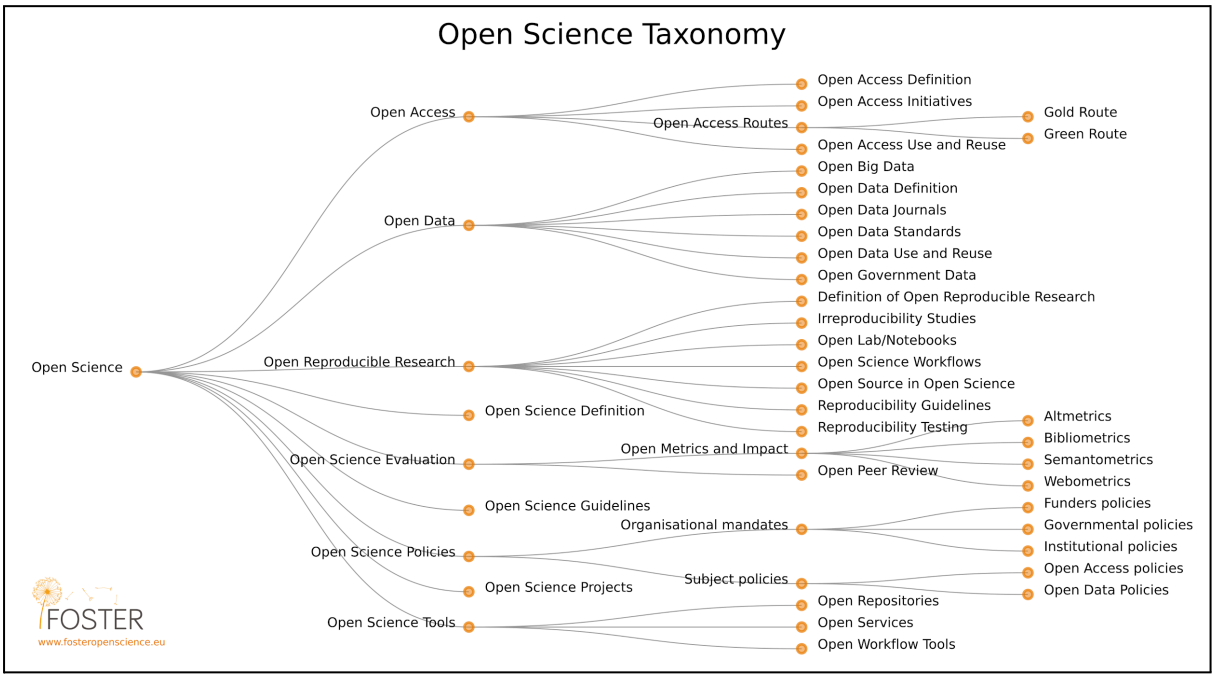

Figure X. An Open Science Taxonomy, from FOSTER (Pontika et al. 2015).

Emphasizing this evolving notion of Open Science, Vicente-Saez and Martinez-Fuentes (2018), as part of a systematic review where seventy-five studies were analyzed, define it as "… transparent and accessible knowledge that is shared and developed through collaborative networks", emphasising sharing and the collaborative process of knowledge production. Watson (2015) takes this a step further and views these aspects as no different from *good* science, akin to arguments that FOSS provides *better* software. Watson here also states that "software should be (and in fact is) driving the open-science movement", clearly holding the position of software in higher regard than that perhaps that implied by FOSTER. Thus, the general consensus appears to be that Open Science is a commitment towards specific practices that lead to the public sharing of research outputs, and that this is a *good* thing, and thus implicitly reinforcing the philosophical basis for openness; something fundamental both to FOSS and for earlier advocates of Open Science (see, for example, (Peters and Roberts 2010). However, even in places where such things are hinted at in Open Science texts, terms such as 'liberty' are still used in their practical, rather than philosophical, sense (Frankenhuis and Nettle 2018); either for personal or a wider public benefit.

Taking these described tendencies as a line of further evolution, more and more scholars acknowledge the similarities and overlaps in the network of scholarly practices

encompassing research, education, communication, and engagement. These practices have an underlying ‘open mindset’ against the backdrop of digitization and its impact on academia, thus re-transforming the understanding of Open Science towards what often is termed “digital” and/or “Open Scholarship” (Peters 2010; Peters and Roberts 2010; Veletsianos and Kimmons 2012; Peter and Deimann 2013; Eve 2014; Cronin 2017; Pomerantz and Peek 2016; Weller 2014). Katz et al. (2018) assert, “Open scholarship is perhaps the most broad term we can use; it includes open science, open humanities, and open research, and can be defined as opening products such as articles, data, software, educational resources, or more broadly, opening the process of scholarship.” Arguing for an integrative and inclusive framework, Veletsianos and Kimmons define Open Scholarship as taking

> “three major forms: (1) open access and open publishing, (2) open education, including open educational resources and open teaching, and (3) networked participation, [...which] refers to scholars’ uses of online social networks to share, critique, improve, validate, and enhance their scholarship.” (Veletsianos and Kimmons (2012, p. 168)).

Most recently, Tennant et al. (2019) echoed this understanding as a framework in a living strategy development document for Open Scholarship, defining it as broadly referring “to the process, communication, and re-use of research as practiced in any scholarly research discipline, and its inclusion and role within wider society” (Tennant et al. 2019). However, a single, unified, comprehensive and widely-accepted definition probably is not sufficient here (or even desirable), given the enormous complexities and diversity involved. Instead, Open Scholarship, Open Research, and Open Science might best be thought of as overlapping/intersecting 'boundary objects' (Moore 2017), indicating a range of principles, practices, and values, as well as motivations and common goals. Such might be preferential in helping to avoid compounding issues further with legal and technical aspects associated with software, data, and other scholarly outputs.

Furthermore, ‘openness’ in scholarship is often used while discussing varying scales (e.g., lab groups or teams, individuals, and institutions) (Tennant et al. 2019). Thus, ‘Open Scholarship’ is often applied indiscriminately to mean very different things, but with the same blanket terminology applied. Linked to the diachronic shift of meaning of Open Science away from the more philosophical underpinnings of the open movement in academia as discussed by, for example, Peters and Roberts (2010), and towards a narrower frame of reference, the widespread and often-synonymous usage of Open Science and Open Scholarship has added confusion. Such usage of these terms has several problems. First, the terms can be off-putting (or exclusive) due to the often dogmatic or purist way they are discussed (in particular on social media) as a seeming solution to all of science’s problems. Very few researchers or ‘advocates’ have been critically reflective on *how* such sharing and transparency are implemented, and on the potential unintended consequences this might have for different sub-communities (Tennant et al. 2019, sec. 7; Bacevic and Muellerleile 2018). Second, it means that people often end up having the same conversations about very different things, or simply talking past one another. An example of this is apparent with peer review, where there are more than one hundred existing definitions of ‘open peer review’ with many competing definitions; e.g., with open identities and open reports (Ross-Hellauer 2017). Third, the language of openness may be deliberately co-opted for non-open purposes if not strictly defined. The term “openwashing” refers to practices considered deceptive

insofar as they purport to be open but do so only to make themselves more attractive and in reality do not adhere to a majority of the principles of openness (Watters 2014; Weller 2014). This term is derived from “greenwashing”, where vendors rush to label their products “green” in some way, to increase sales.

These three issues relate to the broader problem in defining ‘open’ and who has the authority to determine this definition of “open” in various contexts, and across a hugely complex array of disciplines, cultures, and stakeholder groups. The foundational Open Definition, from the Open Knowledge Foundation, defines it as “Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).” However, this is not legally encoded, or enforced should the definition be subverted; for example, if licensing constraints prohibit re-use or modification via non-commercial or non-derivative clauses. Another issue is whether such definitions should ultimately remain static or flexible to wider changes in social culture. This problem is more universal, and can be said of both FOSS and Open Scholarship. One consequence of this lack of precision (or strict adherence to definitions) is that it has enabled proprietary actors, who arguably the Open Scholarship movement started at least in part as a ‘rebellion’ against, to now adopt the term as part of their public relations, marketing, and lobbying strategies, as part of what Mirowski (2018) labels as ‘platform capitalism’. Thus, there remains at least a partial need to rigorously define and enforce the values and principles of Open Scholarship, and how these underpin the practices, and to have these widely accepted by the scholarly community; however, it might already be too late to prevent the co-opting of this space.

