THE COPRODUCTION OF OPEN SOURCE SOFTWARE BY VOLUNTEERS AND BIG TECH FIRMS.

OPEN SOURCE IS EVERYWHERE, FACES AN EXISTENTIAL THREAT | THE 'OPEN SOURCE COMMUNITY' MYTH | FUTURES OF VOLUNTEER LABOUR | INVITED COMMENTS: PERSPECTIVES FROM FRANCE

O'NEIL CAI MUSELLI PAILLER ZACCHIROLI

DCPC21

NEWS AND MEDIA RESEARCH CENTRE

THE COPRODUCTION OF OPEN SOURCE SOFTWARE BY VOLUNTEERS AND BIG TECH FIRMS

Mathieu O'Neil
Xiaolan Cai
Laure Muselli
Fred Pailler
Stefano Zacchiroli

DIGITAL COMMONS POLICY COUNCIL
The DCPC is an international think tank established in 2021. It advocates for initiatives seeking to protect and expand the digital commons and use the digital commons to accelerate the transition to a more ecologically sustainable and fair society. It does so by publishing public reports on the digital commons and related topics, by co-producing how-to guides to assist in the creation of commonly held resources, and by making submissions and recommendations to government.
More information at www.dcpc.info

NEWS & MEDIA RESEARCH CENTRE
The N&MRC advances understanding of the changing media environment. Our research focuses on digital news consumption and the impacts of digital technology on journalism, politics, and society. Research occurs in three hubs: Critical Conversations Lab; Digital News+ Lab; and Media Cultures Lab. The Centre conducts both critical and applied research projects with partners and institutions in Australia and internationally.
More information at https://www.canberra.edu.au/research/faculty-research-centres/nmrc

June 2021

For more information about this report please contact: mathieu.oneil@canberra.edu.au
For more information about the Digital Commons Policy Council please contact: info@dcpc.info

ISBN (print): 978-1-74088-521-8
ISBN (electronic): 978-1-74088-520-1

DOI: https://doi.org/10.25916/r8vg-hd09

Cite as: O'Neil, M., Cai, X. Muselli, L., Pailler, F. & Zacchiroli, S. (2021). The coproduction of open source software by volunteers and big tech firms. Canberra: DCPC / News & Media Research Centre, University of Canberra.

Original cover design: ABM Studio | Report design: DCPC | Report layout: Zita Leung

This research project was funded by the Alfred P. Sloan Foundation and the Ford Foundation's Critical Digital Infrastructure Fund (2019-2020).

REPORT SUMMARY

This report maps how firms are collaborating with communities of unpaid volunteers to produce open source code, used in the 'digital infrastructure' which powers the contemporary networked economy. We map firm employee contributions to top GitHub repositories, finding that though firm employee contributions are dominant, volunteers still play an important role.

We analyse how the IT press portrays this coproduction: the issue of volunteer labour is absent. We show that large and small IT firm employee presentations at open source conferences convey opposed visions of digital infrastructure, business models, and the firm-community relationship.

The IT news media, big tech firms and commercial foundations define firms and projects as a unified 'community.' Yet big tech firms such as Amazon are using cloud computing and Software as a Service to transform open source software, which is intended to be shared and modified, into closed assets.

The report outlines strategic responses to big tech appropriation and reviews current debates about the recognition of volunteer work, money in FOSS, software licenses and universal basic incomes. The report also features invited comments exploring alternative perspectives by French open source specialists from the fields of academia, industry and activism.

DIGITAL COMMONS

Digital resources collectively produced and maintained by communities of diverse actors. They are governed by rules which guarantee their collective and shared nature. Examples include software (Linux), web browsers (Firefox), Wikipedia, and OpenStreetMap.

FREE AND OPEN SOURCE SOFTWARE (FOSS)

The first digital commons: computer software produced by volunteers in self-governed projects, which is meant to be freely shared and improved.

GENERAL PUBLIC LICENCE

The GPL or 'copyleft' is the most famous free license. It guarantees the freedom to copy, use, modify and distribute software.

OPEN CORE

A business model for the monetisation of commercially produced open-source software involving offering a 'core' or feature-limited version of a software product as free and open-source software, while offering commercial versions or add-ons as proprietary software.

BIG TECH (A.K.A. GAFAM)

After hardware in the 1980s, software in the 1990s and 2000s, data is now the key strategic asset, generating considerable revenue. The five GAFAM – Google, Amazon, Facebook, Apple, Microsoft – are the dominant actors in the data and 'cloud' or networked server market.

DIGITAL INFRASTRUCTURE

The name given to the software used to run 'cloud' data storage, networking and analysis services sold by big tech firms. It is based on open source software. Cloud computing has led to the introduction of new delivery models such as Software as a Service.

SOFTWARE AS A SERVICE

In a SaaS mode, a computer program is not transferred onto the user's computer, but executed remotely on the provider's cloud hardware, and used online. Service prevails over use: a subscription to a service is bought, rather than a user licensing agreement being accepted. This creates a SaaS 'loophole' in the FOSS licensing principle, as the service provider is no longer obliged to offer access to the code.

This report and the establishment of the Digital Commons Policy Council were made possible by the Alfred P. Sloan Foundation and the Ford Foundation's Critical Digital Infrastructure fund (2019-2020).

Many thanks to Josh Greenberg at the Sloan Foundation and Michael Brennan at the Ford Foundation for their support.

Some of the report's authors attended a workshop at the Ford Foundation in NYC in early 2019 with other fund grantees, where they received valuable feedback. In particular, Benjamin Mako Hill, Kaylea Champion and Stuart Geiger's helpful suggestions shaped the direction of our research. Our thanks to them also.

For more information on the Critical Digital Infrastructure program, see https://www.fordfoundation.org/campaigns/critical-digital-infrastructure-research/

Mathieu O'Neil is grateful to the University of Canberra's Faculty of Arts and Design for supporting his Outside Studies Program in 2019, which enabled him to focus on this project and to work with colleagues overseas. Special thanks to the Faculty's General Manager Eileen Wise for her support. At Télécom Paris, thanks to Laure Muselli, Benjamin Loveluck and David Bounie. At the Centre Internet et Société / CNRS, thanks to Mélanie Dulong de Rosnay, Francesca Musiani, Céline Vaslin and Olivier Alexandre.

Thanks to Gabriel Alcaras for his fieldwork during the Paris Open Source Summit.

Thanks to the invited commenters for their contributions.

Thanks to Fernanda, Fatima, Sama, Richie, and Laurent.

CONTENTS

AUTHORS

Xiaolan Cai is a PhD candidate at the University of Canberra. She researches community-led smart cities using computational and social network analytic methods.

Laure Muselli is Associate Professor of Information Systems Management at Telecom Paris, where she researches how new logics such as digital transformation or open source change work, occupations, identities and practices within organisations.

Mathieu O'Neil is Associate Professor of Communication at the University of Canberra where he leads the Critical Conversations Lab, and Honorary Associate Professor of Sociology at the ANU. He researches peer production, information literacy, and the trajectories of issues in the online environment. He was a founding member of the VOSON Lab in 2006. He founded the *Journal of Peer Production* in 2011 and the Digital Commons Policy Council in 2021.

Fred Pailler is a post-doctoral researcher at C^{2}DH, University of Luxembourg. He works on the uses of digital platforms and infrastructures in health, sexual, labour and/or political contexts.

Stefano Zacchiroli is Associate Professor of Computer Science at Université de Paris on leave at Inria. His research interests span formal methods, software preservation, and Free/Open Source Software engineering. He is co-founder and current CTO of the Software Heritage project. He is an official member of the Debian Project since 2001, where he was elected to serve as Debian Project Leader for three terms in a row over the period 2010-2013. He is a former Board Director of the Open Source Initiative (OSI).

CONTRIBUTORS

Danièle Bourcier is a legal scholar. She is emeritus Research Director at the Centre National de la Recherche Scientifique and at the Centre d'Études et de Recherches de Science Administrative et Politique (CERSA) at the Université Panthéon-Assas (Paris II).

