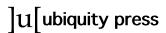




Cui Bono: Do Open Source Software Incubator Policies and Procedures Benefit the Projects or the Incubator?

RESEARCH ARTICLE





*Author affiliations can be found in the back matter of this article

ABSTRACT

Open source software (OSS), a form of Digital or Knowledge Commons, underlies much of the technology that we use in our daily lives. The existence and continuation of OSS relies on the contribution of private resources - personal time, volunteer energy, and effort of numerous actors (e.g., software developers' time as a common-pool resource) - to public goods, the benefits of which are enjoyed by everyone. Nonprofit organizations such as the Apache Software Foundation (ASF) attempt to aid this process by providing various collective services to OSS projects, acting as a second-order actor in the production of the public good. To this end, the ASF Incubator has created policies - essentially rules or norms - that serve to protect its interests and, as they say, increase the sustainability of the projects. Each policy requires investment by ASF (in terms of money or the use of volunteer time) or an incubating project (in terms of taking project personnel time), the benefits of which can accrue to either party. Such policies may impose additional costs on incubating projects, leading to a decreased production of the OSS public good. Using the ASF Incubator policy documents, we construct a dataset that records who -ASF or an incubating project - bears the cost and who enjoys the benefit of each policy and procedure. We can code most policy statements as costing one party and benefiting one party. The distribution of costs and benefits according to party indicates whether the second-order actor is contributing to an increase in the public good and if they are doing so sustainably. Through a two-way ANOVA, we characterize the impact of ASF policies on the production of public goods (OSS). Being a part of ASF imposes some costs on projects, but these costs may make projects more sustainable. Our analysis shows that the distribution of costs and benefits is fairly symmetric between the ASF and incubating projects. Thus, the configuration of policies or the "institutional design" of the ASF could aid in producing the OSS public good by providing services that projects require.

CORRESPONDING AUTHOR:

Anamika Sen

University of Massachusetts Amherst, USA anamikasen@umass.edu

KEYWORDS:

open source software; public goods; commons; commonpool resource; policy analysis; institutional analysis; costbenefit analysis

TO CITE THIS ARTICLE:

Sen, A., Atkisson, C., & Schweik, C. (2022). Cui Bono: Do Open Source Software Incubator Policies and Procedures Benefit the Projects or the Incubator? International Journal of the Commons, 16(1), pp. 64–77. DOI: https://doi.org/10.5334/ijc.1176

1 INTRODUCTION

Open source software (OSS) underlies much of the technology that we use on a regular basis – from bank transfers, to streaming movies, to online shopping. While hard to formalize, one recent European Union study estimated that the economic impact of OSS on the EU economy in 2018 was somewhere between 65 and 95 billion Euros (Leprince-Ringuet, 2021).

To produce OSS, groups of developers act collectively to create software that are licensed in such a way that they become public goods (Benkler, 2008; Schweik and English, 2012). Traditionally, in these forms of digital or knowledge commons (Hess and Ostrom, 2007; Schweik and English, 2013; Alonso de Magdaleno and Garcia-Garcia, 2015), one or more people create OSS by coming up with an idea for a software product, prototyping a solution and, through the use of copyright licensing and the internet, sharing their software solution with other people both for adoption and to encourage additional development (Free Software Foundation, 2021). Developers have multiple reasons for contributing to projects including: their own need for the OSS or the need of the software by their employer; the desire to learn and improve skills through reading the code and interacting with other OSS developers; publicly showcasing their programming ability to others; and, for some, a philosophical commitment to the idea of OSS (Lakhani and Wolf, 2003; David and Shapiro, 2008; Schweik and English, 2012). In effect, this time that developers are seeking to contribute to the production of OSS becomes a common-pool resource from which any OSS project could harvest developer time (non-excludable) that is not available to other projects after it has been used (rivalrous). This has been called the "volunteer energy" common-pool resource in the literature (Brudney and Meijs, 2009).

Software developers produce OSS by contributing their time to creating software that can be freely used by all. Once OSS is produced, it is hosted online and, depending on the particulars of its associated copyright license, it can be used by a new user no matter how many other people are currently using it (i.e., it is non-rivalrous). Furthermore, no one can be prevented from using it (thus, it is non-excludable). This makes OSS a classic public good (V. Ostrom and Ostrom, 1977). The OSS system, therefore, has two sets of "commonly-held" resources: the rivalrous common-pool resource of developer energy and the non-rivalrous public good of the software products themselves.

In the early days of OSS, small groups of like-minded individuals would come together to work on a project of mutual interest. These self-governing projects were left to their own devices for survival, trying to maintain development and grow a user base on their own, operating

under idiosyncratic norms and rules (Schweik and English, 2012). In the last 20 years, organizations have emerged as second-order actors in the production of OSS. They provide various collective services to aid OSS projects, such as legal, technical, and financial support (Riehle and Berschneider, 2012). A recent paper by Izquierdo and Cabot (2020) reports over 100 OSS nonprofit foundations exist today offering differing levels of support services. The Apache Software Foundation (ASF) - the focal nonprofit in this study - is one of the oldest such organizations with one of the largest numbers of associated OSS projects. A key service the ASF provides is the "incubation of podlings",1 which is a mechanism and social process within the foundation to nurture OSS projects that are interested in becoming a project formally associated with the ASF. To achieve their goal of integrating new projects into the ASF community, the ASF Incubator program has created policies or requirements that protect both ASF's interests and, according to Apache, make OSS projects more sustainable (Khudairi, 2019).

This represents an additional layer in the collective action problem of creating and maintaining OSS. Such nonprofits organize developer time with the aim of producing "sustainable" OSS projects. In the context of OSS, a project is sustainable if it continues to build a userbase of the product and a contributor-base to the project.² The nonprofit organizations provide guidance and establish requirements, policies or rules that incubating OSS projects wishing to fall under their umbrella must follow. From their perspective, they very well may have established the policies that they require podlings to comply with to streamline collective-action behavior within the OSS project based on other OSS developer's experience - the people running the ASF incubator program – in growing successful, sustainable projects. However, the people involved in the podling (the developers), may change the distribution of how they invested their time prior to entering the foundation compared to after due to these new policies, rules, and requirements. For instance, joining ASF may require some OSS contributors who had previously worked as developers on the project pre-incubation to switch part of their time toward efforts focused on interfacing with the foundation instead of developing software code for the specific project. OSS developers sometimes refer to this change in time allocation as a result of joining the ASF Incubator as the foundation "tax" (ASF, 2021c). This could also be seen as a transaction cost of incubation.