The absence of sustained and integrated education and alignment of the specific elements of Open Scholarship has fragmented the idea of openness and made it ambiguous. It is also problematic in that the lack of common understanding has impeded the widespread adoption of the strategic direction and goals behind Open Scholarship. This prevented it from becoming a true ‘movement’ and separated researchers into disintegrated groups with differing, and often contested, definitions and levels of adoption of openness (Tennant et al. 2019). A potential consequence of this is that the future of Open Scholarship remains insecure, potentially and inadvertently increasing the strength of commercial players in this space, and becoming entirely divorced from any sort of value- or philosophically-driven basis. There are a number of additional potential contributing factors to this too, including inadequate resources and governance structure, lack of coherent communication, and the fact that researchers have an incredible workload to deal with already, and thus less time and energy to dedicate to challenging faulty parts of the scholarly ecosystem. One potential solution to this is to foster a deeper sense of community within Open Scholarship, sustained by bi-directional learning and integration with the FOSS community. This would help both to be more flexible and adaptive to changing external factors, as well as provide more cross-sectorial support and training. Communities of practice can be sustained, based around the different core elements underpinning open research practices, including reproducibility and data stewardship or management. However, this would require a commitment by research funders and institutes towards financially supporting the growth and maturation of such communities; something which is almost entirely absent at the moment in spite of the vast sums spent annually on global research and development.

What aspects can be transferred between FOSS and Open Scholarship

FOSS and Open Scholarship as communities of practice around sharing

The software and research communities are both vast and diverse. Within FOSS, a key and defining component is their community (Katz et al. 2018). Here, software development and re-use is inherently communal and ideologically democratic based on its core methodologies, and self-organised either hierarchically or modularly. Version control systems allow developers to independently make changes to a copy of the code and share the changes with collaborators. Remarkable momentum towards open collaboration models has been provided by Git and GitHub. This modern, distributed version control system allows decentralized development of large-scale FOSS projects. It was initially developed by Linus Torvalds to facilitate Linux kernel development and has subsequently been widely adopted by the FOSS community. While harder to learn and use than earlier popular systems, such as SVN and CVS, Git similarly facilitates collaborative workflows, and also adds additional support for decentralized development. This allows developers to copy the initial code repository, work independently on a fork of it, and then share the changes or the full changed version on platforms such as GitHub, GitLab, Bitbucket, or Sourceforge, leading to the rapid popularisation of such practices.

At least part of the success of FOSS was that the methods and products created were proven viable alternatives, practically and commercially, to more proprietary software. Another major part is that, in theory, anyone who shows an interest is able to collaborate and contribute code, write documentation, add examples or tutorials, discuss, give feedback, and more within a project that is generally receptive to this. What this means is that, even in expert-driven and software-intensive projects, there is a low barrier to entry (particularly when using Web interfaces to the version control system) even for those with non-coding skills, as there is substantial value in, and appreciation for, generalist skills in documentation and communications. It also helped to divide up the labour required for larger projects, opening up new working environments and group formation structures (Palazzi et al. 2019). FOSS can also be thought of as a powerful method for social, community-driven or peer-to-peer dynamic learning, where the distinction between users and developers is more blurred than in a proprietary environment. This leads to the emergence of new governance structures around communities based on a shared sense of collective participation and authority (O'Mahony and Ferraro 2007). These modes of collaboration are essentially continuous, iterative, and progressive methods of verification. The underlying assumptions of this type of process are very much idealised: public disclosure of bugs and issues that get resolved sequentially, usually by a core team plus external volunteers, until a piece of software is released, with the principle of many eyes helping to increase the validation and verifiability of work. A consequence of this is a further assumption that this process also

results in *better* software (Kelty 2001); or an alternative viewpoint that perhaps only 'better' software is worth being shared under an Open Source license (Pauliuk et al. 2015).

Much of this sharing and collaborative culture was driven by reciprocal motivations around sharing and repurposing, also sometimes called 'hacking' (Orr 2006; Coleman and Golub 2008). When applied specifically to software, hacker culture leads naturally to FOSS as a community of practice, as the freedom to re-use and modify (i.e., 'hack') is built into the software through licensing requirements. Raymond puts hackers at the heart of a gift-giving culture:

> "You become a hacker when other hackers call you a hacker. A "hacker" in this light is somebody who has shown (by contributing gifts) that he or she has technical ability and understands how the reputation game works. This judgment is mostly one of awareness and acculturation, and can only be delivered by those already well inside the culture." (Raymond 2001, 31)

A 'gift-giving' culture strongly emphasizes the ideology that underpins the focus on "community" for FOSS, in that sharing is intrinsically valued as part of wider hacker culture, in which reciprocity is key[2]. It is essential that scholars understand how to apply the successful collaborative practices from open software development, especially now that much scholarly research is becoming more and more reliant on computation. This has the potential to improve the efficiency, diversity, productivity, and reliability of research processes, decoupled from any formalised journal-bound peer review system (Frassl et al. 2018; Heller, The, and Bartling 2014; Tennant, Bielczyk, et al. 2019).

Here, however, it is important to clarify some important differences between participating in FOSS and Open Scholarship activities. These differences lie in the process more than the product. Let us start with end products. Granted the adoption of similar licenses (e.g., those that are Open Definition compliant), there is no substantial difference in the openness of OA,

[2] A study in 2002 (N=81) showed that participation in FOSS can be broadly divided into internal factors that are not easily influenced, such as intrinsic motivation and altruism, and external rewards, such expected future returns, skill expansion, and personal needs (Hars and Ou 2002). Additional factors included peer recognition and self-marketing (both extrinsic), and community identification (intrinsic). Contrary to the proposed stance of openness for the greater good. Interestingly, only 16% of the participants rated high on altruism. Another survey based on closed and open-ended items of (N=110) FOSS contributors (Baytiyeh and Pfaffman 2010) revealed that intrinsic motivation factors, including the belief to be working for the greater good, providing something valuable to others, and the belief of helping others are the strongest factors motivating individuals. These motivators are followed by extrinsic (external) human capital factors such as learning new skills, tools, strategies, and methods, and factors related to increasing own academic and personal success. A further study (Bagozzi and Dholakia 2006) of (N=402) active Linux User Groups participants, who completed a survey based on the Theory of Planned Behavior and the attitude-theoretic model, found that the strongest factors to contribute with active participation to FOSS are social identity (identification with the movement, cognitive and affective social identity), attitudes, and perceived behavioral control (i.e., the perceived difficulty in actively participating given norms and rules).

Open Data, and FOSS. End users are typically able to access the end products without any restrictions, and, according to the license, they are able to derive works from these products as well as redistribute them. There are, however, residual issues in that just because something is 'open' does not make it accessible and immune from other inequities and ethical concerns that might influence its access and usability (Nicol et al. 2019). On the process side, it is the way products are openly available for collaboration during development that differs. With FOSS, it is typical that a small, core group of people starts openly developing a software product; however, this depends on the scale of the project, and it can range from individual, networked, to large corporate or governmental ventures. Other people can join (and leave) the effort over time, contributing as they see fit. Development iterates over several cycles and releases, and the product is openly available from the very beginning. The process is very much idealised, and best practices can include aspects such as using open repositories, version control, and issue trackers; but such collaborative practices are neither strictly inherent to FOSS, nor are they limited to it. The production of knowledge goods that is based on sharing both resources and outputs, performed by loosely connected individuals who collaborate in a decentralized, modular and often non-hierarchical way has been identified as a new way of organizing production. Yochai Benkler coined the term commons-based peer production for this mode of production and identifies projects such as Wikipedia and OpenStreetMap as further peer production examples. Through this lens, it becomes clear that the use of open licenses might be a necessary condition for the success of peer production endeavours such as FOSS, Wikipedia and OpenStreetMap, but it is not sufficient. Rather, a community of contributors is needed to sustain such efforts.

Distinct to this, for Open Scholarship activities the process is, most of the time, closed - as with more traditional ways of conducting research. Exceptions do exist where the actual outputs from a research project are software or models (e.g., Astropy, NEURON). Scholars can join a collective effort, produce several research artifacts, and open them up at the time of publishing (i.e., the process is closed, but the products are open) (Masuzzo and Martens 2017). Although new approaches have been proposed lately, such as post- and open-publication peer review where authors publish a "to be reviewed" version of their manuscript and its revisions (Ross-Hellauer, Deppe, and Schmidt 2017; Tennant et al. 2017), we mostly still only see snapshots of research artifacts in the open, not their development[3]. One example is where the process itself is more exposed, for example by using electronic laboratory notebooks (Myers et al. 1996), or registered reports (Nosek and Lakens 2014). These are still not as open to participation or collaboration as a FOSS workflow, but provide more of a window into the process of the research itself. Recently, the concept of a 'Massively Open Online Paper', or MOOP, was proposed to attempt to reflect a FOSS-style workflow within scholarly manuscript writing (Tennant, Bielczyk, et al. 2019), presenting new collaborative opportunities for peer-to-peer learning and interactive knowledge generation. Such exceptions aside, Open Scholarship in its current form is more similar to proprietary

[3] One exception to this is many papers where the first author is involved, where calls for research happen openly in social media websites and the resulting papers are available from the beginning in publicly shared documents (e.g., as was the process for the present article). Such a process of collaborative authoring and editing is now simple and popular using Google Docs (or other similar applications), which was enhanced when the Open Source reference manager Zotero was integrated with it.

software companies who decide to render their products Open Source. The willingness to participate in FOSS development as well as Open Scholarship in general seems to be more around the sharing of common beliefs and attitudes.