Sébastien Broca is a sociologist. He is Associate Professor in the Media and Communication Department at Université Paris 8 Vincennes Saint-Denis. He researches digital capitalism and the digital commons. He is working in the EnCommuns and TAPAS research projects, and is co-editor of the *Anthropology & Materialism* scholarly journal.

Thierry Carrez is the Vice-President of Engineering at the Open Infrastructure Foundation. A Python Software Foundation fellow, he was previously the Technical lead for Ubuntu Server at Canonical, an operational manager for the Gentoo Linux Security Team, and an IT manager in various firms.

Pierre-Yves Gosset trained as an economist. After working for several universities, then at the CNRS, he became in 2008 Co-Director and General Delegate of the popular education association Framasoft. In particular, he is in charge of coordinating various projects such as https://degooglisons-internet.org, https://chatons.org and https://contributopia.org.

Célya Gruson-Daniel is an associate researcher at COSTECH (Technology University of Compiègne) and works as a consultant at Inno^3, where she is developing action-research projects at the crossroads between open source, open data, and open science. She is also one of the coordinators of the Open Science and Society working group at the Internet & Society Centre in Paris.

Benjamin Jean is the founder of Inno^3. He is a lawyer specialising in intellectual property management within the framework of open models. He is also active in free software communities as an administrator of Framasoft and as co-founder of Veni Vidi Libri and of the European Open source & Free Software Law Event (EOLE) conference series.

Hervé Le Crosnier is a publisher at C&F éditions. He previously worked as a librarian, then as a lecturer and researcher in computing and digital culture at the University of Caen.

Camille Moulin is an open source and open standards advocate. He is currently working as a consultant at Inno^3, a French consultancy firm specialising in intellectual property and organisational open infrastructure.

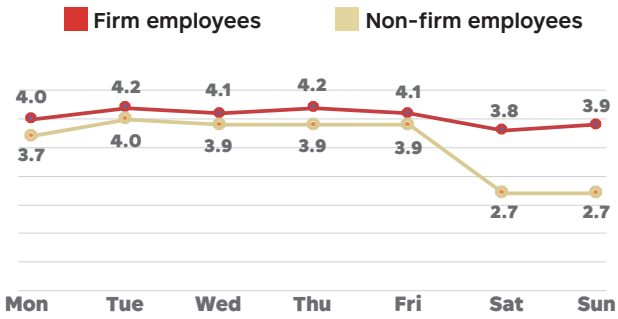
KEY FINDINGS

OPEN SOURCE IS EVERYWHERE

FOSS IS USED IN 92% OF ALL THE APPS!

Open source software is used in all digital devices and infrastructures: a survey of 1,200 Information Technology professionals found that 92% of applications contain FOSS libraries.[1]

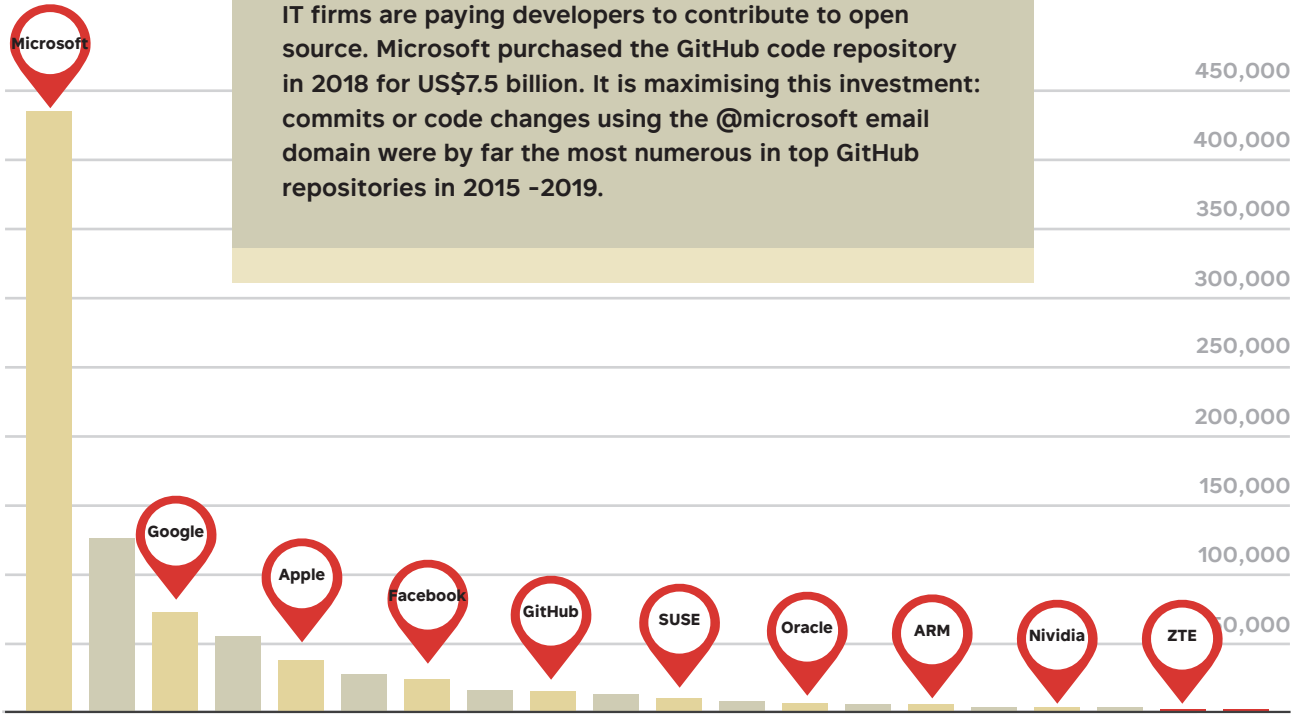
ROLE OF VOLUNTEERS



FOSS projects still rely on volunteer labour. On GitHub, the average number of commits produced by firm employees and non-firm employees per day between 2015 and 2019 were almost identical. Firm employee contributions dip during the weekend, whereas volunteer contributions remain constant.

IT FIRMS MASSIVELY CONTRIBUTE

IT firms are paying developers to contribute to open source. Microsoft purchased the GitHub code repository in 2018 for US$7.5 billion. It is maximising this investment: commits or code changes using the @microsoft email domain were by far the most numerous in top GitHub repositories in 2015 -2019.



[1] Szulik, K. (2018, April 12) Open source is everywhere. https://blog.tidelift.com/open-source-is-everywhere-survey-results-part-1

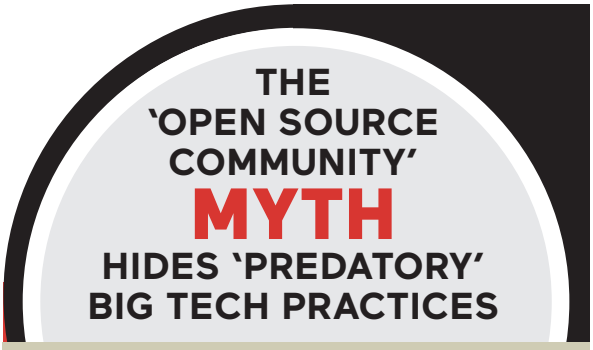
A DIVISION IN THE INFORMATION TECHNOLOGY FIRM SECTOR

Large IT firms ('big tech') promote the collection of user data, cloud computing and Software as a Service. Their concerns are that projects adopt consistent performance, safety, documentation and technical standards, and diversify their contributor base. They do not address the cost of labour. In contrast small IT firms produce open source directly following the 'open core' model. They have financial and ethical concerns such as the sustainability of open source business models, costs of labour, 'free riding' (using open source without contributing) and the issue of control over open source.

HOW THE IT NEWS MEDIA PORTRAYS OPEN SOURCE

The central issues in articles about firm-project coproduction in ZDNet, PCWorld and Slashdot were 'data' and 'cloud.' Firm-project coproduction was described in terms of professionalisation and 'career development.' The role of volunteers in FOSS and the value of their free labour are not mentioned.