This relates the analysis of such foundations to the study of nonprofit organizations that gather individual donations (a common-pool resource; see Brudney and Meijs, 2009) to contribute to a public good such as environmental quality (Grant and Langpap, 2019). Whether or not nonprofit

organizations efficiently contribute to the production of public goods has long been of interest (Sugden, 1984; Roberts, 1987). Most existing research looks at individual motivations for contributing resources to what may be termed the "fundraising commons" (MacQuillin, 2015), though recent work has also examined if nonprofits are able to efficiently harvest from this common-pool resource to produce public goods (Grant and Langpap, 2019).

With all of this in mind, the key overarching question that we address in this paper is: How can second-order actors harvest from a common-pool resource (developers' and other contributors' time) to positively contribute to the creation of a public good? To this end, we analyze the policies or rules placed by ASF on entering OSS projects and use qualitative methods to study the cost-benefit distribution of these policies.3 On the one hand, the developers associated with the entering podling could completely absorb the costs – time and effort – of complying with the incubator rules, and ASF could receive all the benefits of abiding by those rules, as they would have a high-quality project associated with their nonprofit. This would represent ASF harvesting from the common-pool of OSS developer time in order to grow itself as an organization rather than the public good. Alternatively, the costs for implementing rules could fall largely on the role of the actors running the ASF Incubator, and the developers with the incubating podlings could be reaping all the benefits. In this case, ASF would be contributing to the public good independent of developers' contributions. Yet another possibility is that the costs and benefits are evenly distributed between the OSS podlings and ASF Incubator. Our findings provide support for a fairly symmetric cost-benefit distribution between the two parties.

The question of who incurs the costs and benefits of incubator rules is of significance as the model of OSS projects associating with an overarching nonprofit is widespread (Izquierdo and Cabot, 2020). Thus, understanding how operational requirements help or hinder each side can provide valuable insights into whether an incubation program works well or not. This paper adds to the existing literature that explores the factors that influence the costs and benefits of contributing to OSS, such as the motivations that guide participation in social movements (Hertel et al., 2003) and the adoption of new software (Islam et al., 2017).

The rest of this paper is organized as follows. Section 2 provides a brief background of the ASF Incubator and describes the conceptual framework of our analysis. Section 3 looks at the data and methods used in our analysis. Section 4 and Section 5 describe our findings and their implications. Section 6 draws some conclusions and provides thoughts on future research possibilities.

2 BACKGROUND AND CONCEPTUAL FRAMEWORK

2.1 BACKGROUND ON ASF AND THE ASF INCUBATOR

According to ASF's overview webpage (ASF, 2021d), the mission of the organization established in 1999 is to provide software for the public good, through the provision of services and support to software projects "... consisting of individuals who choose to participate in ASF activities." OSS projects "in the wild" that either request or ASF recruits to join the organization receive services and guidance from experienced OSS developers and ASF members on intellectual property (e.g., copyright licensing), financial contributions that help reduce legal exposure for OSS project developers, access to ASF technical resources, as well as project specific mentorship by experienced software developers who are ASF members with an eye toward longer-term project sustainability. OSS projects who ASF admits as incubating projects are required to abide by ASF's policies and incubation rules, thus giving up some of their operational freedom to comply with the so-called "Apache Way" policies (ASF, 2021b).

Established in 2002, the ASF Incubator acts as the entry point for OSS projects that wish to be a part of the ASF. The main roles of the Incubator are to help incoming projects adopt Apache's principles of governance and operation as well as guide them to successful graduation out of the Incubator to become full-fledged "top-level projects" that are formally associated with the ASF. According to the Apache Incubator website, as of November 2019, this program has helped 315 incoming OSS projects of which at least 200 of them have graduated (ASF, 2021a).

The ASF adds several layers of management to the collective action problem of creating and maintaining OSS compared to OSS projects "in the wild". Projects are hosted within the Incubator, which itself is a top-level project under the ASF umbrella. Consequently, the Incubator has its own Project Management Committee (PMC) who oversee its activities. ASF allows projects (both graduated OSS projects and the Incubator) to bring on new members to their PMCs through a vote of the PMC members on that project - and the Incubator is no different. Uniquely, however, the Incubator PMC approves all ASF member requests to be added to the Incubator PMC. But membership to the ASF is invite-only and via election, and typically reflects an enduring commitment to the principles of OSS development adhered to by Apache. Members ensure the continued success of ASF and are from where the ASF Board of Directors are drawn. This means that OSS development within the Incubator in ASF consists of multiple levels of collective action: contributions to individual projects, the Incubator, and the ASF at large.

2.2 CONCEPTUAL FRAMEWORK OF COST-BENEFIT ANALYSIS

Economic analyses of regulations often use quantitative assessments of costs and benefits to determine if they will improve the overall welfare of the society. While this type of analysis has been commonly used for environment, health, and safety (EHS) regulations (OMB, 2020), similar frameworks have also been proposed for financial regulations (Siegel et al., 2009; Posner and Weyl, 2013). In this paper, we adapt the cost-benefit framework to allow for a qualitative analysis of the policy statements that govern the incubation process of the ASF.

Nonprofit organizations such as the ASF aim to aid with the sustainability of the OSS commons by providing projects with a variety of services that could help them survive and even thrive. In exchange, however, the Incubator expects projects to abide by certain policies and procedures that the organization has in place. These policies require some investment of time or other resources – costs – by either the project or ASF, the benefits of which could accrue to either party. How the parties distribute costs and benefits could have important implications not only for the sustainability

of the OSS commons but also for the necessity and survival of the nonprofit organization itself. By using cost-benefit tables, we can summarize these distributions and gain a better understanding of how ASF organizes OSS production. Additionally, the cost-benefit distributions could provide us with insights on whether the incubation requirements of the ASF incentivizes continuous contributions to the OSS commons, thus resolving the collective action problem of maintaining the public good of OSS, or adds to the costs of producing OSS, thus making the collective action problem more acute. A balance between the costs and benefits accrued from following policies designed to protect natural resource commons is found to have important implications for encouraging participation in collective action (Cox et al., 2010). Given the similarities between the OSS commons and a natural resource commons in terms of incentives to contribute, we expect that organizations that aim to resolve this collective action problem will need to follow similar principles while designing their policies.