Regarding the motivating factors for Open Scholarship, it seems that intrinsic motivations might play a stronger role than extrinsic ones. A survey of 300 participants (Oreg and Nov 2008) among FOSS developers and Wikipedia authors (that is, both code and knowledge contributors) revealed that software developers tend to place greater emphasis on reputation and self-development motivations, whereas knowledge contributors place greater emphasis on altruistic motives. Yet, extrinsic factors should not be underrated, in particular those related to human capital; this is especially the case as, unlike much of FOSS development, in many fields a large quantity of research is funded by the public. This presents a strong moral case for making the outputs, at least, of research publicly accessible. Once it is empirically demonstrated that Open Scholarship activities build stronger peer networks and help building one's reputation and resume, extrinsic factors (e.g., promotion processes) might encourage otherwise less motivated scholars (see the section below on systems of valuation in openness). The Open Access citation advantage is a powerful example of such (Norris, Oppenheim, and Rowland 2008; Swan 2010; Wagner 2016; McKiernan et al. 2016).

The extent to which intrinsic factors are more powerful than extrinsic factors remains yet to be fully illuminated for the wider case of scholarly knowledge diffusion. The Open Scholarship community is responsible for being aware of the social contexts surrounding knowledge transfer (Morgan et al. 2018; Gibbons 1999), and use this to create the best cultural environment and shared values possible to attract these motivations of scholars. Given some effects of openness such as wider readership and access (Levin 2015), boosting collaboration and transparency (Peters and Roberts 2015), democratisation of science (Brown 2009) and promoting accountability (Resnik 2006), one would like to think that all stakeholders involved in the process of science communication and research publication should equally be in favor of this trend. However, that is not always the case. As outlined in the Foundations for Open Scholarship Strategy Development document (Tennant et al. 2019), a number of challenges and threats to this ideal still remain. This includes acknowledging the limitations of ‘openness’, the appearance and inclusivity of the community, and possible disincentives to sharing and open collaboration. There are also direct threats that openness brings to be mindful of; for example, potential mis-use of publicly shared genomic data to discriminate against marginalized communities. However, much of the existing research indicates that Open Scholarship is more aligned with fundamental incentives for researchers (e.g., visibility, collaborations, reputation, career advancement, grant awarding) (McKiernan et al. 2016). Therefore, there is space for additional insight into the motivators and barriers for the uptake of open research practices.

Leaving the classic self-determination theory of intrinsic and extrinsic motivation aside, other studies have also shown how social identity and altruism might play a role in engaging with Open Scholarship, but also that tool support is important. An exploratory study of self-archiving practice and inhibitors among computer scientists (Graziotin 2014) found that authors often infringe the copyright transfer agreements of the publishers when self-archiving and stop self-archiving when they discover they are infringing. The study calls for tools from the FOSS community for facilitating author-based self-archiving that take authorship rights

into account. Lack of automation and time required were found to be the strongest factors inhibiting self-archiving of publications and sharing of datasets. Similar research has identified widespread copyright infringement of scholarly articles (Green 2017), and there are a number of ongoing legal cases in this space (Manley 2019), indicating that more awareness and support are needed for researchers regarding the legality and restrictions associated with licensing and copyright of scholarly works.

All together, this suggests that there are complex barriers to enter an Open Scholarship mindset and to take up Open Scholarship practices. Open activities carry norms and culture that some individuals might find hard to understand even before engaging with them. This would seem to manifest itself in a divergence between the attitudes towards (generally positive) and practices of Open Scholarship (Zhu 2019; Smith et al. 2017; Ross-Hellauer, Deppe, and Schmidt 2017). What emerges from this is that the culture is not simply binary (i.e., open versus closed), nor is it confined to the realm of software or scholarship. Openness can mean very different things in different contexts, reflecting values and principles, practices and processes, and outputs; which themselves are not discrete but form a complicated feedback loop with external and internal constraints. Although intrinsic motivation can be important, identifying with the community of practice is also key here (Eghbal 2016). Extrinsic motivators, such as money, prestige, patents, citations (i.e., 'incentives') may lead us away as a community from Open Scholarship practices. Research indicates, however, that extrinsic and intrinsic motivations are complex and interrelated. Ryan and Deci suggest that both types of motivation can be traced to basic human needs for autonomy, competence and relatedness (Ryan and Deci 2000). Such tensions between motivations and practices can manifest themselves between selective communication of knowledge, an inherent willingness to share, and the strange duality of scholarly knowledge often being simultaneously public and not publicly accessible (Levin et al. 2016; Hilgartner 2012).

It is contingent on the Open Scholarship community to develop and coordinate strategies for transitioning to OA, Open Data, and FOSS in academia (Tennant et al. 2019) and these strategies should encompass social and cultural issues. The lesson here for Open Scholarship may be that an inherent personal love of science and discovery must be nourished, as well as foundational ideals around 'hacking' and 'gift-giving', and communities that can affect the principles of Open Scholarship must also be cultivated around this. As we saw with the FOSTER taxonomy, these critical and fundamental aspects of scholarship are almost completely absent, more with a focus on communication of outputs. The taxonomy implicitly neglects the role that connectivity (i.e., cooperation and collaboration) might have to achieve greater efficiency, impact, and influence (Gunnarsson 2019). Therefore, in the future, it seems like it might be useful to think of Open Scholarship in two fundamental ways, as with FOSS. One element would focus on production and participation (e.g., collaboration through sharing), and the other based on dissemination (e.g., due to licensing constraints). This exposes two different viewpoints on what constitutes being 'fully open', while also embedding core values around equal and free participation, while implicitly excluding proprietary or exclusive systems.

Integrating FOSS into an Open Scholarship workflow

There now exist a number of practical manifestations of FOSS being built explicitly for supporting Open Scholarship. Open Journal Systems, created by the Public Knowledge Project and launched in 2001, is FOSS for the management of peer reviewed journals, and released under the GNU general license. As of 2010, around half of the journals it helped to publish were based in the developing world, showing the value of FOSS for leveraging global participation with Open Scholarship (Edgar and Willinsky 2010). The Collaborative Knowledge Foundation similarly offers FOSS that reduce costs while increasing interoperability and flexibility in the scholarly publishing process, while avoiding vendor lock-in associated with proprietary systems (McGonagle-O'Connell and Ratan 2019). These two examples represent major attempts to reform scholarly publishing and bring it into line with more Web-native functionalities (Morris 2009; Tennant 2018b), while also helping to secure the future of open and community-controlled scholarly infrastructures (Joseph 2018; Bilder, Lin, and Neylon 2015). Such developments are critical during a time when there is increasing capture and enclosure of vital components of the global scholarly infrastructure by private entities, which has further consequences on inclusive participation and epistemic diversity, particularly against researchers from the 'global south' (Posada and Chen 2018).

So, to what extent can the FOSS 'hacker culture' and Open Scholarship be further aligned as communities of practice? Hacker culture is based on the assumption that the best software is the best because it works, and fulfills its intended purpose. Other aspects can also include appreciation of efficiency or style of the code. Translating this to research practices is more complex, based on how we perceive its functionality. There are now a number of noteworthy examples of communities of practice coming together, and attempting to collaborate upon different aspects of Open Scholarship, irrespective of the seeming tensions with current incentives. They are leveraging rare traits within scholarship, such as open development, networked collaboration, strong, value-based frameworks, and inclusivity, as a sort of counter-culture to traditional research processes. Examples of this include The Carpentries, the Software Sustainability Institute, and indeed the broader research software community (Hettrick 2016). The fundamental elements of Open Education and FOSS seem to underpin developments in these initiatives, leading to values-based communities, open collaboration, re-use, and sharing, and cross-pollination of ideas and practices, and a truly interdisciplinary and inclusive approach to scholarship (Tennant, Becker, et al. 2019).

On a more practical level, many of the more social elements associated with FOSS culture could be adopted by Open Scholarship communities to help increase the efficiency and productivity of research, based around several key elements (Millman and Perez 2014):

1. Version controlled workflows, for tracking data collection, analysis runs, manuscript writing, peer review, and up-scaling collaboration;

2. Automated execution and continuous integration, where the code, data, and full computational environments are inherently reproducible, and proposed where possible in-browser or via peer-to-peer networks;

3. Testing, to detect and remove errors and bugs prior to peer review;
4. Readability with the so called ‘clean-code’ philosophy, where well written and documented codes are easy to access at different coding levels based on existing standards for self-documentation; and
5. Developing infrastructure capacity for tools and procedures essential to research workflows.

These benefits could help to define a successful new community model based around sharing, collaboration, and innovation. This development model is increasingly being adopted by the software industry who, among the many advantages, see it as an economic improvement. Importantly, parts of this community still have their roots from inside of academia, whereas many also exist outside of it and move at a faster pace. The ideas around openness are quickly spreading from academia to the wider coding community: software companies today are eventually selling their expertise more than the code or the code subscription. Users can freely access the code and verify the robustness of the latter, improve it for specific cases when necessary and make sure all bits are in place for better company performances. Not to be forgotten is the key aspect that code-sharing and adaptation skip, almost entirely, the problem of standards by letting anyone use a ‘common language’. One of the key aspects here that underlines the community is that research is a process that continuously evolves, and is therefore not captured and defined appropriately with just discrete outputs (i.e., peer reviewed articles). Research projects could contain numerous components and ‘releases’, akin to version control and FOSS, with an emphasis on the value of the process and documentation. This could create a tendency for projects to become more distributed, and collaboration to be recognised as a fundamental trait. The Open Science MOOC, for example, follows such an openly collaborative workflow (Tennant, Becker, et al. 2019).