THE COMMUNITY MYTH



IT media, big tech firms and 501(c)(6) foundations define firms and projects as a unified 'community.' Yet cloud-based systems such as Software as a Service (SaaS) introduced by firms like Amazon transform FOSS, a common resource meant to be shared freely, into a proprietary resource to be sold: the opposite of 'community'?

STRATEGIES AND SOLUTIONS

The identity of 'free riding' firms and the existential threat posed by big tech to FOSS should be publicised and debated.

The contributions of researchers to FOSS sustainability should be publicised and debated.

Public authorities and regulatory bodies should support alternatives to big tech services.

The rise of automation and predicted job losses require creating more connections and recognition between the volunteer sector and state institutions.

1

THE INTEGRATION OF OPEN SOURCE SOFTWARE INTO THE FIRM ECOSYTEM

- Free and open source software, a shared resource developed by volunteers in projects and communities, has been adopted by IT firms and is used in every sector of the digital economy, sometimes under the name ‘digital infrastructure.’
- Two entities facilitated the adoption of free and open source software by firms: the GitHub collaborative development platform allows developers to showcase their coding portfolios; the Linux Foundation celebrates the open source community and communicates firm requirements at trade conferences.
- Big Tech firms such as Amazon are using cloud computing and ‘Software as a Service’ built with open source software to transform scientific and technological knowledge intended to be shared, into closed assets to be sold.
- Free licenses such as the GPL failed to prevent this appropriation by big tech.
- Alternative perspectives and solutions exist, including some from France.

WHY DOES 'DIGITAL INFRASTRUCTURE' MATTER?

Today everything from hospitals to stock markets to newspapers runs on software using free or open source code. This is the infrastructure of the digital economy, described by Nadia Eghbal in the *Roads and Bridges* report.[2] This software is produced by volunteers in self-governed projects, and anyone is free to access and modify it thanks to 'copyleft' licenses such as the General Public Licence (GPL). Free and open source software is ubiquitous online. The foundational LAMP open source web application acronym (Linux, Apache, MySQL, Perl/PHP/Python) is a good example: Google owes its dominance to Linux (used in Android and Chrome OS). Apache powers 40% of the Internet's web servers. Without the MySQL database, there would be no online commerce (Paypal, Amazon), social media (Facebook, Twitter, LinkedIn), or 'sharing economy' (Uber, Yelp). Perl/PHP/Python are popular programming languages.

[2] Eghbal, N. (2016) *Roads and bridges: The unseen labor behind our digital infrastructure.* NYC: Ford Foundation

THE ADOPTION OF OPEN SOURCE SOFTWARE BY FIRMS

How did open source come to occupy this position, given it was originally perceived as a major threat by firms because licenses such as the GPL contradict traditional understandings of intellectual property? The reasons are many. Outsourcing labour to volunteer projects lowers production costs, and firms can also take advantage of the excitement generated by FOSS projects, particularly when it comes to recruitment, as IT firms employ developers to produce open source code. In addition to IT firms, significant end-user firms such as Sony have created Open Source Program Office (OSPOs) which act as internal liaisons between stakeholders in firms and external communities of volunteers. Firm adoption of open source can also cause risks, such as personnel issues when key maintainers drop out or move on from projects.

A SHORT HISTORY

- 1985 Free Software Foundation
- 1989 General Public Licence
- 1991 Linux
- 1993 Debian
- 1998 Microsoft Halloween documents declaring Linux a 'major threat' leaked
- 1998 At the Open Source Summit in Palo Alto, key players in the community decide to use the business-friendly term 'open source' in preference to the term 'free software'
- 2000 Linux Foundation
- 2005 GitHub
- 2008 First mobile devices running Linux-based Android
- 2010 Oracle buys Sun Microsystems
- 2011 Firefox and Chrome open source web browsers more popular than Internet Explorer
- 2018 Microsoft buys GitHub for US$7.5b, IBM acquires Red Hat for US$34m

TECHNOLOGY ENABLERS: GITHUB, STACK OVERFLOW

The adoption by IT firms of open source licenses, a significant change in corporate practice, enabled the integration of FOSS into the IT industry. Online platforms such as GitHub and Stack Overflow also played a part. Created in 2005, GitHub is a code hosting platform based on the Git version control system which facilitates large-scale collaborative development and makes individual contributions to FOSS projects highly visible. It has grown to become the most popular FOSS collaborative development platform, now hosting 40 million users and 190 million repositories. Its position as a central hub made it difficult for developers to consider leaving it when it was purchased by Microsoft. The Stack Overflow platform, launched in 2008, has become a key resource for mentoring and advice.

CORPORATE ENABLER: THE LINUX FOUNDATION

Open source's adoption by firms was also facilitated by the emergence of mediating entities such as the Linux Foundation. This foundation was originally created in 2000 to employ Linux's creator Linus Torvalds in order to prevent him from being attached to a single firm which would thereby gain inordinate influence. It has grown into a key facilitator of software interoperability and firm-project coproduction. Firms who rely on open source technology require both employees with open source skills and solutions that are compliant with their security and safety standards. The Linux Foundation met these requirements in two ways: it sought to change the way developers work, by advocating for stricter norms and rules, thus enabling better quality releases; and it provided professional certifications.

THE CLOUD MENACE: SOFTWARE AS A SERVICE

Data analysis generates considerable revenues, provided these data can be massively aggregated in terms of volume and diversity, and analysed with velocity ('3 Vs of big data'). The data market is dominated by web giants such as Facebook, Google and Amazon who use their own platforms, or by historically dominant IT firms such as Microsoft which have acquired platforms such as Skype, LinkedIn or GitHub. Data management and services occur in the 'cloud,' on servers whose digital infrastructure is based on free and open source software.

FOSS is shareable following the 'copyleft' principles of the GPL. Cloud computing includes a new value proposition, Software as a Service (SaaS), which allows the outsourcing of software applications. In a 'traditional' mode, a software program is downloaded and executed by customers on their own hardware. In a SaaS mode, the program is not transferred, but executed remotely on the provider's hardware and used online (e.g., within a Web browser). With SaaS, service prevails over use: a subscription to a service is bought, rather than a user licensing agreement being accepted for software downloaded onto the user's computer. This creates a SaaS 'loophole' in the FOSS principle, as the service provider is no longer obliged to offer access to the code: copylefted software used as SaaS is not 'distributed' - since only a service is being provided - and therefore fails to trigger the reciprocal character of the GPL.

OTHER EXAMPLES OF BIG TECH APPROPRIATION OF SHARED RESOURCES

This subversion of FOSS principles is part of a range of 'predatory' practices which also includes establishing scientific collaborations with research institutions whilst seldom sharing intellectual property, releasing libraries to open source in order to set standards, enabling start-ups to integrate their applications into GAFAM platforms and - in case they are acquired - facilitating their integration into the GAFAM. A striking example of 'predatory' GAFAM behaviour is the re-appropriation of co-authored research: 78,3% of Microsoft's 17,405 publications between 2014 and 2019 were co-authored with university researchers; during the same period Microsoft applied and was granted 76,109 patents, 0,2% of which were co-owned.[3]

[3] See Rikap, C., & Lundvall, B-A. (2020). Big tech, knowledge predation and the implications for development. *Innovation and Development*.

GPL: WHAT HAPPENED?