Figure 1 provides a set of hypothetical distributions of the costs and benefits of policy statements between the two parties: (1) the incubating project and (2) ASF officials who

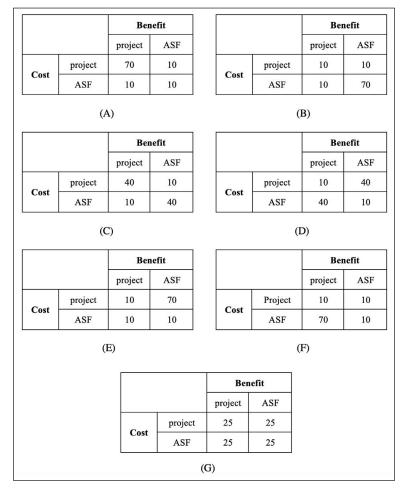


Figure 1 Alternative distributions of costs and benefits of policy statements.

are affiliated with the ASF Incubator (hereafter referred to simply as "ASF"). In Panel A, incubating projects incur a significant majority of the costs but also receive most of the benefits. Meanwhile in Panel B, we have the opposite scenario where ASF bears most of the costs but also enjoys most of the benefits of its policies. In both these instances it is unlikely that there are many contributions being made to sustain the overall OSS commons as each party is mainly engaging in activities that benefit itself. Furthermore, such a distribution of costs and benefits of policies may not be favorable for the existence of long-term relations between projects and the ASF as one party is far more invested in the process than the other. For instance, when ASF bears most of the costs and also enjoys most of the benefits (Panel B), projects may be more likely to leave the Incubator sooner as they neither are benefiting much from it, nor do they have much to lose in terms of resources invested. Conversely, when projects undertake most of the costs and also benefit significantly more from the policies, the ASF can easily shirk from its responsibilities without many losses.

Panel C illustrates a scenario that is fairly similar to Panels A and B. Here, the diagonal elements are more balanced indicating that the party that has incurred the cost of a policy is also likely to enjoy its benefits. However, in this case too, the nonprofit's policies may not help in sustaining the OSS commons as the actors are primarily involved in activities that directly benefit themselves. This could also reflect a situation where each party is mainly interested in protecting their own interests.

In Panel D, we observe the exact opposite of the above cases. In this case, it is the off-diagonal cells that contain most of the policy statements. This implies that when projects incur the costs of a policy, it is very likely that its benefits accrue to the ASF and vice versa. Such a distribution of costs and benefits establishes a pattern of mutual obligations between the two parties and possibly encourages risk sharing.

In Panels E and F, we observe two extreme cases of altruism. In the first instance, projects bear most of the costs of policies, the benefits of which accrue to the ASF. This distribution of costs and benefits could adversely impact the production of the OSS public good. In such a scenario, projects spend significant resources complying with policies that do not benefit them instead of engaging in activities that directly lead to the creation of more, higher quality, or more sustainable OSS. In Panel F, it is the ASF that incurs the majority of the costs of its policies for the benefit of the incubating projects. While this arrangement may be beneficial for ensuring the sustainability of the OSS commons, it may not be feasible for the ASF to persist in the long run under such an allocation of costs and benefits.

The final panel illustrates a scenario where the costs and benefits of policies are evenly divided between the ASF and incubating projects. Such an allocation could be an indication that the nonprofit organization has policies and procedures in place to provide projects with skills and resources that they need but do not possess. At the same time, ASF gives projects the freedom to engage in activities that they do best without requiring them to divert significant amounts of resources towards complying with policy requirements. From an overall perspective, such a distribution of costs and benefits could aid in the sustainability of the OSS commons.

3 METHODOLOGY

3.1 DATA

The dataset for this paper has been constructed using publicly available policy documents of the Apache Software Foundation Incubator (ASF, 2021a). There are a total of eight policy documents in our dataset. The sources of the original documents can be found in Table A1 in the Appendix. These documents are briefly described below:

- i. Incubation Policy: An overarching reference on the policies and procedures used in the incubation process.
- ii. The Apache Incubator Cookbook: A document that helps podlings – software projects in incubation – determine if joining the ASF Incubator is a good fit for them or not.
- iii. Podling Project Management Committee (PPMC) guide: Guidance on the role of the PPMC that supports each podling in incubation to learn how to govern itself and how they interact with the ASF.
- iv. Guide to Successful Community Building: Information to podlings emphasizing the importance of building a community to support the project and providing guidelines on how to bring in new developers ("committers") and PPMC members as well as how to encourage increased community engagement.
- v. Release Management guide: A release in OSS is making a new version of the software available to the public. This document outlines the rules that podlings need to comply with related to building ASF-compliant software releases.
- vi. Guide to Successful Graduation: Graduation from the Incubator is the ultimate goal for the podlings, meaning that they have proven themselves to the ASF as a project with a high probability of being sustained and that they understand the so-called "Apache Way" of building community, and approaching software development and release management.

- This document provides podlings with information on what is expected for graduation and the process of graduation.
- vii. Guide to Retirement: A retired podling is one which will no longer be associated with the incubation program for various reasons. This policy document describes retirement as a concept and as a process.
- viii. Mentors' guide: This document provides rules and guidelines for mentors on how to get a podling operational with special attention to the technical infrastructure (e.g., podling webpage, version control system, etc.) to support it.

While these are not all the documents hosted on ASF's website that offer instructions for how podlings can successfully go through the Incubator, they include all the formal policy documents as well as the documents that new podlings are directed to upon joining the Incubator. Other documents primarily reorganize this information to help podlings better understand the policies in place.

To construct our dataset, we first identified the institutional statements present in each of the above policy documents. We define an institutional statement in accordance with the literature as a "shared linguistic constraint or opportunity that prescribes, permits, or advises actions or outcomes for actors (both individual and corporate)" (Crawford and Ostrom, 1995, 583; see Table 3 for examples of institutional statements). We identify the operational-level institutional statements, which are policies that deal with day-to-day processes and functions (E. Ostrom, 2009). Such statements comprise the majority (approximately 90-95 percent) of identified institutional statements across the ASF Incubator policy documents.4 Our final dataset has a total of 259 operational-level statements. We summarize the distribution of these operational statements among the policy documents in Table 1.