Regardless of self-identified membership of the “Open Scholarship Community”, one could argue that performing any (or a combination of) the multitude of practices associated within Open Scholarship, contributes to its political and practical goals. Open Scholarship, as a practice, will thus greatly exceed Open Scholarship as a community. This does present problems, especially in the context of governance issues. Consequently, this is also where Open Scholarship falls short of being defined as a single, organic ‘movement’, as it appears to lack common, shared values, strategic coordination, and well-defined goals. This issue is similar to FOSS, in that democratic values and inclusivity appear to be core, underpinning elements to Open Scholarship, but it remains unclear if these are upheld in practice, and whether these practices are inherently ‘better’ for scholarship.

Fundamental principles and values

One fundamental element that FOSS, OA, and Open Scholarship all have in common is that they believe that the outputs of the wider process of organised knowledge creation, including intellectual properties that they represent, deserve to be a public, rather than proprietary, property (Willinsky 2006; Bartling and Friesike 2014). They are also based around the simple fact that existing work is there to be used (for what diverse purposes) and built upon, both as

developers and/or researchers “standing on the shoulders of giants”. For Open Scholarship, there has been relatively little explicit work into the underlying principles (with the exception of the dozens of declarations and charters aimed at OA), and especially regarding those that are universally accepted. A recent review of Open Science demonstrated that it can help to advance diversity, justice and sustainability in the psychological sciences (Grahe et al. 2019). Beyond this, OCSDNet is almost unique in this space, in that it proposes seven principles for Open and Collaborative Science as part of its Open Science Manifesto. Here, it outlines the importance of representation and inclusivity within Open Science, and the importance of these in challenging the core values of more traditional forms of science. Principle 6 also states: “It incentivizes inclusive infrastructures that empower people of all abilities to make, and use accessible open-source technologies”, thereby embedding FOSS within the principles of open and collaborative science. Similarly, the Vienna Principles for scholarly communication have a strong focus on Open Science, based around 12 principles including accessibility, discoverability, and reusability. Finally, the Scholarly Commons was designed to embed strong principles in the wider landscape of research communication, rather than with an explicit focus on Open Scholarship (Bosman et al. 2017); indeed, these principles do not mention Open Science/Scholarship at all.

However, within the Free Software movement, the principles of freedom were much more deeply encoded than they appear to be for Open Scholarship or Open Science, which both appear to be more similar to Open Source in their current focus on pragmatism, with the exception of the OCSDNet and Vienna Principles mentioned above. The result of this for software communities is that, by being based on essential freedoms, this helped to promote social aspects such as frictionless sharing and co-operation - factors that are increasingly important in a digital world. Often a moral basis is argued for OA around fundamental rights of access and re-use, akin to arguments made for democracy. Similarly, a parallel argument is that Open Scholarship is a more efficient or effective way of performing and communicating research than traditional forms. It's clear that these complementary arguments have not had the same sort of widespread impact as those within FOSS, the reasons for which currently remain quite elusive. Thus, although FOSS and Open Scholarship might have similar foundations (i.e., around ideas of free sharing and common goods), the differences in the pace of technological and cultural changes between the two mean that ‘openness’ has made more impact in the software world than in research.

There are a number of key differences here between software and research that might explain this distinction in uptake rates between open practices, and the translation of values and principles into practices. One possibility here is that the consequences of FOSS can be seen almost immediately in improved software, while observed changes in research due to Open Scholarship practices manifest themselves in less tangible ways or after a longer period of time (although see McKiernan et al. (2016)). Another possible cause is the difference in support and training for ‘open’ skills between the software and scholarly research environments. Much of the infrastructure and technology for scholarly communication has been outsourced to either publishers or other third-party vendors, rather than within research institutes or academies themselves. This means that hard skills required to learn how to do open scholarly communication are difficult to acquire from within academies, which makes translation of the practices from the underlying principles problematic. Additionally, there is the temporal duration; FOSS has become normative

across global software communities, whereas diffusion of Open Scholarship practices appear to be more recent in many aspects and develop at a slower pace. Furthermore, there is the core issue of incentives around open research practices. The strong core principles and values of Open Scholarship practise that lead to 'better' research, and more advantages for individual recognition, is still set against what is often termed a 'publish or perish' culture in modern academia. This tension pitches the values of high quality scholarship against other systems of value such as citation-based metrics or journal brands (McKiernan et al. 2019; Alperin et al. 2018), which are often in control of proprietary vendors and directly tied to career advancement or financial recognition. Finally, a key difference could simply be in the communication within the Open Scholarship community, and how this translates to the wider scholarly community. If Open Scholarship is treated as something distinct than just 'good' scholarship (Watson 2015), then this creates a divide and tension between different camps.

A relevant question to better understand these causes is whether or not open research is equivalent to 'better' research, based on either the values, practices, or principles. Does sharing code and the coding environment, being more inclusive, making the work OA, all make research better? What if it can be reproduced, because it solves a particular problem, leads to an innovation, or makes a big jump forward in a field (Sandve et al. 2013; Fanelli 2018)? This depends on several factors. Firstly, many of these common notions of openness do not translate well beyond empirical research disciplines, and can be better considered here as a 'boundary object' (Moore 2017). Critically here, it depends on the perception of what is seen as 'better' or 'higher quality' research. If sharing data and software environments, detailed and rigorous analytical/methodological protocols, and other elements can be considered to increase the rigour, reproducibility, and accountability for research then it is difficult to argue against openness increasing intrinsic research quality. Furthermore, adopting such practices allows individuals and communities to automatically adhere to the coupled principles and values of openness, and thus become included as part of the Open Scholarship community by default. If one considers science and research to be self-correcting, which often is assumed, then openness is necessary for others to repeat and verify, and thus, enable correcting to function at a system-wide scale (Ioannidis 2012). Open Scholarship as a process becomes decentralised, distributed, and a peer-to-peer process of production, based on strong scientific principles; rather than a competition for spurious measures of success. A follow-on question is, once this process becomes better defined for Open Scholarship communities, how do we make these traits relevant to the things that we value (i.e., incentives), and thus transform them into more widely-adopted behaviours and practices.

Thus, it appears there is a substantial place for refining the core values of Open Scholarship/Science, in particular around collaboration and participation; or perhaps even refining two separate 'camps' again based around different concepts. For example, Open Science already appears to be well understood to be more equivalent to Open Source in its pragmatic, process and output focus (Levin and Leonelli 2017). The OCSDNet and Vienna Principles seem to provide a strong basis from which to work on this, and to begin generating a more coherent discourse around open scholarly practices, challenging the issues with the present incentive system, and creating a more fair academic advancement system. Without such a foundation, it seems difficult to equate Open Scholarship or Open Science with a

formal movement on par with FOSS. Ultimately, this could be refined to be a more effective narrative based on the inherent motivators for different stakeholder groups; for example, how openness can catalyse technological innovation and economic growth (for industry and policymakers), or how openness can challenge unethical conduct while increasing the reliability of research through a more effective and collaborative creation method. In time, we would likely see such more effective communication, based around a combination of principles and practices, close the 'apathy gap' and resistance towards openness through appealing to a sense of common interests among stakeholder groups.

The role of commercial players

As we have discussed above, commercial players play an important role in both FOSS and Open Scholarship. In the latter, we have seen how they contribute to both our understanding and practices of openness, and maintaining the present system of evaluation and reputation. However, many of the commercial aspects of the wider scholarly communication ecosystem remain in direct conflict with any notions of openness. As a key example, traditionally many scholarly publishers operated around a subscription-based business model that has created an oligopolistic 'market' structure (Larivière, Haustein, and Mongeon 2015; Bergstrom et al. 2014; Shu et al. 2018), particularly based on anti-competitive practices such as non-substitutability, micro-monopolies, and non-disclosure clauses (Tennant and Brembs 2018; Armstrong 2015). This is clearly in violation of fundamental freedoms towards openness, as many of them actively prohibit sharing through mechanisms such as copyright and creating the illusion of artificial scarcity in order to preserve revenues. Such rights are mostly signed away by authors in exchange for publication, and for remuneration of these primary investors and value creators to be effectively nil, despite scholarly publishing being one of the most profitable industries that exists. A general lack of transparency has meant that the commercial elements of the system are able to maintain great power and control over the flows of knowledge and finances in scholarly communication (Björk et al. 2018).