Why couldn't copyleft licenses prevent GAFAM appropriation? Copyleft licenses such as the General Public Licence (GPL) need to be associated with dominant software to become 'attractors' of contributions. This was the case for Linux, but occurred less for other free software. Most importantly, copyleft has been subjected to relentless legal attacks from GAFAM, and especially from Google. Linux is the central operating system of the open source world. Google has largely built its current dominance on Linux, which is the foundation of Android phones, so the firm is forced to share the source code of its Linux modifications, which is distributed under GPL license. It has therefore worked to develop an alternative core to Linux, developed *ex nihilo* in order to control it and to associate a non-copyleft license to it: Google Fuchsia.[4] The most effective copyleft licenses against GAFAM 'cloudification' such as the GNU Affero General Public License (AGPL), if adopted massively, would force Google and its ilk to share the source code of software running on their servers, even for users who interact with this software remotely; Google banned its employees from using the AGPL.[5]

[4] See https://en.wikipedia.org/wiki/Google_Fuchsia

[5] See https://opensource.google/docs/using/agpl-policy/

ALTERNATIVE SOLUTIONS AND PERSPECTIVES: VOICES FROM FRANCE

Clearly, new perspectives are needed. This report gathers diverse invited comments by French open source specialists from the fields of academia (D. Bourcier, S. Broca, H. Le Crosnier), industry (T. Carrez, C. Gruson-Daniel, B. Jean, C. Moulin) and activism (Framasoft's P.-Y. Gosset). Framasoft originally sought to build bridges between teaching and free software values such as sharing knowledge, transparency, and mutual aid. It has since become the main provider of concrete alternatives to big tech services, with campaigns such as Degooglisons Internet and CHATONS having significant impact. Two other notable French digital rights organisations are La Quadrature du Net[6] and the Association pour la Promotion et la Recherche en Informatique Libre (April).[7]

[6] See https://www.laquadrature.net/en/

[7] See https://www.april.org/

INVITED COMMENT:
SEBASTIEN BROCA

TIME FOR PUBLIC AUTHORITIES TO STOP BLINDLY SUPPORTING BIG TECH

The report rightly tackles the issues stemming from the fact that large IT corporations greatly benefit from free and open source software. However, it is worth noting that the issues at stake, aptly described by the authors, are multi-faceted. One of them is the lack of remuneration for voluntary contributions, even if firm involvement in free and open source software (FOSS) has risen in the past few years.

Another issue – quite a different one – is the predominance of large IT firms in the governance of FOSS projects, i.e. the shift from community-driven projects to projects closely controlled by a firm or by an industrial consortia. In the first case, the problem seems to be that firms do not participate enough (they act as free riders); in the second, the problem seems to be that they participate too much (they act as if they 'own' the project). This ambivalent situation leaves FOSS advocates in a somewhat ambiguous position. Should they ask large IT firms to contribute more or to contribute less? Or, to put it differently: what is the right contribution from large IT firms to the FOSS community?

There is no easy or simple answer to this question. The least these firms can do is to be more transparent about their use of FOSS. They should also contribute code somewhat proportionally to the benefits they obtain from free software. One could also expect that they would adopt and favour open models of governance, considering software as a commons managed and developed by different stakeholders inside and outside the firm.

These proposals may however appear as wishful thinking, given the fact that the main problem is structural. It is fundamentally the dominant position of big tech in the FOSS ecosystem (and in society at large) that must be questioned. This is when and where the state enters the picture. Public authorities and regulatory bodies currently do not act in support of the numerous existing alternatives to Big Tech: small tech, coops, or decentralised services based on free software.

Worse still, public authorities have not even levelled the playing field between big tech and their competitors, a reality in stark contradiction with the liberal mantra of a competitive market economy. Public authorities, be it in Europe or in the US, have in fact actively reinforced the predominance of big tech, refusing to block mergers and only mildly addressing monopolistic concerns; letting big tech avoid paying their fair amount of taxes; insufficiently regulating their use of personal data; and directing public funding (in France, through the Banque Publique d'Investissement for instance) chiefly toward tech start-ups, ignoring alternative forms of entrepreneurship and commons projects.

Public bodies have adopted and reinforced the narrative on technological and social innovation advanced by Silicon Valley and its organic intellectuals. In a time where deep economic changes aiming for a more just and sustainable society are needed, this vision must change. Cooperative platforms, commons projects and smaller tech firms (in France, actors like Framasoft, IndieHosters, Mobicoop, Coopcycle, Oiseaux de passage, to name just a few) are in urgent need of public support. Although they occupy niche positions in today's economy and are still largely ignored by the general public, they can serve as a prefiguration of what a more just and sustainable digital future might look like. That is why we should pay attention to what they are building, and ask public authorities to do the same.

2

CONTRIBUTIONS OF FIRM EMPLOYEES TO TOP GITHUB REPOSITORIES

- GitHub is the hub of the firm-volunteer coproduction network.
- Developers contribute ‘commits’ or source code changes to repositories.
- We analyse 135 top GitHub repositories.
- A minority of projects concentrate a majority of firm investments and a minority of firms massively contribute.
- Linux is the central project and Microsoft is the main corporate contributor.
- Volunteers play an important role, and their contribution level remains constant, in contrast to firm employees.

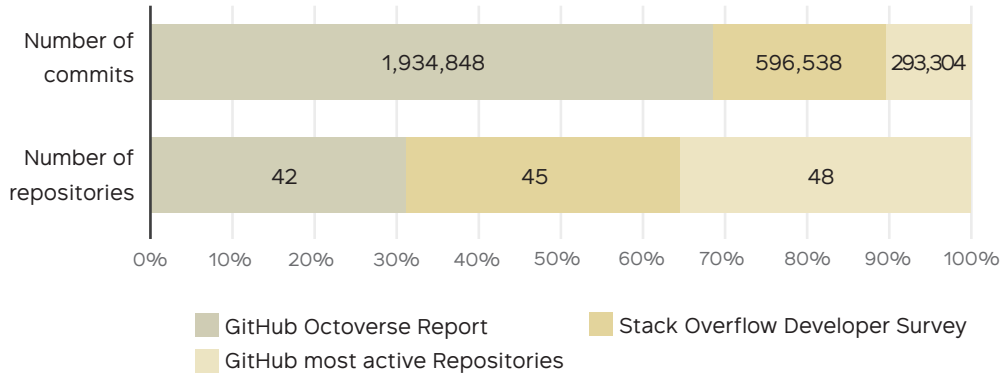
IDENTIFYING ACTIVE GITHUB REPOSITORIES

Between early 2015 and mid-2019 approximately 10 billion commits were published on GitHub. These numbers are deceptive. Despite being nominally oriented towards 'social coding', and while many repositories are being actively developed on GitHub, most of them are simply personal, inactive repositories.[8] For this reason, we targeted a restricted number of highly active repositories. 135 GitHub repositories were selected in three ways, summarized in Table 2.1 and Figure 2.1. We collected commit information from these 135 repositories: 113,614 committers with 26,459 unique email domains contributed 2,824,690 commits to these 135 repositories between 01.01.2015 and 31.05.2019.

Table 2.1. Constitution of GitHub repository dataset[9]

	GitHub Octoverse Report	Stack Overflow Developer Survey	Most active GitHub Repositories
Number of repositories	42	45	48
Source	Open Source section	Technology section	Self-identified open-source repositories
Metric	Count of committers, commits, and forks	Vote by developers	Count of commits and stars
Time range	November 2015 to October 2018	January 2016 to December 2018	Repository creation date to 30 April 2019
Number of commits	1,934,848	596,538	293,304

Figure 2.1. Constitution of GitHub repository dataset



[8] S Kalliamvakou et al. (2014) found the number of committers per GitHub repository is highly skewed: 72% of repositories have one committer, 91% have 2 or less, and 95% have 3 or less.

[9] Source: GitHub Octoverse Report https://octoverse.github.com/, Stack Overflow Developer Survey: https://insights.stackoverflow.com/survey/. The most active GitHub repositories were collected by researchers during May 2019.