3.2 METHODS

As the first step in our analysis, we classified the institutional statements in our dataset based on three attributes:

- i. Focus Area: Using the grounded theory approach proposed by Glaser and Strauss (2017), we categorize each institutional statement into one of the following focus areas risk mitigation, infrastructure, release management, project governance, Apache governance, roles and responsibilities, graduation criteria, communication, prerequisites, and Apache protection. These areas were identified by reading the policy documents, collecting recurring themes, and agreeing on them before coding individual statements. The definitions for these focus areas can be found in *Table 2*.
- **ii.** Benefit: Institutional statements are categorized as benefiting the ASF (OSS project) if the majority of the benefits from the policy accrue to ASF (OSS project).
- iii. Cost: Institutional statements are categorized as costing the ASF (OSS project) if the majority of the costs from the policy are borne by ASF (OSS project).

We base the assessment of "the majority of benefits" and "the majority of costs" on an in-depth understanding of the workings of the ASF Incubator. We arrived at this understanding through reading a large number of emails, reports that projects file regularly, reports that the Incubator files with the ASF, reviewing presentations given by experienced Incubator members to new incubating projects, and extensive one-on-one discussions with experienced members of the ASF.

It is important to note that this coding answers two questions independently: (i) Does it cost ASF or the podling more to comply with this policy? (ii) Does this policy benefit ASF more or the podling? – and not the single question: Does ASF benefit more than it costs them and is that benefit more or less than the podlings? Furthermore, we are concerned

POLICY DOCUMENT	NO. OF OPERATIONAL STATEMENTS
The Apache Incubator Cookbook	51
Guide to Successful Graduation	49
Podling Project Management Committee guide	42
Mentors' guide	34
Incubation Policy	28
Guide to Successful Community Building	26
Release Management guide	25
Guide to Retirement	4

Table 1 Distribution of operational statements across policy documents.

FOCUS AREA	DEFINITION
risk mitigation	Policies designed to make projects more sustainable
infrastructure	Policies describing technical rules of how to do things
release management	Policies describing how to make Apache-compliant code releases
project governance	Policies describing how things are done in a project
Apache governance	Policies describing how things are done in Apache
roles and responsibilities	Policies describing the duties of people occupying certain positions
graduation criteria	Requirements to graduate out of the Incubator
communication	Policies governing the exchange of information
prerequisites	Requirements to occupy a particular position
Apache protection	Policies designed to protect Apache from potential issues (legal, etc.)

Table 2 Definitions of focus greas.

INSTITUTIONAL STATEMENT	POLICY DOCUMENT	FOCUS AREA	COST	BENEFIT
"Each podling in incubation must report to the incubator PMC (Project Management Committee)."	Incubator Policy	risk mitigation	project	ASF
"A major criterion for graduation is to have developed an open and diverse meritocratic community."	Graduation Guide	graduation criteria	project	project

Table 3 Coding examples of operational statements.

only with the marginal cost of each operational-level policy statement, not the total cost of being able to get to that policy statement (e.g., we do not incorporate the cost to ASF for maintaining a Board of Directors into judgments regarding infrastructural requirements). *Table 3* provides a few examples of coded policy statements to illustrate this process in greater detail.

The first statement in *Table 3* describes the reporting requirements for incubating projects. We classify the focus area of this statement as risk mitigation as the objective of this policy is to ensure that any issues projects are facing are brought to the attention of Apache in a timely manner. The majority of the costs of this policy is borne by the incubating project (podling) as it has to file a report that meets ASF's requirements. The primary beneficiary of this policy is the ASF Incubator as these reports keep them apprised of any changes in the projects, thus allowing the ASF to encourage projects to retire or graduate in a timely way. While ASF also incurs costs to review a report (i.e., find mentors for a project, setup reporting infrastructure, read the report, etc.), we are concerned only with the marginal cost of this policy. Most of the costs borne by the ASF are contained in policies establishing them, such as the requirement for podlings to have mentors.

The second statement outlines an objective that projects need to meet in order to graduate out of the Incubator.

Graduation criteria is the focus area for this statement. Both the costs and benefits of this policy accrue to the incubating projects (podlings). Projects have to invest the time to build a community of contributors who are not all affiliated to a single company. However, they also enjoy the benefit of having such a community as they are no longer dependent on the time and interest of a single contributor or company for survival.

Two independent coders carried out this classification of operational statements. They subsequently resolved coding discrepancies in a separate discussion where the coders agreed on a common classification. After initial coding, Krippendorf's alpha (Hayes and Krippendorff, 2007; Hughes, 2021) was 0.66 (95% confidence interval = [0.63, 0.69]), which can be deemed "substantial" agreement. Following discussion, there remained only one disagreement out of 1,405 coded statements.

Once coded, we organized our data into 2×2 tables with each cell denoting the parties to which the costs and benefits accrue (see below, *Tables 4-9* and Tables A2–A14 in the Appendix). We then analyze whether there exists a relationship between which party incurs the cost and which receives the benefit. In other words, is the column (benefit) in which a case is independent from which row (cost) it is in.

Despite the high level of agreement between coders and a coding process that uses extensive knowledge of

Apache and projects to reach a final classification, we recognize that it is possible that other researchers could disagree on the assessment of which party takes on the costs or enjoys the benefits of any particular operational institutional statement. Consequently, we provide a measure of robustness for each finding (R). To calculate R, we generate a hypothetical table with the fewest number of changes in the cells that will produce the opposite result (e.g., make a non-significant result significant), and the number of changes is then divided by the total number of cases. The resulting R is the share of total cases that would need to change to give the opposite result. If this value is small, then it is possible that other researchers coding costs and benefits may generate a different finding. If this value is large, however, we have confidence that the finding is robust to mis-coding and legitimate disagreements.