This tension has been one of the driving forces behind the entire modern push for Open Scholarship, with demands for a new, modern, open scholarly infrastructure (McGonagle-O'Connell and Ratan 2019; Bilder, Lin, and Neylon 2015; Joseph 2018; Okune et al. 2018). Now, there exist two major camps: the non-profit side, comprising services such as those via the Joint Roadmap for Open Science Tools (JROST); and proprietary alternatives offered by publishers such as Elsevier and Springer Nature (via Digital Science) and service providers like Clarivate Analytics. A major criticism of the latter is that the services they offer around data and analytics represent their ownership of key elements of scholarly infrastructure that create lock-ins for users and their workflows (Campfens 2019; Posada and Chen 2018). The result of this has been that, despite the best intentions of OA and Open Scholarship, and the foundational principles upon which it has been based, much of the infrastructure it relies on has now been captured by commercial entities and subject to market control as a system of organised competitiveness (Anderson and Squires 2017; Odlyzko 2013). The scholarly publication process now primarily exists to reward and regulate a reputation economy (Hyland 2015; Lariviere and Sugimoto 2018), and which poses one of the greatest single threats to modern scholarship (Brembs, Button, and Munafò 2013).

Similar to FOSS, this represents a key problem that key elements of ‘open’ infrastructure do not receive adequate sustainable funding and commitment for essential projects and services (Chang, Mills, and Newhouse 2007; Chengalur-Smith, Sidorova, and Daniel 2010), something with which Open Educational Resources also struggles (Wiley 2007). Exceptions to this do exist, such as Mozilla, but often it is the case that many of the people engaged in FOSS and maintaining its infrastructure operate from within large commercial organisations, rather than being committed to such projects full-time. Within the wider realm of scholarship, we know that there is ‘enough money within the system’ to support an almost total ‘flip’ to a more open system, especially regarding OA (Schimmer, Geschuhn, and Vogler 2015; Björk and Solomon 2014). However, this does not appear to have manifested in a way that is disruptive, or affordable to many researchers and institutes around the world (Green 2019; Barić et al. 2017), and may now even be creating a second financial crisis, akin to the serials crisis, through hyperinflation of charges for OA and a lack of price sensitivity (Khoo 2019). Furthermore, this highlights the problem that such forms of ‘closedness’ are often more difficult to see because they are more about violating principles than any specific practices, with the unintended consequence that ‘openness’ can actually lead to new forms of enclosure (Tkacz 2012). It might be the case that substantial intervention is now required from public administration bodies, seeing as this involves the transfer of billions of euros/dollars of public funds each year (Di Cosmo 2006), in parallel with the mobilisation and strategic coordination of research communities around these issues (Tennant, Beamer, et al. 2019).

Systems of valuation in openness

At the present, few of the things which we might use to rationally define ‘better’ scholarship (whether ‘open’ or not) are what we seem to value in terms of how research is evaluated or rewarded (Alperin et al. 2018; McKiernan et al. 2019; Schimanski and Alperin 2018). Indeed, this divergence is probably one of the key factors underlying the fairly widespread, but complex, issues around reproducibility in research, as well as more ‘open research practices’. Here is where a number of problems bind and intersect under the umbrella of ‘Open Scholarship’, around the credibility and legitimacy of scholarly research, the relationship between prestige and peer review, the role of science in society, the incentive or reward system, and how communities form around these tightly woven concepts. For example, it remains eminently unclear whether communities form around practices, software, shared ideas and values, challenges, or other processes or objects of value. This presents a major challenge for academics (and academies) in how they can adapt and evolve into a system that is defined by fundamentals of Open Scholarship (e.g., re-use and collaboration). Indeed, such strategic implications around openness are only now just beginning to be explored within some communities (e.g., AI) (Bostrom 2017), and represent major challenges in defining ‘value’ in [open] scholarly research..

Perhaps one of the most critical issues here is the exploitation of many elements of the peer review system, in that it is often anonymous, unrewarded, and yet in high demand, and often in the control of commercial, third-party entities (Fyfe et al. 2017). This is further exacerbated by the fact that peer review, in its current journal-coupled implementation, is considered to be

a necessity for virtually all forms of scholarship, irrespective of the diversity of processes that lead up to it. It remains virtually unknown how such peer review operates as a quality control mechanism, whether it actually performs to the expectations with which it is regarded, or what actually even happens as part of the process; and yet, it remains almost universally highly regarded as the process which assigns most value to a piece of scholarly research (Tennant and Ross-Hellauer 2019). Other knowledge generation systems, including FLOSS and Wikipedia, have more continuous 'review' systems that allow for greater differentiation of value, and on a more granular basis, than a binary state of 'peer reviewed' and 'non-peer reviewed' that continues to defy (Tennant et al. 2017). Measures for reputation of reviewers are yet to find formalized expression, although some progress in this domain is being made (Bianchi, Grimaldo, and Squazzoni 2019). Given that traditionally, information about reviewers and their reviews are kept private by the publishing bodies, it is hard to gauge the reviewer quality from just their training and experience, or indeed the quality of the review itself (Callaham and Tercier 2007). There are already many journals out there making the reviewer information and their reviews more open (Bravo et al. 2019), which is an encouraging sign here.

Beyond peer review, any renewed understanding of Open Scholarship would seem to encourage recognition of a wider array of research 'products' rather than simply research papers. Indeed, in the five Open Science schools of thought put forth by Fecher and Friesike (2014), measuring scholarly impact is considered a major school by itself. However, it seems that congruent changes in reward or 'incentive' structures are typically lagging behind this at the policy, institutional, and practical levels (Levin et al. 2016). At present, the primary reward system for professional advancement in academia in many STEM fields seems to be focused on peer-reviewed articles published in reputable scholarly journals and conference proceedings, author position/order, and perceived value of a publication venue. For example, being published in a 'top' journal like Science or Nature can define the future success of one's career; however, by and large there is no equivalent in the Open Source world. A focus on articles as the primary output of scholarship implies opening articles would provide the most value within the research process. Alongside articles, assessing reputation in scholarly communication has traditionally relied upon metrics such as the 'Journal Impact Factor' (Lariviere and Sugimoto 2018), which is based on the annual citations to articles in the Journal. On the side of the author, the number of publications and citations are often converted to a 'h-index' (Bornmann and Daniel 2007). What is often mistaken in this context is that, while such journals and other measures of success or reputation might be of comparative use (simply because they can be measured), this does not actually make them a useful or objective measure of reputation or prestige itself. While the limitations of paper and journal-based evaluations are broadly understood, their use remains at the core of almost all issues and challenges facing social and technical changes around Open Scholarship (Brembs, Button, and Munafò 2013a). The continued use of these inappropriate metrics and proxies for value has led to a culture widely known as 'publish or perish', and represents a vastly unscientific approach towards research evaluation with a number of well-known deleterious effects (Brembs 2018). Social movements such as the San Francisco Declaration on Research Assessment (DORA) have attempted to implement a new broader form of research evaluation, and while approval of this seems to be increasing, active institutional adoption and support is still clearly required.

Due to this focus on journal articles and related measures, other essential components of research and academic cycles, including software, materials, teaching, infrastructure, public engagement, and data often remain unacknowledged or viewed as ‘second class’ value objects and processes. This is now also becoming codified into policies as a new norm within Open Scholarship, where OA to research articles has undoubtedly had the biggest focus in the last 3 decades (Piwowar et al. 2018). This is peculiar as research articles could not exist without many of these other things, and yet they are consistently undervalued in terms of socio-cultural appreciation, evaluation, and recognition for funding (Alperin et al. 2018).

Citation, in its rawest form, is usually an acknowledgement of an intellectual debt (Garfield 1979), and a hierarchical genealogy of ideas, dependent on a number of factors. Scholarly research again is peculiar in that citations only between articles seem to have any form of currency, and not between different networks of objects, despite the principle and practice being essentially the same. For example, principles for software citation now also exist as a reflection of this (Smith, Katz, and Niemeyer 2016). Citation typically links recognition (as measured by citations) with reputation; this becomes problematic in the often dogmatic association of this reputation in one way or another to career and funding structure (i.e., success), which leads to cycles of self-reinforcement (more grants, better publications, thus more grants; i.e., the Matthew Effect) (Merton 1968, 1988). Not only is this a misuse of the purpose of citations and a failure to understand what they are conceptually, it also misappropriates them in assuming that every citation is of equal value, as in a form of economic currency.