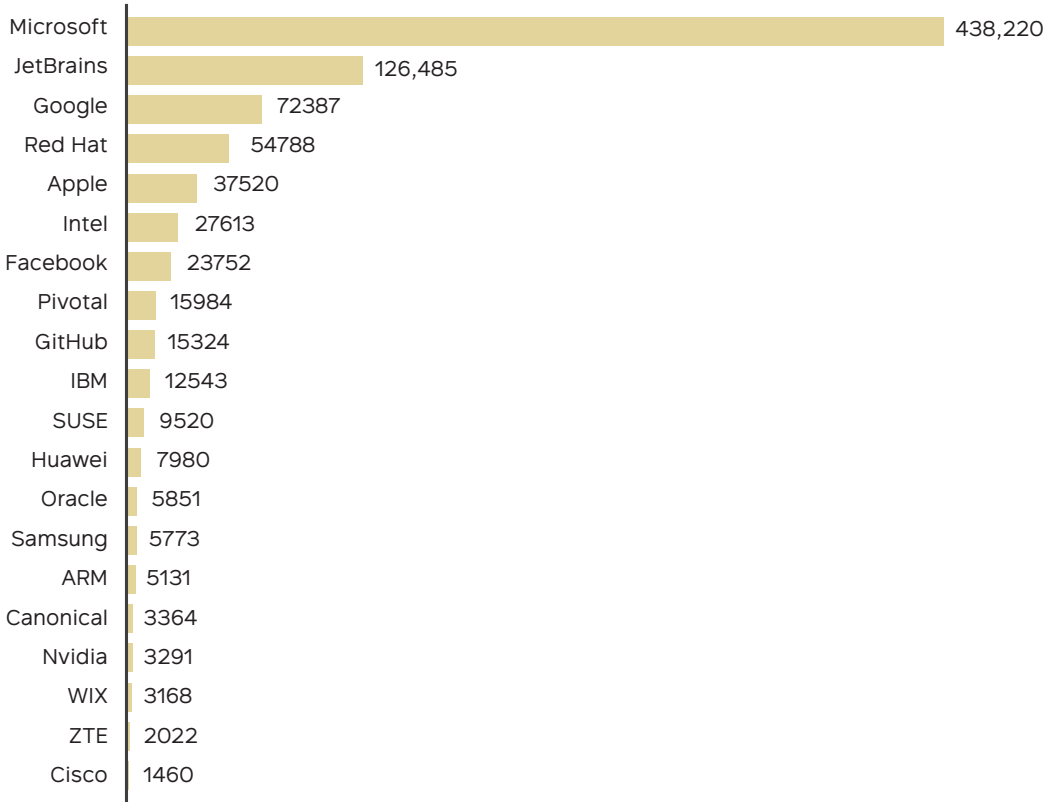
MOST ACTIVE FIRMS IN TOP GITHUB REPOSITORIES

We collected logs of contributions to the 135 repositories and used email address domains as proxy of employment. For example, a contribution by someone with a microsoft.com email address is deemed to be employed by Microsoft. Table 2.2 shows the ten, and Figure 2.2 the twenty, largest corporate contributors.

Table 2.2. Top-10 most active firms by number of commits submitted to selected 135 repositories, 01.01.2015 to 30.04.2019

Rank	Firm	Number of commits
1	Microsoft	438,220
2	JetBrains	126,485
3	Google	72,387
4	Red Hat	54,788
5	Apple	37520
6	Intel	27613
7	Facebook	23752
8	Pivotal	15984
9	GitHub	15324
10	IBM	12543

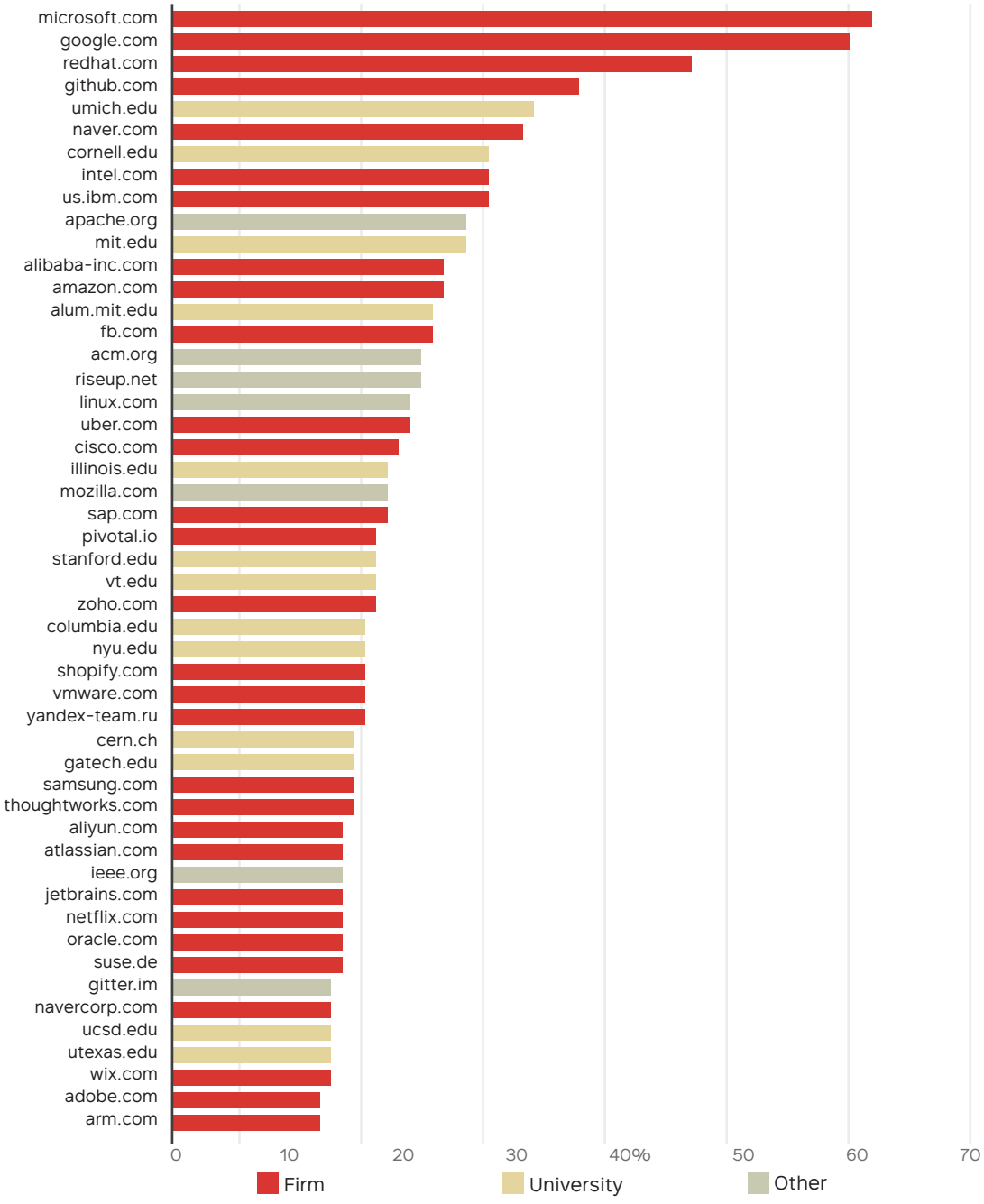
Figure 2.2. Top-20 most active firms by number of commits submitted to selected 135 repositories, 01.01.2015 to 30.04.2019.



NUMBER OF REPOSITORIES CONTRIBUTED TO BY FIRMS

Microsoft's strategy is clear: it bought GitHub in 2018 and is now maximising this investment. Microsoft was the lead firm contributor on GitHub during the 2015-2019 period, followed by other large IT firms seeking to secure technological leadership and to mutualize part of the development costs, according to the classic open source 'coopetition' model in which firms both compete and collaborate with each other.

Figure 2.3. Number of repositories a firm or other organization contributed to within the 135 GitHub repositories.



DISTRIBUTION OF FIRM CONTRIBUTIONS TO TOP GITHUB REPOSITORIES

Table 2.3 and Figure 2.4 show the projects most-contributed to by firm employees. Linux had the highest number of commits in our sample (340.000), 73% of which were authored by firm employees, yet employees of the leading firm (Intel) contributed 0.07% of commits: contributions to Linux are diverse, and community support is high.

Table 2.3. Number and proportion of commits in top-20 projects, firms contributing largest number of commits in 135 GitHub repositories, 1.01.2015 to 30.04.2019.