Using the organized 2×2 tables, we can analyze our coded operational statements to determine whether there exists a dependency between which party incurs the costs and which party enjoys the benefits of these policies. In our analysis of the overall distribution of institutional statements, we use the chi-squared approximation of the binomial distribution to determine the p-value and reach a conclusion regarding statistical significance (Pearson, 1900). However, the chi-squared approximation diverges substantially from the exact distribution when the number of expected cases is less than five in one cell of the table, which often occurs in our data given the small number of cases in some tables. These instances call for an exact test. Researchers frequently use Fisher's exact test, though that test is conditional: the p-values are calculated assuming the row and column totals are known. When this is not the case, the test is overly conservative (Barnard, 1947). Boschloo (1970) developed an unconditional test, which has been found to have the best power of the unconditional exact tests (Berger, 1994). As such, all results presented here in which a cell of the table has fewer than five expected cases or which have fewer than 50 cases use the Boschloo exact test to derive the p-value. While the p-value allows us to identify tables with dependency, it breaks down in highly skewed tables. Specifically, tables in which the multiplication of the diagonal elements and the off-diagonal elements are both equal to 0 return a p-value of 1. This, however, can hide interesting cases such as, if all the benefits accrued to one party or one party bore all the costs. Consequently, in addition to the p-value we add our measure of robustness (R), which tells us the share of statements from a given table that would need to be coded differently to bring the p-value above or below 0.1, as the case may be.

4 RESULTS

4.1 COSTS AND BENEFITS ACROSS ALL ASF INCUBATION POLICY STATEMENTS

Table 4 summarizes the costs and benefits for all ASF incubation policy statements. The chi-squared test indicates no significant differences in the distribution of costs and benefits. This implies that at the aggregate level, the distribution of costs and benefits is almost symmetric between the projects and ASF. With a R of 0.054, 14 policy statements currently coded as costing the project and benefiting the project would need to be distributed among the other cells to bring the p-value below 0.1.

4.2 COSTS AND BENEFITS OF INDIVIDUAL POLICY DOCUMENTS

In addition to the aggregate analysis shown in *Table 4*, we undertook similar analyses of the distribution of costs and benefits for each policy document individually. Most documents show no evidence of a relationship between the costs and benefits accrued to ASF and incubating projects (see Tables A2–A6 in the Appendix). However, there a few notable exceptions. From *Table 5*, we observe that the distribution of costs and benefits of the policies in the Release Management guide is not independent of the parties. What sets this document apart from the others is that the majority of its policy statements fall under one of the off-diagonal classifications. This implies that when the cost of a policy statement falls on the incubating project,

		BENEFIT		TOTAL
		PROJECT	ASF	
Cost	project	101	60	161
	ASF	59	39	98
Total		160	99	259

Table 4 Summary of costs and benefits of Apache incubation policies.

Chi-squared test statistic = 0.075, p-value = 0.78, R = 0.054.

		BENEFIT		TOTAL
		PROJECT	ASF	
Cost	project	2	8	10
	ASF	11	4	15
Total		13	12	25

Table 5 Summary of costs and benefits of Release Management quide.

Boschloo test statistic = 0.015, p-value = 0.01, R = 0.12.

		BENEFIT		TOTAL
		PROJECT	ASF	
Cost	project	26	0	26
	ASF	0	0	0
Total		26	0	26

Table 6 Summary of costs and benefits of Guide to Successful Community Building.

Boschloo test statistic = 1, p-value = 1, R = 0.12.

		BENEFIT		TOTAL
		PROJECT	ASF	
Cost	project	2	0	2
	ASF	0	2	2
Total		2	2	4

Table 7 Summary of costs and benefits of Guide to Retirement. Boschloo test statistic = 0.33, p-value = 0.11, R = 1.

it is very likely that the benefits accrue to ASF. Conversely, when ASF bears the cost of a policy statement, its benefits likely go to incubating projects. The Release Management guide establishes a pattern of mutual obligations (as shown in *Figure 1*: Panel D) between ASF and projects, who need to work together to make successful releases. With an R of 0.12, 3 policy statements currently coded as costing ASF and benefiting the project would need to be distributed among the other cells to bring the p-value above 0.1.

Table 6 and 7 show distributions of costs and benefits that match panels in Figure 1. The policy statements in the Guide to Successful Community Building (Table 6) all require the costs to be borne by the project and provide the benefits to the project. This is an extreme example of Panel A from Figure 1. The policy statements in the Guide to Retirement (Table 7) are such that the party incurring the cost also receives the benefit. This is an example of Panel C from Figure 1. While this result is statistically insignificant, this is only because the small number of policy statements in this document barely prevents it from rising to a p-value of 0.1.

4.3 COSTS AND BENEFITS BY FOCUS AREA

When we analyze the distribution of costs and benefits by focus area – risk mitigation, infrastructure, release management, project governance, Apache governance, roles and responsibilities, graduation criteria, communication, prerequisites, and Apache protection – we find that in almost all cases there is no evidence supporting

		BENEFIT		TOTAL
		PROJECT	ASF	
Cost	project	0	10	10
	ASF	2	1	3
Total		2	11	13

Table 8 Summary of costs and benefits of Apache Protection policies.

Boschloo test statistic = 0.039, p-value = 0.02, R = 0.62.

a relation with status as ASF or incubating project (see Tables A7–A14 in the Appendix). There is one exception – policies classified under the Apache protection category. As shown in *Table 8*, the cost- benefit distribution of policy statements in this focus area is not independent of ASF or project status. This result arises as incubating projects do not receive any benefits from the ten policy statements that are found to impose a cost on them. The benefits of almost all policies in this category accrue to Apache. This pattern is very similar to what we observe in Panel E of *Figure 1*. With an R of 0.62, 8 policy statements currently coded as costing the project and benefiting ASF would need to be distributed among the other cells to bring the p-value above 0.1.

In addition to this, one focus area matches a panel given in *Figure 1*. Of the 10 policy statements that describe roles and responsibilities, all of them benefit the project while 60 percent of them impose a cost on ASF. This is an example of Panel F in *Figure 1*.

		BENEFIT		TOTAL
		PROJECT	ASF	
Cost	project	4	0	4
	ASF	6	0	6
Total		10	0	10

Table 9 Summary of costs and benefits of Roles and Responsibilities.

Boschloo test statistic = 1, p-value = 1, R = 0.60.

5 DISCUSSION

When looking across all policy documents and topical areas, the distribution of who incurs the cost of and who gets the benefit from ASF Incubator policies is fairly symmetric between ASF and incubating projects (*Table 4*). This is not to say that ASF and projects have the same number of policies in which they incur costs (98 and 161, respectively) or benefits (99 and 160, respectively), but rather that who

gets the benefit is independent of who incurs the cost. This distribution of costs and benefits matches Panel G from *Figure 1*. From the digital or knowledge commons perspective of OSS, this indicates that this policy regime is likely aiding in the sustainability of the commons by giving projects the freedom to engage in activities that they do best at the expense of minimal resources while ASF provides projects with skills and resources they need but do not possess. Such a distribution of costs and benefits allows a second-level actor to maximally contribute to the OSS commons while ensuring their own sustainability. This represents an efficient use of the common pool resource of developer effort to produce the OSS public good.