Data regarding these metrics and indicators are typically maintained by commercial bibliometric databases, such as the Web of Science and Scopus, that, in turn, also set the baseline for incentivization. The problem with using such a simplistic metric to measure impact and reputation is that it reinforces tendencies of risk aversion, lack of data transparency and many others, thus negatively affecting the Open Scholarship environment (Benedictus, Miedema, and Ferguson 2016; Haustein, Bowman, and Costas 2015; Wilsdon et al. 2017; Wilsdon 2017). This has led to the search for “Alternative Metrics” (popularly known as ‘altmetrics’) that can better assess the diversity inherent within the wider publishing ecosystem beyond citations and peer reviewed articles (Priem et al. 2010). Examples of these include blogs, media reports, white papers and other indicators of this discussion or use of a given piece of research. If new metrics are employed and a few of those metrics become prominent in determining the quality of research, it could potentially kick start a new set of biases within scholarly publishing. One way to circumvent this to an extent is to responsibly employ a combination of quantitative metrics and qualitative assessment (Hicks et al. 2015), so the system becomes less dependent upon metrics. The most important factor while choosing these metrics are the values which we want them to reflect (e.g., responsibility, trust, consistency), and how this is reflected in scholarly practice. To complicate matters further, Open Scholarship includes multiple roles, such as authors, publishers, reviewers, and editors, with a range of complex network interactions. To generate a robust framework for measuring impact and therefore reputation, different metrics and indices would be required to reflect this participant diversity (Nichols and Twidale 2017). Thus, any future appropriate valuation system has to capture the inherent diversity in dimensions between the practitioner, the practice, and the output.

As Cronin notes for the realm of Open Education, “complexity resides in determining and negotiating the value of open practice at an individual level, and structural and cultural barriers to openness persist within higher education” (Cronin 2018); but this surely also holds true for other dimensions of Open Scholarship. Thus, the question raised is how to measure reputation and impact within Open Scholarship. This critical factor has also received relatively little attention in the FOSS space (Hu, Zhao, and Cheng 2012), especially when compared to participation motivations. In FOSS, reputation is generally gathered based on one’s ability to be a good coder, and the practices that come with that in an openly participatory environment (Kelty 2001). In this way, it is similar to academia in that prestige functions as its own sort of capital, and drives an ‘economy’ based on reputation. However, a study in 2016 demonstrated that additional factors such as past reputation based on positive evaluations, social network effects, and other demographic traits were important in defining reputation within an Open Source community (Cai and Zhu 2016). Furthermore, such positive reputations tended to lead to greater success and performance as a group, thus re-emphasising the collaborative basis for reputation building. Applying this to Open Scholarship, it seems that there is therefore a substantial amount of scope in realising it as a community-based movement focused on gift-giving (Feynman 1974), especially around the tools that allow others to create, akin to FOSS, rather than a commercially-governed system focused primarily on outputs. This would help to realign the focus of where the value lies most in research, exposed around the general process of collaborative creation, and not just purely focused on outputs.

This renewed view of reputation accrual through collaboration and sharing can also have a secondary benefit, in that issues to do with appropriate authorship and accreditation[4] can be resolved by having explicit and tangible guidelines for contributing to research projects. This is because the concept of authorship shifts radically from being something bound to an article, to something in which attribution is fundamentally transparent and easy to resolve, as well as dynamic based on ones’ contributions, irrespective of what they might be. Thus, Open Scholarship workflow could seek to find new ways of quantifying or measuring contributions as different forms of ‘impact’ on research, including how the wider community interacts with processes and outputs. For example, these can be simple things borrowed from FOSS, such as project stars, forks, commits, and watchers (e.g., via GitHub). Each of these help to expose value in a different way, based on how content is perceived and digested by the wider community.

This would have several effects. First, it credit could be obtained and granted for more than just authorship of research papers. Second, through valuation of a greater diversity of processes and outputs, these things become more widely incentivised. Third, the alignment of wider recognition would bring the reward system in line with that of Open Scholarship, and thus ‘better’ research or best research practices. It ultimately would mean that research that does not follow convention can be rewarded, thus fostering diversity, freedom of exploration, and creativity (Frankenhuis and Nettle 2018b). Prestige and reputation thus would return to principles of “good” scholarship, and not be outsourced to inappropriate and proprietary measures. Elements of such a system can already be seen in a range of digital platforms that leverage the power of networked communities to gain a deeper understanding of people

[4] CRediT - CASRAI.

and processes (Tennant 2018). The bottom line is that the right measures of reputation across all levels of Open Scholarship could incentivize the right motivation. For this to happen, a broad exploration of altmetrics must ensue to ensure a substantial shift in how impact and reputations are measured.

However, this focus on incentives, and in particular metrics, as a mechanism for cultural change towards openness could also have dangerous and unintended consequences. While there is little doubt that rewards should be provided for more diverse forms of process and sharing beyond published articles, there is a risk that such evaluation procedures, and an increasing focus on ‘alternative metrics’, could lead to greater monitoring of researchers (and thus less freedom, innovation, and creativity), while still suffering from the same bureaucratic misuse as the impact factor does (Kansa 2014). One thing that OA, and Open Scholarship, has failed to do is to fully understand the significance and impact of such an entrenched paradigm that couples reputation, evaluation, and funding systems together, and how to alleviate or reform this. And, by extension, Goodhart's Law, which is commonly paraphrased as "When a measure becomes a target, it ceases to be a good measure."

If the focus remains on individual researchers adopting the principles and practices of Open Scholarship, the present evaluation system can potentially create resistance and thus become difficult for any single individual to overcome. This is distinct from FOSS, where individuals could more easily form communities to help to fix elements of the system and solve problems, without the need for substantial infrastructural or institutional support. Open Scholarship remains constrained by institutional forces that control academia (e.g., the reputation and evaluation process). This institutional hurdle is one of the challenges where FOSS and Open Scholarship significantly differ. Both software and research have, or had, the knowledge prohibition problem, but software development is more dynamic and independent of the institutional layer that the research community is rooted in. Recent research in North America is beginning to shed some crucial light into the extent of this (McKiernan et al. 2019), and provides the basis for strategic moves to evaluation reform. Specifically, structural reorganisation from the ‘top-down’ by those who are responsible for decision making and forming the hierarchical structure of research and academia. Without such a top-level systemic reform, change towards Open Scholarship will be slow.

All of these things tie Open Scholarship to much broader issues around credibility and legitimacy of published research, and thus to the wider impact of scholarship on society.

Reproducibility and data sharing

This final section will look at the intersections between data sharing, FOSS, and issues around research reproducibility. Perhaps one of the strongest pragmatic things that unites both FOSS and Open Scholarship is that of reproducibility, where reproducibility is defined here as the agreement between the results of measurements of the same measure and carried out with the same methodology. Reproducibility is inherent to the efficiency and

reliability of research, and is thus pragmatically underpinned by ‘open research’ practices (Munafò et al. 2017; Sandve et al. 2013a; ter Riet, Storosum, and Zwinderman 2019). Scholarship that is lacking in openness to facilitate reproducibility can be compared to only making compiled source code available, where reverse engineer to reproduce, at the mercy of the methods section), and open scholarship = equal to sharing the source code itself (reproduction immediate). Similarly, is scientific research really scientific if it cannot be reproduced, computationally or otherwise? In software environments, computational reproducibility is now becoming more frequently considered throughout all stages of the research development pipeline, when executed efficiently; for example, with strong community management, good documentation, and using container technologies (Boettiger 2015). Open Source software can help to facilitate transparency and reproducibility of experiments within a specific environment, as the entire process is exposed and ‘forkable’. However, while Open Source might lead to better software, it alone does not necessarily lead to reproducible research. Reproducibility is also affected by factors such as various operating systems, inputs, parameter ranges and data availability, and testing environments (Sandve et al. 2013; Millman and Perez 2014). Within the scholarly ecosystem, reproducibility is focused on how the software is being used rather than being an intrinsic property of the software itself. Thus, reproducibility manifests in several different forms within the software-scholarship ecosystem.

[**software alone is insufficient to reproducibility**] The advent of Open Science has radically affected the research life-cycle, in the way science is performed, its results published, assessed, discovered, and monitored. Research infrastructures (RIs) are the timely opportunity for reproducibility and openness of research results to blend with the mission of science itself, and “FAIRness of science” (not only of data) is acting as practical (and inspirational) recipe to achieve such objectives. In the last decade in Europe there has been a lot of hype around the construction of research infrastructures (RIs) and e-infrastructures, initiatives where research communities can identify a governance body to plan policies, best practices, and digital services to be adopted by scientists to leverage better scientific workflows. Within RIs, scientists can find and grow their *digital laboratories* (e.g. tools, services, standards, models, VREs) which they use to perform experiments that lead to the creation of new digital objects such as research data (e.g. time series, sensor data, maps) and research software (e.g. R scripts, executable workflows, scripts). Accordingly, digital laboratories are the digital twin of traditional scientific laboratories. Computer scientists make use of digital resources to perform their experiments, ranging from their laptops, to the Internet itself, cloud resources, and shared on-line services or software as a service solutions. Accordingly, to be as FAIR* as possible, when they perform an experiment and share the relative results they should also share the digital laboratory assets they employed, together with the experimental condition (e.g. configurations, software versions).