Project	Total	Firm-made commits	Proportion firm-made	Leading firm by commits	Proportion leading firm
torvalds/Linux	340710	247864	0.73	Intel	0.07
NixOS/nixpkgs	125205	63042	0.50	LogicBlox	0.01
Homebrew/homebrew-core	108709	54352	0.50	Charcoal Design	0.001
apple/swift	67197	40351	0.60	Apple	0.56
kubernetes/kubernetes	74201	40041	0.54	Google	0.30
Microsoft/vscode	49418	37366	0.76	Microsoft	0.72
tensorflow/tensorflow	56656	29515	0.52	Google	0.46
dotnet/corefx	32884	25660	0.78	Microsoft	0.71
DefinitelyTyped	54801	17920	0.33	Microsoft	0.10
aspnet/AspNetCore	34946	16486	0.47	Microsoft	0.39
spring-projects/spring-boot	17855	16357	0.92	Pivotal	0.89
ansible/ansible	31544	16252	0.52	Red Hat	0.06
elastic/elasticsearch	33983	16051	0.47	Elastic	0.23
rust-lang/rust	57790	15897	0.28	Red hat	0.003
facebook/react-native	16908	12704	0.75	Facebook	0.63
moby/moby	24472	11743	0.48	Docker	0.16
home-assistant/home-assistant	18876	10756	0.57	Affolter Engineering	0.04
pytorch/pytorch	17717	10492	0.59	Facebook	0.49
apache/spark	15180	8584	0.57	Databricks	0.30
storybooks/storybook	18968	7297	0.38	Dependencies	0.12

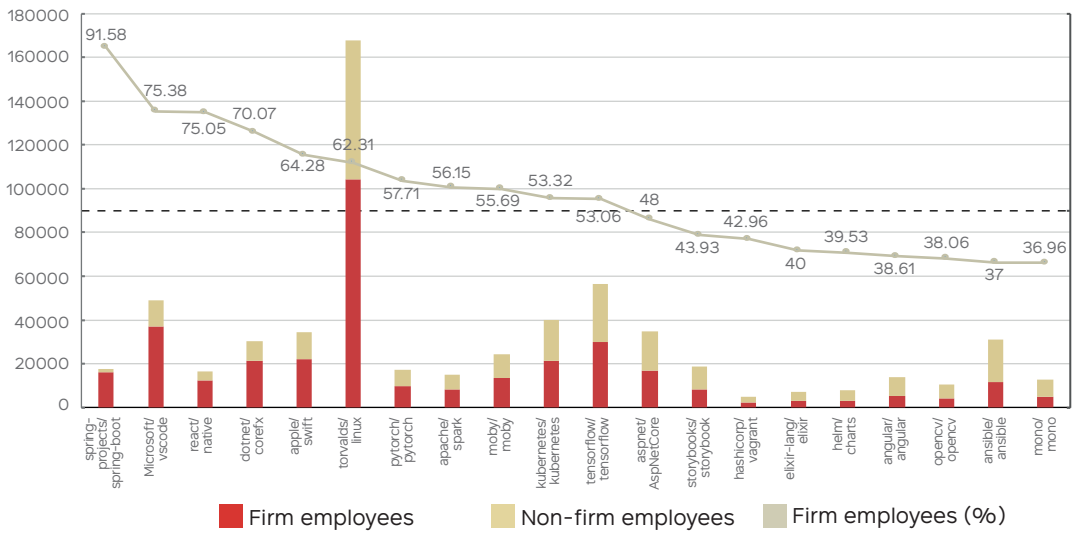


Figure 2.4. Count and percentage of commits by firm employees and non-firm employees to the top-20 repositories, descending by percentage of commits made by firm employees, 01.01.2015 to 30.04.2019.

FIRM CONTROL VS COMMUNITY CONTROL

Apart from NixOS, Homebrew, pytorch, and Apache Spark, which are community-managed projects, Table 2.3 shows that the top projects are either entirely developed by firms via their employees or managed by industrial consortia formed by firms with common market interests, whose boards are controlled by firms. With rare exceptions, in contrast to community-driven projects, technical governance in those consortia is not in the hands of developers but derives top-down from industry interests.

The technical development of single-firm projects is directly controlled by the strategic interest of the controlling firm, as can be seen in Table 2.3 in the case of spring-projects (Pivotal, 89%), or vscode and dotnet (Microsoft, 72% and 71%). In consortium projects inter-firm 'coopetition' balances the interests of firms. Who-contributes-how-much determines to what extent a project is controlled by a single firm, a group of firms, or a much larger group of stakeholders - as is the case for the Linux Kernel.

VOLUME OF FIRM EMPLOYEE CONTRIBUTIONS

In order to determine whether firm contributions are significant, we collected commit Source lines of code (SLOC), a software metric used to measure the size of a computer program by counting the number of lines in the text of the program's source code.

Figure 2.5 and Table 2.4 show to what extent firm employees contributed to projects. SLOCs contributed by firm employees are significant in volume, signifying that employees are not just gatekeepers harnessing volunteer effort, they do the actual work.

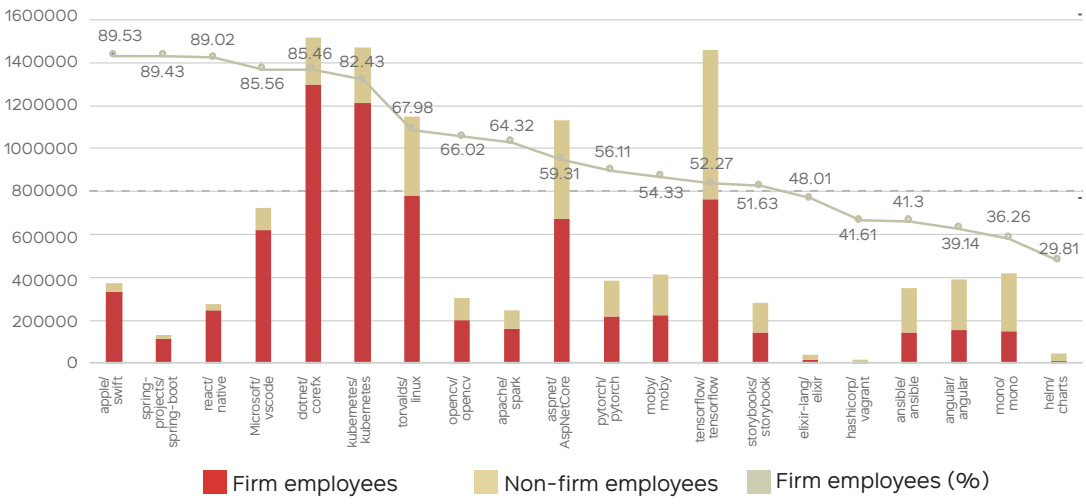


Figure 2.5. Count and percentage of SLOC by firm employees and non-firm employees of the top-20 repositories (descending by percentage of SLOC made by firm employees), 01.01.2015 to 30.04.2019.

Table 2.4. Count and percentage of commits and SLOC by firm employees and non-firm employees of the top-20 repositories, descending by percentage of commits made by firm employees, 01.01.2015 to 30.04.2019.

Project	Commits			SLOC		
	Firm employees	Non-firm employees	Firm employees (%)	Firm employees	Non-firm employees	Firm employees (%)
spring-projects/spring-boot	16045	1475	91.58	1162062	137322	89.43
Microsoft/vscode	36919	12059	75.38	6213205	1048348	85.56
react/native	12464	4143	75.05	2452807	302470	89.02
dotnet/corefx	21227	9066	70.07	12976731	2207941	85.46
apple/swift	22007	12230	64.28	3336874	390357	89.53
torvalds/linux	104431	63155	62.31	7792547	3670596	67.98
pytorch/pytorch	9922	7272	57.71	2156521	1686623	56.11
apache/spark	8443	6594	56.15	1603642	889692	64.32
moby/moby	13612	10831	55.69	2234220	1878238	54.33
kubernetes/kubernetes	21349	18689	53.32	12105955	2580226	82.43
tensorflow/tensorflow	30042	26573	53.06	7631597	6967787	52.27
aspnet/AspNetCore	16764	18162	48.00	6728611	4615271	59.31
storybooks/storybook	8231	10507	43.93	1454999	1363101	51.63
hashicorp/vagrant	2153	2859	42.96	65825	92383	41.61
elixir-lang/elixir	2891	4337	40.00	182327	197479	48.01
helm/charts	3100	4743	39.53	140820	331542	29.81
angular/angular	5336	8486	38.61	1521341	2366043	39.14
opencv/opencv	4049	6589	38.06	1991537	1024937	66.02
ansible/ansible	11543	19653	37.00	1441504	2048983	41.30
mono/mono	4771	8138	36.96	1509989	2654873	36.26

Table 2.4. provides detailed insights into the extent to which firms contribute to free and open source projects. Only SLOC for the top-20 projects with the largest proportion of commits made by firm employees were collected.