At a finer scale, however, individual policy documents have distributions of costs and benefits that are examples of different dynamics illustrated in Figure 1. The Release Management Guide (Table 5) is an example of Panel D, where costs incurred by one party go to benefit the other party. This establishes a pattern of mutual obligation between the parties, which could encourage risk sharing. The Guide to Successful Community Building (Table 6) is an extreme example of Panel A, where one party incurs all the costs and gets all the benefits. In isolation, such a distribution would demonstrate a loose affiliation between the involved parties. In this context, however, this highlights ASF's stated goals of helping projects grow communities, the members of which do not necessarily participate in the broader ASF community. The limited number of statements in the Guide to Retirement (Table 7) resemble the distribution in Panel C in *Figure 1*. Such a distribution tells us that each party is acting in their own interest and that the affiliation between parties is low. As retirement is the process whereby a project leaves the Incubator either through abandonment or via the community removing it from ASF but continuing development in the wild, this distribution reflects a situation where each party is protecting themselves and their interests.

There are also researcher-derived focus areas whose distribution of costs and benefits are examples of dynamics shown in *Figure 1*. The Apache Protection focus area (*Table 8*) captures all statements that are designed to protect Apache from potential issues (legal, etc.). The distribution of costs and benefits for these statements closely matches Panel E, where the costs are largely borne by projects and the benefits go to ASF. While in general projects benefit from being associated with ASF, at the operational level the onus to not abuse the ASF name and its trademarks falls on projects, ensuring that ASF benefits by being protected from potential issues such as lawsuits. The Roles and Responsibilities focus area (*Table 9*) captures statements that describe the roles of people occupying certain positions. While it establishes roles that must be

staffed at both the project and ASF level (showing in the near equal distribution of costs across the two parties), the aim of all those roles is to help projects (meaning that all benefits go to projects). This is an example of Panel F from *Figure 1*, which, in isolation, would indicate complete altruism on the part of ASF and might mean that the organization is not sustainable. As these are part of a larger body of policies, however, it does not indicate that in general.

6 CONCLUSION

The existence and continuance of the OSS commons relies on the conversion of a private good – developer time and effort - to the collective public good of OSS. Contributors have many reasons for doing so, ranging from a need they have to a desire to learn or signal skill. This pool of available developer time creates a common-pool resource of "volunteer energy" (Brudney and Meijs, 2009). There are now well over 100 non-profit organizations that aim to organize the efforts of individual software developers and "in the wild" OSS projects by bringing them under their umbrella through programs like the ASF Incubator (Izquierdo and Cabot, 2020). These programs provide OSS projects with services and education that the nonprofit sees as important for them to become sustainable (i.e., continued maintenance and development). However, the nonprofit foundations also require the OSS projects and the developers affiliated with them to comply with their incubation policies and procedures. These policies and procedures could be such that they draw from the common pool resource of developer time to build the nonprofit, leading to a decrease in the production of the OSS public good. Thus, when a project decides to incubate with ASF, developers divide their finite time between following new rules established by ASF while also engaging in their regular software development activities that directly contribute to OSS creation. This could reduce the rate of production of OSS, if a significant share of developer time shifts to rule compliance without adequate benefits.

The switch in time allocation may affect the efficiency of the OSS commons – how developers translate their time into sustainable open source software (Sickles and Zelenyuk, 2019). With developers dedicating more time to complying with the policies and procedures of the software foundations that they are a part of, less time is available to activities more directly related to OSS development and maintenance. In this manner, nonprofit foundations that support OSS projects could affect the efficiency of converting the common-pool resource of developer time and effort to the public good of OSS, in that developers

devote the same amount of time, but they produce less (or less productive, or lower quality) OSS. On the other hand, the additional cost of complying with these policies could also be an investment in the future sustainability of the project, meaning that time spent complying with policies will ultimately lead to more (or more productive, or higher quality) OSS.

In this paper, we develop a methodology to analyze the institutional design of OSS incubator programs. Specifically, we apply a cost-benefit analysis framework to investigate whether the operational-level institutional statements - the rules, policies and procedures - of the ASF Incubator explicitly benefit the OSS projects or the ASF Incubator program. Our results show that, at an aggregate level, the distribution of costs and benefits of the ASF's incubation policies is approximately symmetric between the projects and the Incubator. This result also holds true for most individual policy documents related to the incubation process as well as the identified focus areas of the operational statements. Deviations from balanced costs and benefits show interesting dynamics in different parts of the relationship between ASF and projects, both validating and showing the utility of this methodology.

For the ASF Incubator, this is a positive sign as these findings demonstrate that the rules they have dynamically created over time to guide their incubation program appear to be relatively equal in terms of costs and benefits accrued to ASF and to their incubating projects. This could indicate a sense of fairness and collaboration between the two parties towards the common goal of OSS sustainability. At an overall level, the policies and procedures of the ASF Incubator do not hinder the efficiency with which developers create OSS and thus represent an effective use of the common pool resource of developer energy.

While this approach of coding helps us conduct a qualitative analysis of the costs and benefits of the policies of second-order actors involved in the commons, it is subject to some limitations. In the absence of quantitative estimates, we are unable to make more concrete calculations of the costs that specific incubating projects incur in complying with these policies and the subsequent impact that this may have on their long-term success and sustainability. We can envision several ways of countering this such as, using the digital traces of OSS projects to look at how the distribution of effort changes from engaging with a second-order actor, and using surveys of developers to estimate how the amount of time invested in complying with policies changes over time. This method is also time consuming as coding is done manually by at least two independent researchers and requires consistency between them. One way to overcome this is to use text analysis methods that automate the coding of costs and benefits of institutional statements. Such methods could build on techniques that have been developed for institutional grammar analysis (Rice et al., 2021).

From a broader perspective, this study raises the issue of whether the relative balance of costs and benefits of operational rules holds true for other nonprofit OSS foundations. Is the balance between costs and benefits of operational rules a factor that influences whether OSS projects successfully graduate to completely affiliated projects with other nonprofits? Or are OSS nonprofits with incubator programs themselves more "successful" if and when their incubation program rules are more balanced from a cost and benefit perspective? These are questions that we intend to investigate in the next stage of our research. Additionally, we plan to quantify the costs and benefits identified in our analysis using surveys of incubating projects as well as data gathered from project reports and email communications.