One critical factor that is inherent to scholarly research, and not so much in pure software, is that of data. More recently, the Open Data movement has taken off within Open Scholarship, particularly around the concept of ‘FAIR’ - that data is Findable, Accessible, Interoperable and Reusable (Wilkinson et al. 2016). With this in mind, the fundamental elements

underpinning research reproducibility are complete within an Open Scholarship system - articles, data, software, and all relevant research environment data are exposed and executable to help increase fundamental reproducibility. Open Data is now becoming an increasingly important and core aspect of Open Scholarship. This refers to the sharing of data, either raw or treated, that permits others to analyse and reach the same conclusions for a piece of research-thus allowing research reproducibility - or to provide resources (e.g., ideas, data, code) for new unintended research. However, despite the clear importance that software and computation plays in modern research, a similar demand or imperative for open sharing of software/code seems slower to reach some research fields yet, at least with the same strength as Open Data or Open Access.

One additional benefit of this process is that Open Source software can help to facilitate transparency and reproducibility of experiments within a specific environment and conditions, as the entire process is exposed and 'forkable'. Therefore, reproducibility sits at the interface between Open Source and Open Scholarship. Reproducibility is inherent to the efficiency and reliability of research, and is thus pragmatically underpinned by 'open research' practices (Munafò et al. 2017; Sandve et al. 2013a; ter Riet, Storosum, and Zwinderman 2019). Reproducible research then has grown at the intersection of the 'open movements' as a sort of hybrid ideology around making something available, but in the wider context of academic scholarship.

As data and code sharing are viewed as fundamental to reproducibility (along with other factors), this presents a strange tension in research, in that what is often best for research is pitted against that which is best for individuals or communities. In the field of Industrial Ecology, for example, one recent paper stated "[editors of journals]...should consider asking authors for a major revision of manuscripts whose claims are not reproducible because of lack of documentation, data, or software. This would encourage authors to document their scientific claims with higher transparency and reproducibility." (Pauliuk et al. 2015). However, this focuses only on one element of reproducibility (i.e., results), and there are other concerns such as analytical, computational, and methodological reproducibility that should probably be considered too. All of these are very complex concepts in themselves, both philosophically and in terms of practical implementation, and they add an additional dimension or barrier to the adoption of 'open research practices'.

A range of generalist or discipline-specific data (or 'object') sharing platforms or repositories now exists, including Figshare, GitHub, Zenodo, the Open Science Framework (OSF), 4.TUResearchData, and Dryad; although of these examples, only the latter three are built on Open Source software. Often, the intrinsic value of these services is focused around 4 things:

1. Opening data increases trust, verifiability, and reproducibility of research (Munafò et al. 2017);

2. Opening data tends to lead to more citations (Piwowar, Day, and Fridsma 2007; Piwowar and Vision 2013);

3. Opening data can lead to increased innovation and economic growth (Moedas 2015; Zuiderwijk et al. 2014).

4. Having open and FAIR data and subject to evaluation works as a good control mechanism.

Open sharing of data is perhaps not viewed with the same universal imperative as OA, as there are legitimate concerns regarding sensitive and personal/private data that cannot be made open or public by default – unless anonymised. Another example is research collaboration between academia and industry, where different values regarding openness and release policies can cause tensions. Furthermore, sharing data can be seen as a burden and a technical hurdle for some researchers, who may be unprepared for the challenge of creating, treating and sharing data for external consumption rather than private use. As a consequence of this, one of the major drivers behind sharing data has been the emergence of the FAIR Principles for scientific data management and stewardship (Wilkinson et al. 2016), which have now been adopted at a number of levels.

Similar to what has happened for OA, there is now a growing concern that open data too is being 'hijacked' by commercial organisations, who increasingly offer a range of services including repositories and management plans for researchers or research institutes as part of service packages (Posada and Chen 2018). This is part of a growing trend within the publishing sector, towards recognition of the commercial value behind data - which is produced by the research process, as well as metadata and usage data - and thus an increasing corporate-driven desire to control this element of scholarship. As such, this presents a new challenge that solutions and infrastructure around data sharing and storage need to become more widely supported by non-commercial entities, including possibly government-backed institutes. These issues around data sharing also factor into core principles around reproducible research, as well as the incentive/reward system that governs research. Reproducibility can often be seen as a barrier to recognition (often based around journal publications), for several reasons: reproducibility takes time, thus slowing publication down; reproducibility often is not a requirement for publication; failure to reproduce could be seen as taking value (e.g., originality) away from research, thus preventing its publication; reproducibility can challenge one's research, including assumptions and hypotheses, thus being undesirable; and there is less perceived value in reproducing someone else's work than in making one's own discoveries.

Computational reproducibility is not necessarily something simple or easy to learn. One thing which is generally agreed upon is that computers and software are foundational to many aspects of modern research. Yet, despite this, computational reproducibility does not appear to receive due attention or credit in the training of researchers, or in the process of daily research. Generally, computational reproducibility and reliability are not considered perhaps as much as they should be, given the impact that they can have on comprehensibility and reliability of research results. The same can perhaps also be said often for data processing and management, with both data and software often being relegated to 'second class citizens' with little documentation, organisation, testing, version control, or even whether software used are proprietary/closed source or open source. However, there are now many good examples of academic projects that try to improve this by promoting community

standards, code review, and reproducible software, such as Bioconductor and rOpenSci (Forero 2019; Boettiger et al. 2015).

A question that arises from this is whether Open Source software and Open Data are fundamental, and prerequisites for reproducibility and Open Scholarship. This opens up additional issues by exposing research processes, for example allowing auditing of the methodologies, allowing attempts at reproducibility as part of the peer review process, and opening up participation in research processes to those beyond those with access to proprietary data/software (thus tending towards 'citizen science'). It also motivates or encourages different forms of 'talent' within scholarship, something which can only be exposed through transparency and inspection of the entire research process. Such transparency, in theory, also facilitates more productive criticism, evaluation, and diversity in forms of peer review. As such, this element of Open Scholarship has many parallels to the latter development of the Open Source movement (David 2008; Vermeir and Margócsy 2012; Lahti, Silva, et al. 2017).

Limitations and future research

In this article, we have attempted to investigate the cross-over between Open Source and Open Scholarship, in the contexts of common understanding, communities of practice, underlying principles, issues around reputation and incentives, reproducible research, and the role of commercial players. Given the enormous complexity surrounding both software and research, this is clearly non-exhaustive. Likewise, we also did not consider additional overlaps with wider 'open culture', such as the dynamic histories behind Open Education and Open Educational Resources, Open Governments, Open Policy, Open Pharmaceuticals, among others. This explicit limitation in scope allowed us to focus on critical issues that remain pervasive within modern software and research cultures.

Such has the potential then for the underlying principles and practices, based around sharing and collaboration, to become normative within scholarly cultures. However, we recognise that both in Open Source and Open Scholarship, code, data and other outputs are not value-free. Algorithms, programmes and more, allow interventions in public life. The values inherent in openness need to be carefully balanced with other values (e.g., human rights, conservation, population health), meaning that one of the things Open Scholarship can learn from Open Source is that there is a need for a critical study of the ethics of openness.

Based around this, future research could also look into the personality traits, attitudes, and intrinsic and extrinsic motivators that seem to drive the Open Scholarship community, and look at the factors that seem to catalyse participation in the practices. A technology that could possibly capture most of the values we want in our research cycle and Open scholarship environment is Blockchain. Given its decentralized and transparent nature and appending of attributes such as immutability and provability to digital goods that circulate within it, it is a strong candidate for future academic systems. Decentralized and distributed structures are useful as they increase tolerance towards points of failure (meaning there are

multiple copies of a digital asset present) which could include things like storage of our academic output. Now imagine we have our academic repositories in the hands of a few private individuals/bodies (which we do in lots of cases). This exposes all this information to potential permanent tampering. This is where an immutable system is useful [Topol EJ 2016]. We want our information to evolve but not without a trace. Some form of version control and time stamping is necessary to ensure provability and integrity of our scientific data (Huprich 2017; Bell et al. 2017).

These systems however, are useful for more than just data transparency. Once data is acquired, Smart contracts can ensure pre-registration of our hypotheses regarding our studies and keep our analytical pipelines in check. For work with confidential/sensitive data attached to it, these networks and Smart contracts allow provability of work while simultaneously allowing us to include privacy for aspects of it [Kosba A et al 2016]. They can also work together with consensus mechanisms to democratize how pieces of research are validated, how scientific reputation in measured and through research tokens, also incentivize more credible knowledge creation. When it comes to evaluating pieces of research, the lack of proper metrics has resulted in the penning of thousands of signatures in the San Francisco Declaration on Research Assessment.(San Francisco Declaration on Research Assessment. url:http://www.ascb.org/dora/). While many alternate reputation systems were discussed in previous sections, one type of evaluation that could benefit from securing by decentralized systems is the use of Academic Endorsement Points (AEP) [b8d5ad9d974a44e7e2882f986467f4d3 2016. doi:http://doi.org/10.5281/zenodo.60054]. This is akin to a form of 'Research Currency' allocated to each researcher based on their work, which they can then use for endorsing other pieces of research. The best part being what you endorse, can range from anything from a whitepaper to a blog article. This helps diversify what we consider as an academic contribution.