Linux has significant non-firm contributions. Kubernetes, which produces a popular open source solution for managing cloud infrastructure, has relatively few commits from firm employees (52%) but these contributions are more massive (82% of SLOC): volunteers contribute frequently, but to marginal sections of the product. Mono, though currently controlled by Microsoft, has strong volunteer involvement, likely stemming from the project's origins as a community-driven implementation of Microsoft's .NET technology. Pytorch, a scientific project, has relatively few firm contributions, just like Tensorflow (Google's machine learning framework).

TIME OF INDIVIDUAL COMMITS

When do firm employees and non-firm employees submit commits? Figure 2.6 shows that firm employee contributions dip during the weekend, whereas volunteer contributions remain constant.

Figure 2.7. shows how Source lines of code or SLOC relates to time of commit. Just like the median number of commits (Figure 2.6), the median SLOC number dips for employees during the weekend, whereas it remains constant for volunteers.

Figure 2.6. Median number of commits submitted by individuals in a week, 01.01.2015 to 30.04.2019.

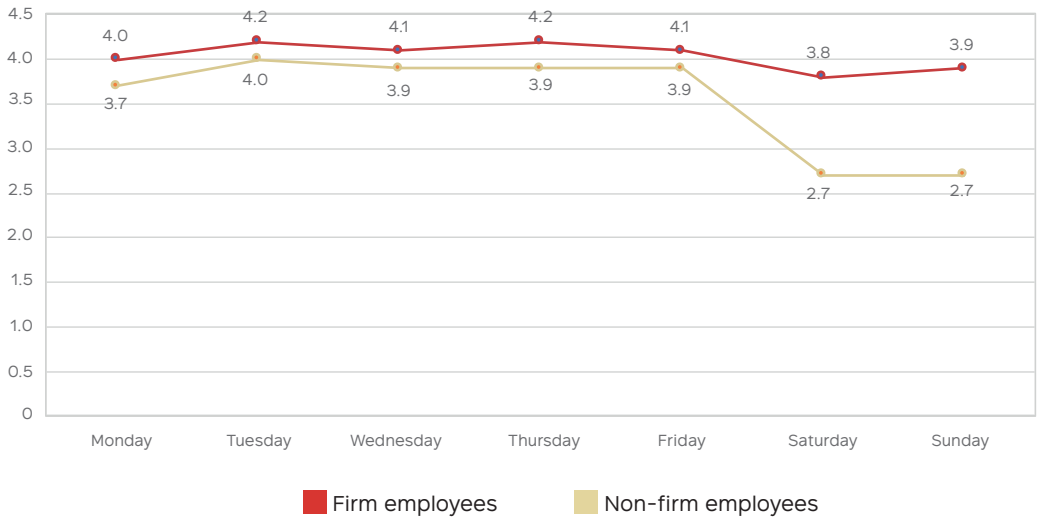
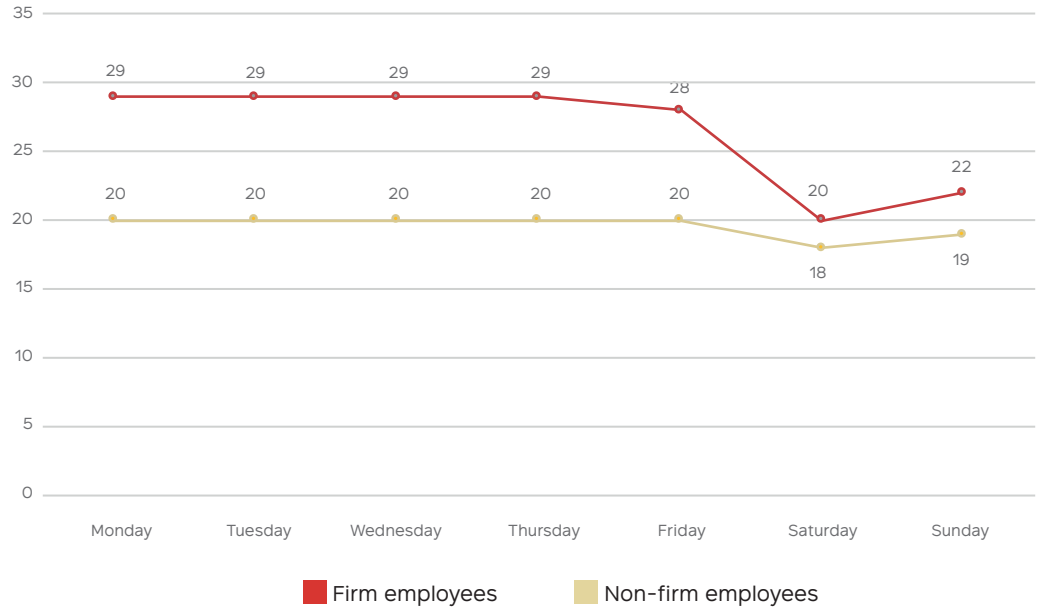


Figure 2.7. Median SLOC committed by individuals in a week, 01.01.2015 to 30.04.2019.



INCREASED FIRM INVOLVEMENT IN OPEN SOURCE

Rising firm involvement in FOSS is demonstrated in Figure 2.8: the number of non-firm email addresses remains constant, but there is a clear increase in the number of firm email addresses between 2015-2016 and 2017-2019. Firm email addresses are more numerous than non-firm addresses during the work week, and there is a sharp decrease of firm email account use on the weekend. Non-firm email addresses use also decreases on the weekend, but much less than that of firm email addresses. Volunteer labour is not just free, it is also constant.