The approach of aggregating costs and benefits of policies across parties can be used to study institutions within a common-pool resource management regime as well as compare across regimes. If there is a common-pool resource with many second-order actors organizing and restricting the action of lower-level actors, this method could allow us to characterize the relative contribution of each actor to the overall health of the commons. This would allow one to make recommendations to lowerlevel actors regarding which second-order institutions they may join. Additionally, we can develop a set of metrics that consider the relative distribution of costs and benefits (e.g., benefits to projects divided by costs incurred by the second-order actor) to characterize how secondorder actors contribute to the health of the commons. Such metrics could also be used to compare the effect of second-order actors on different common-pool resource management regimes.

Finally, from a methodological perspective, our analytic approach provides a relatively low-cost way to assess the distribution of costs and benefits for entire regulatory frameworks which can be applied to other commons settings as well. Determining the balance of costs and benefits for each individual policy (e.g., rules or norms) and then aggregating across them allows us to assess the burden of compliance with these institutional or regulatory requirements. This method may be especially useful in evaluating large regulatory frameworks that have emerged over time. The costs and benefits of new rules and regulations are typically measured on a piecemeal basis when the regulations are proposed (Hahn, 2004).

When a regulatory or institutional framework is built this way, imbalances in the distribution of costs and benefits that seem small for each policy may add up to unduly benefit one party at the expense of the other. By evaluating the entire institutional regime at once, this method allows us to evaluate the fairness of large-scale regulatory frameworks.

NOTES

- 1 In this paper, the term "podling" means OSS project in the Apache Software Foundation Incubator.
- 2 For a deeper discussion on this topic, see Chapter 7: Schweik and English (2012), or Schweik and English (2013).
- 3 There are at least two levels of benefits (and costs) in ASF. One is at the individual level, which looks at how developers benefit from being part of or involved with the OSS project and ASF. For example, developers can benefit from the experience of working on a project by learning from more experienced OSS contributors. At another level, there are rules guiding podlings in the incubation process and there are benefits and costs for podlings or the ASF Incubator at this "rule level". The OSS projects could benefit from having an ASF mentor guide them over the incubation process. ASF might benefit by having the OSS project associated with it and building the ASF brand. It is this latter level the study of who benefits or incurs costs to following rules that we study in this paper.
- 4 Other institutional statements could be related to constitutionallevel design of ASF or collective-choice rules specifying how ASF changes operational rules and practices. See E. Ostrom (2009) for more information.

ADDITIONAL FILE

The additional file for this article can be found as follows:

 Appendix. Tables A1 to A14. DOI: https://doi.org/10.5334/ ijc.1176.s1

ACKNOWLEDGEMENTS

We would like to thank Vladimir Filkov and Seth Frey along with the entire NSF GCR "Jumpstarting Successful Open-Source Software With Evidence-Based Rules and Structures" team for stimulating this investigation and comments on the manuscript. Brenda Bushouse, in particular, provided a key theoretical connection to the idea of "volunteer energy" discussed in the nonprofit literature. We are grateful to two anonymous reviewers for their comments and feedback. We also thank the participants of the IASC 2021 Knowledge Commons conference for helpful suggestions on an earlier draft of this paper. Santiago Virguez Ruiz assisted with data collection. This material is based upon work supported by the National Science Foundation under GCR Grant #GCR-2020900 and Grant #SES-1917908.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

University of Massachusetts Amherst, USA

Anamika Sen Dorcid.org/0000-0001-9066-0544
University of Massachusetts Amherst, USA
Curtis Atkisson Dorcid.org/0000-0003-3575-6871
University of Massachusetts Amherst, USA
Charlie Schweik Dorcid.org/0000-0002-0199-3740

REFERENCES

- Alonso de Magdaleno, M. I., & Garcia-Garcia, J. (2015).

 Sustainability and social responsibility reporting in open source software. International Journal of the Commons, 9(1), 369. DOI: https://doi.org/10.18352/ijc.503
- ASF. (2021a). The Apache Incubator. https://incubator.apache.org/
- **ASF.** (2021b). Briefing: The Apache Way. https://www.apache.org/ theapacheway/
- **ASF.** (2021c). Incubation Policy. http://people.apache.org/~geirm/incubator/incubation_policy.html
- **ASF.** (2021d). What is the ASF? https://www.apache.org/ foundation/
- **Barnard, G. A.** (1947). Significance Tests for 2 × 2 Tables. *Biometrika*, 34(1/2), 123. DOI: https://doi. org/10.2307/2332517
- Benkler, Y. (2008). The Wealth of Networks: How Social Production Transforms Markets and Freedom. Yale University Press. DOI: https://doi. org/10.12987/9780300127232
- **Berger, R. L.** (1994). Power Comparison of Exact Unconditional Tests for Comparing Two Binomial Proportions. 16.
- **Boschloo, R. D.** (1970). Raised conditional level of significance for the 2 × 2-table when testing the equality of two probabilities. *Statistica Neerlandica*, 24(1), 1–9. DOI: https://doi.org/10.1111/j.1467-9574.1970.tb00104.x
- Brudney, J. L., & Meijs, L. C. P. M. (2009). It Ain't Natural: Toward a New (Natural) Resource Conceptualization for Volunteer Management. *Nonprofit and Voluntary Sector Quarterly*, 38(4), 564–581. DOI: https://doi.org/10.1177/0899764009333828
- Cox, M., Arnold, G., & Tomás, S. V. (2010). A Review of Design Principles for Community-based Natural Resource Management. *Ecology and Society*, 15(4). JSTOR. DOI: https://doi.org/10.5751/ES-03704-150438
- Crawford, S. E. S., & Ostrom, E. (1995). A Grammar of Institutions.