Different contexts may call for different ways to allocate and disburse funding for research. While current systems make it hard to modify and implement new ways, flexible Blockchain frameworks can also help disburse funding in many varieties. This could involve coming up with tokens for individual tasks that are exchangeable with established tokens like Bitcoin. Moreover, these decentralized funding systems can also take advantage of an entire community to decide what counts as fair disbursement [Bollen J et al. 2014]. Finally, one of the more discussed applications of Blockchain in the academic context is in Peer review. A blockchain powered peer-review journal would involve tokenizing rewards to contributors, reviewers, editors and other service providers (Swan, 2015) to uphold a standard, below which there would be penalties, thus ensuring a push towards higher quality revisions and outputs (Tennant JP et al. 2017).

From Mikael Laakso: I don´t want to blow up the manuscript but I think Wikipedia could warrant some more emphasis in at least some other part of the manuscript, it is kind of a "mid-step" between the two arenas of consideration. Why does Wikipedia "work" while Open

Scholarship has been slow to get off the ground? Why is Wikipedia one of the most visited domains in the world, while open scholarship is still in its kindergarten phase of maturity? There is some interesting dissonance there that likely boils down to 1)incentive systems, 2) wikipedia contributions can be made whenever based on existing information, open scholarship requires the execution of a research project to adopt open scholarship practices from beginning to end (and during) to retain the capability to produce various open outputs as a result. There is probably more there to think about but in whatever way I would much enjoy seeing Wikipedia being featured is some more emphasised way, if no one else goes for it I can take a stab.

There are many aspects of this technology such as:

· mode of identifying researchers in the network

· whether the network should be secured by the Public or by private agencies

· kind of token systems to be used

that are still being debated by communities. While Blockchain has existed for a long time, its applications in Open Scholarship are being considered only recently and is definitely in Open Scholarship's future.

Several studies over the years have attempted to understand what motivates individuals to contribute to Open Source software. To our knowledge, the body of knowledge on the motivation for performing Open Scholarship activities is lacking. If we assume a similarity between the two phenomena, we can draw some parallels from the former to understand the latter.

Here, a further critical question to examine at what point during the scholarly process should it become 'open', to what extent, and to benefit whom. At the present, much of scholarship ultimately ends up being public at some stage, but only in the sense of not being secretive (Kelty 2001); however, from this it often ends up being privatised or obfuscated in ways that are not beneficial to wider society. There is also likely an enormous 'file drawer' problem with many aspects of historical scholarship, and little in the way of understanding the scope of this at the present, or how to potentially resolve it. This is related to the wider issue of who scholarship is supposed to serve. If it is for the wider benefit of society, so that the maximum number of people can exploit and use it, then it has to be as rigorous and reliable as possible. It clearly does also not meet this need if scholarly works are owned by private entities that actively prohibit its sharing and re-use. Thus, there is a direct tension here between Open Scholarship and organisations that maliciously violate fundamental freedoms underpinning this. However, without a solid principles-based foundation to Open Scholarship, and one that is widely accepted and enforced, then such violations will continue unchecked.

Conclusions

Here, similar to Free Software, the ‘enemy’ is not Open Source, but proprietary or non-free software. This is complicated further by the diversity and fragmentation of the scholarly community, and therefore providing a cohesive basis is difficult. Here, a key difference is that Free Software was defined based on freedoms before Open Source, and this catalysed the movement. However, Open Scholarship has already began and is now in need of such principles and values, to avoid unintended consequences around capture of this space by proprietary entities.

However, similar to the concepts underpinning Open Source, it might be that this focus on pragmatism and practical values has led to a failing to widely grasp or appreciate the underlying principles behind Open Scholarship; or at least to a sufficient degree to challenge the problems with the commercialisation of research and the evaluation system. It might also be this, coupled with the inherent hyperdiversity of research, that seems to have created a highly fragmented and heterogeneous ‘community’ around Open Scholarship.

If we consider Open Source to be more more about processes and methodologies, and Free Software more of a community-driven ideological movement (with some overlap), which of these best describes Open Scholarship? It seems that at its roots, Open Scholarship is a combination of both. Often it is described as a more efficient, effective way of doing research, injecting more transparency and rigour/reproducibility into the process. Other times, it is described as an issue of equity and social justice. However, at least part of this divergence or ambiguity is due to the fact that Open Scholarship seems to encompass a diverse myriad of practices, outputs, and principles, which have not been connected together into a single, unified and unambiguous understanding. Thus, it appears that there is also no clear cut boundary - either in terms of principles/values, practices, or outputs - between ‘open’ scholarship and ‘closed’ scholarship[5].

Openness must include accessibility beyond scholars, and free ability to re-use, as fundamental properties.

Potential add on: it should not have taken a cataclysmic event such as Sci-Hub to wake up the world to see that there is a real access issue. The moral, and practical, arguments behind fundamental openness for scholarship have been there for some time now, and funding an infrastructure that supports this should be imperative.

[5] What is open science, C. Titus Brown.

For the first few years of its existence, the Internet had the potential to form the technical and institutional basis for a new system of ‘openness’ in science. It has failed to achieve this for several reasons:

1. Evaluation metrics have not sufficiently changed from the pre-Internet world, and reputation and success are conflated;
2. Abuse of copyright (i.e., used counter to its initial intentions);
3. ‘Closed’ scientific practices, from the pre-Internet world, have been institutionalised.

From DG to integrate: what if universities did that, even on a national (governmental-backed infrastructure) level, perfectly legal, and using their established repositories (PID:s, metadata, data for assessment, research data, citations) with immediately and full access for everyone. Commercial profits from publishers can finance this shift, the money exists in the system. Also, an important aspect, this would increase speed for publications making them more interesting for further research and surrounding society, since current research cycle is slow compared to industry cycles with the present publishing system. Besides ownership will stay where it belongs, the researcher. Researchers can then spend more time on important research than looking for submissions and adopt to different formats for different publications and several review processes (the death of predatory journals will be a side effect, killing that argument against open scholarship).

Open Scholarship is a relatively new research term and paradigm, but with deep historical roots in the very foundations of modern scholarship. Indeed, openness in one form or another has been part of the scientific ideal for centuries now. However, in the present way it is being discussed and implemented, it could paradoxically lead to a more enclosed, monopolistic scholarly system dominated by ‘siloes’. It seems that open scholarship, if based more on morality around freedom, would align itself more closely with Mertonian norms (Bowman and Keene, 2018); as would thus open source and open scientific practices.

What seems clear is that the fundamentals that underpin open source and open societies (i.e., freedom) are either lacking, superficially addressed, or being utterly corrupted by commercial agendas in open science. We are witnessing this already. The advent of open platforms, open data, and open access have all been corrupted by major players such as Elsevier, Wiley, Sage, and Springer Nature (quote Posada and Chen). Open scholarship was intended to create a more rigorous research system that corrects the damages of contemporary publishers. However, it too risks succumbing to the same issues. This capture of open scholarship will not be due to any inherent value that these players add or because it is inherently good for scholarship. Their ongoing success in this space is simply due to the financial strength that they have and their capacity to invest (acquire) in key infrastructures and services in this space.

Such privatisation of the research process is clearly not in the spirit of science and scholarship as fundamental to creation of learning and development of a commons around knowledge.

In the future, it also worth exploring how open scholarship and open source intersect with other ‘open movements’ such as open education, open society, openness in public administration, and open cultures.

Note also CC licensing. Software developers now must be familiar with licensing as an essential part of their work. Research shows that researchers are generally unfamiliar with either copyright or licensing, clearly a catalyst for one of the three major issues mentioned above.

If policymakers wish to engage with open scholarship in any meaningful way, then they should ensure that any developments are based on fundamental democratic values like equity and freedom. At present, this is largely not the case. Instead, capitalist values seem to be interfering at the highest levels. Thus, it is in the best interests of the open scholarship community to organise politically and voice this case strongly at the political level.

Will the *glass house* of open scholarship be to *ivory towers* what the bazaar was to the cathedral in open source?

Author contributions

JPT conceived of the idea for the project, which was collaboratively written as a MOOP. All co-authors provided text, edits, ideas, and discussion for this paper, which was written in the open and collaboratively using Google Docs. It received hundreds of contributions as part of an iterative and organic procedure, and trying to define specific roles within this would be impossible and a pointless waste of time.

Acknowledgements

We thank Charles Dickens for articulating duality in words far beyond ours. Thanks to Arfon Smith for sharing some excellent presentations that helped to inspire this paper.

Others: Glenn Hampson, Brian Ballsun-Stanton, Tal Yarkoni, Bianca Kramer, Nadine Levin

To members of the ‘open community’ who every day provide a constant source of inspiration. We also acknowledge that, because of the open writing process, there are some people who might have contributed thoughts, discussion, or comments, who have not been credited in the final author list. We would like to extend our thanks to everyone who contributed to this in one way or another.

Conflicts of Interest

Like, billions probably.

JPT is the founder of the Open Science MOOC.

David Brassard produces a podcast about FLOSS in science as well as open science called FLOSS for Science.

Bibliography

Automatic citation updates are disabled. To see the bibliography, click Refresh in the Zotero tab.