 American Political Science Review, 89(3), 582–600. DOI:

 https://doi.org/10.2307/2082975

- David, P. A., & Shapiro, J. S. (2008). Community-based production of open-source software: What do we know about the developers who participate? *Information Economics and Policy*, 20(4), 364–398. DOI: https://doi.org/10.1016/j. infoecopol.2008.10.001
- **Free Software Foundation.** (2021). What is Copyleft? https://www. gnu.org/licenses/copyleft.en.html
- **Glaser, B. G.,** & **Strauss, A. L.** (2017). Discovery of Grounded Theory: Strategies for Qualitative Research. Routledge. DOI: https://doi.org/10.4324/9780203793206
- **Grant, L., & Langpap, C.** (2019). Private provision of public goods by environmental groups. *Proceedings of the National Academy of Sciences*, 116(12), 5334–5340. DOI: https://doi.org/10.1073/pnas.1805336115
- **Hahn, R. W.** (2004). The Economic Analysis of Regulation: A Response to the Critics. *The University of Chicago Law Review*, 71(3), 1021–1054. JSTOR.
- Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. Communication Methods and Measures, 1(1), 77–89. DOI: https://doi.org/10.1080/19312450709336664
- Hertel, G., Niedner, S., & Herrmann, S. (2003). Motivation of software developers in Open Source projects: An Internetbased survey of contributors to the Linux kernel. Research Policy, 32(7), 1159–1177. DOI: https://doi.org/10.1016/S0048-7333(03)00047-7
- Hess, C., & Ostrom, E. (2007). Understanding Knowledge as a Commons: From Theory to Practice. MIT Press. DOI: https://doi.org/10.7551/mitpress/6980.001.0001
- **Hughes, J.** (2021). krippendorffsalpha: An R Package for Measuring Agreement Using Krippendorff's Alpha Coefficient. *ArXiv*:2103.12170 [Stat]. http://arxiv.org/abs/2103.12170
- Islam, M., Miller, J., & Park, H. D. (2017). But what will it cost me? How do private costs of participation affect open source software projects? Research Policy, 46(6), 1062–1070. DOI: https://doi.org/10.1016/j.respol.2017.05.005
- **Izquierdo, J. L. C.,** & **Cabot, J.** (2020). A Survey of Software Foundations in Open Source. *ArXiv Preprint ArXiv:2005.10063*. http://arxiv.org/abs/2005.10063
- **Khudairi, S.** (2019, March 19). The Apache Way to Sustainable Open Source Success. The Apache Software Foundation Blog. https://blogs.apache.org/foundation/entry/the-apache-way-to-sustainable
- Lakhani, K., & Wolf, R. G. (2003). Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. SSRN Electronic Journal. DOI: https://doi. org/10.2139/ssrn.443040
- **Leprince-Ringuet, D.** (2021, February 8). How much are opensource developers really worth? Hundreds of billions of

- dollars, say economists. ZDNet. https://www.zdnet.com/ article/how-much-are-open-source-developers-really-worthhundreds-of-billions-of-dollars-say-economists/
- MacQuillin, I. (2015, September 4). OPINION: The fundraising commons not quite the tragedy we might think. http://blogs.plymouth.ac.uk/criticalfundraising/2015/09/04/opinion-the-fundraising-commons-not-quite-the-tragedy-we-might-think/
- OMB. (2020). 2018, 2019, and 2020 Report to Congress on the Benefits and Costs of Federal Regulations and Agency Compliance with the Unfunded Mandates Reform Act. https://www.whitehouse.gov/wp-content/ uploads/2021/01/2018_2019_2020-OMB-Cost-Benefit-Report.pdf
- Ostrom, E. (2009). Understanding Institutional Diversity. In Understanding Institutional Diversity. Princeton University Press. https://doi.org/10.1515/9781400831739. DOI: https://doi.org/10.2307/j.ctt7s7wm
- **Ostrom, V.,** & **Ostrom, E.** (1977). Public Goods and Public Choices. In E. S. Savas (Ed.), *Alternatives for Delivering Public Services*. Routledge.
- **Pearson, K.** (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50(302), 157–175. DOI: https://doi.org/10.1080/14786440009463897
- **Posner, E.,** & **Weyl, E. G.** (2013). Benefit-Cost Analysis for Financial Regulation. *American Economic Review*, 103(3), 393–397. DOI: https://doi.org/10.1257/aer.103.3.393
- Rice, D., Siddiki, S., Frey, S., Kwon, J. H., & Sawyer, A. (2021).

 Machine coding of policy texts with the Institutional

 Grammar. Public Administration, 99(2), 248–262. DOI: https://doi.org/10.1111/padm.12711
- Riehle, D., & Berschneider, S. (2012). A Model of Open Source Developer Foundations. In I. Hammouda, B. Lundell, T. Mikkonen, & W. Scacchi (Eds.), *Open Source Systems: Long-Term Sustainability*, 378, 15–28. Berlin, Heidelberg: Springer. DOI: https://doi.org/10.1007/978-3-642-33442-9_2
- **Roberts, R. D.** (1987). Financing Public Goods. *Journal of Political Economy*, 95(2), 420. DOI: https://doi.org/10.1086/261463
- Schweik, C. M., & English, R. C. (2012). Internet Success: A Study of Open-Source Software Commons. MIT Press. DOI: https://doi.org/10.7551/mitpress/9780262017251.001.0001
- Schweik, C. M., & English, R. (2013). Preliminary steps toward a general theory of internet-based collective-action in digital information commons: Findings from a study of open source software projects. *International Journal of the Commons*, 7(2), 234. DOI: https://doi.org/10.18352/ijc.397

Sickles, R. C., & Zelenyuk, V. (2019). Measurement of Productivity and Efficiency: Theory and Practice (1st ed.). Cambridge University Press. DOI: https://doi.org/10.1017/9781139565981

Siegel, P. H., O'Shaughnessy, J. J., & **Franz, D. P.** (2009). The Sarbanes-Oxley Act: A Cost Benefit Analysis Using the U.S.

Banking Industry. SSRN Electronic Journal. DOI: https://doi.org/10.2139/ssrn.1608962

Sugden, R. (1984). Reciprocity: The Supply of Public Goods Through Voluntary Contributions. *The* Economic Journal, 94(376), 772–787. DOI: https://doi. org/10.2307/2232294

TO CITE THIS ARTICLE:

Sen, A., Atkisson, C., & Schweik, C. (2022). Cui Bono: Do Open Source Software Incubator Policies and Procedures Benefit the Projects or the Incubator? *International Journal of the Commons*, 16(1), pp. 64–77. DOI: https://doi.org/10.5334/ijc.1176

Submitted: 31 December 2021 Accepted: 11 April 2022 Published: 29 April 2022

COPYRIGHT:

© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

International Journal of the Commons is a peer-reviewed open access journal published by Ubiquity Press.