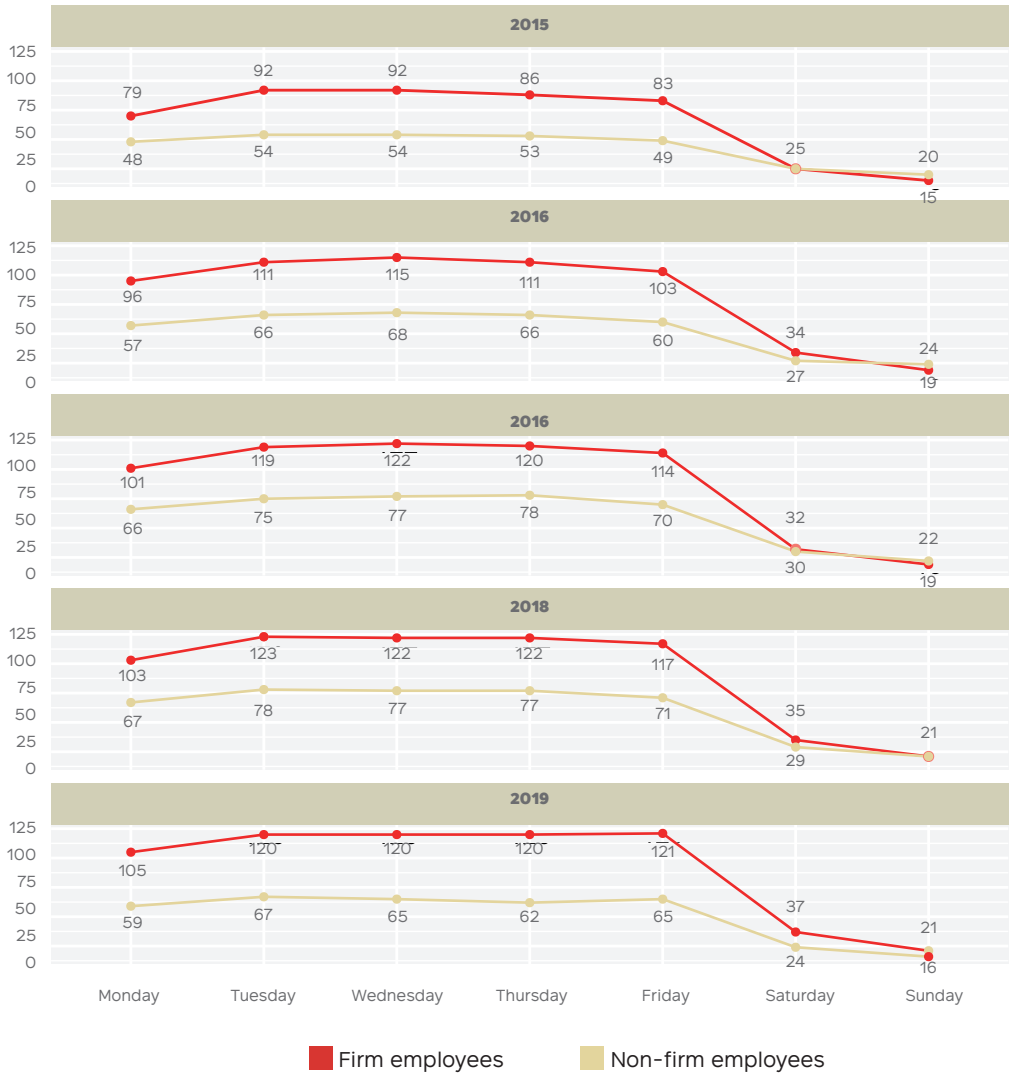


Figure 2.8. Average number of firm or personal email accounts of the 3,279 developers using both firm and personal email addresses to commit per week by year, 2015-2019.

LINUX IS THE CENTRAL NODE

Table 2.5 provides a summary of the firm and project network. The number of ties directed towards projects (indegree) and ties issuing from firms (outdegree) are separated into four quartiles, each containing an equal number of nodes. The firms and projects in the highest quartiles concentrate connections: a minority of projects is attracting the overwhelming majority of commits; a minority of firms are committing.

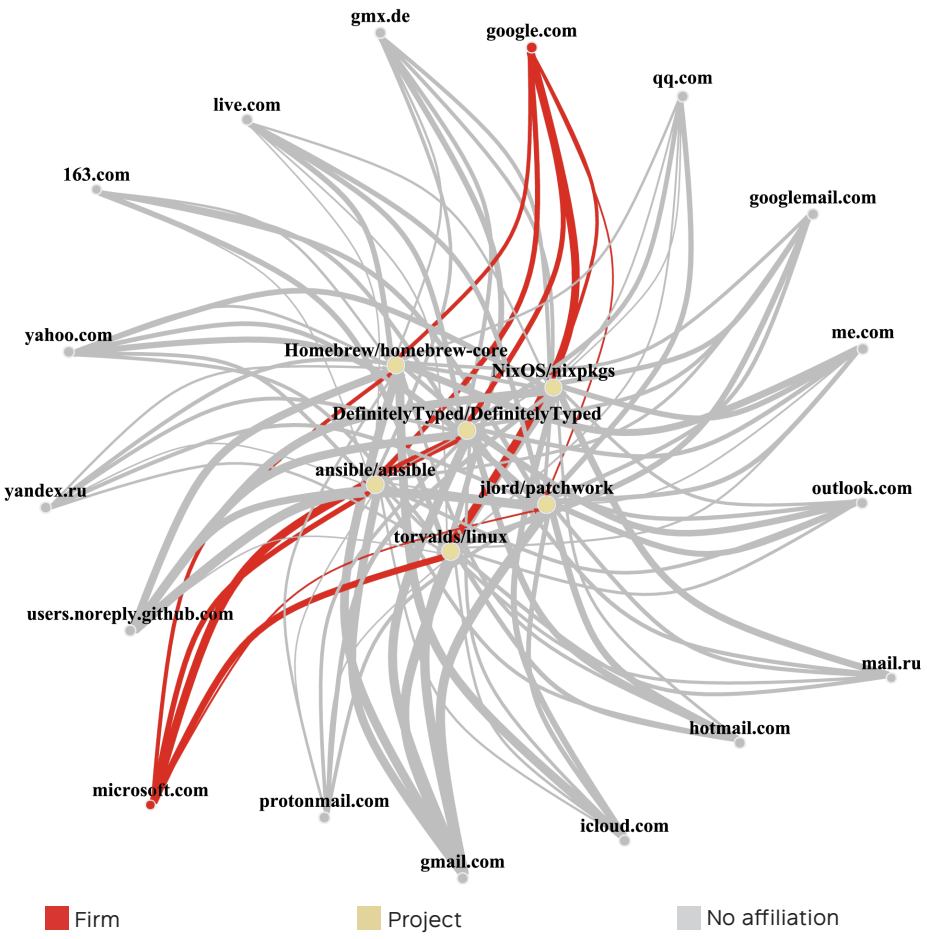
Figure 2.9 shows Linux is the central node in our coproduction subnetwork. This is due to its historical importance as the first massively collaborative free software project. In addition, a kernel is mandatory in both hardware machines and virtual machines in the 'cloud,' making a high-quality open source kernel hard to avoid.

Table 2.5. Summary of network of contributors and project on GitHub, 2019/05/31

Directed-weighted network
Size: 26,594 Projects: 135 Contributors: 26,459
No. of ties: 36,980

	Mean	min	1st Qua.	2nd Qua.	3rd Qua.	4th Qua.
Indegree of projects	275	1	49	131	271	3,378
Outdegree of firms	1.4	1	1	1	1	133

Figure 2.9. Subnetwork of projects and contributors on GitHub, indegree ≥ 1000 and outdegree ≥ 50, 2019/05/31



TOP CONTRIBUTORS

The top 100 email domains with the greatest number of commits were classified in five categories (Table 2.6). There are 41 firm domains among the top 100, corresponding to 11,538 GitHub accounts, which are responsible for more than 1 million commits.

The other significant block of commits is authored by unaffiliated email domains, linked to a much higher number of GitHub accounts (68,913): proportionately, firm employees contribute more.

Table 2.6. Top-100 contributors to selected 135 repositories on GitHub.

	Category	Email domains	GitHub accounts	Number of commits
Shared email domains	Firms	41	11,538	1,025,681
	FOSS foundations	19	1,069	149,293
	No affiliation*	20	68,913	1,060,168
	Research institution	1**	565	42,451
Individual domains	Individual developers	19	19	68,133
Total top-100	N/A	100	82,104	2,345,726
All domains	**N/A**	**26,459**	**113,614**	**2,824,690**

* Comprises email domains such as users.noreply.github.com, outlook.com, qq.com (common among Chinese developers), and mail.ru (a popular Russian email provider). **CERN.

The top-1,000 individual contributors contributed 56.6% of our total sample. In terms of SLOC, these top-1,000 accounts, or 0.88% of the total, disproportionately contributed 47.5% of the entire SLOC. Table 2.7 shows how these top-1,000 contributors are distributed. The most numerous (504) are firm email accounts. A smaller number (366) of non-affiliated email accounts made the most commits (44.5%, against 41% for firm accounts) though these commits were of a smaller size than those of firm accounts: 28% of total SLOC for non-affiliated accounts, against 61% for firm email accounts.

Table 2.7. Classification of top-1,000 individual contributors

	Firm	Foundation	Researcher	Individual developer / No affiliation	Total
No. of contributors	504	108	22	366	1000
(%)	50	11	2	37	100
No. of commits	659,116	191,435	32,414	716,271	1,599,236
(%)	41	12	2	44.5	100
SLOC	88,302,347	15,762,297	501,496	40,439,833	145,005,973
(%)	61	11	0	28	100

TOP CONTRIBUTORS USING BOTH FIRM AND NON-FIRM ACCOUNTS

Each account on GitHub can be linked to multiple email addresses. Among the 113,614 developer accounts who contributed to the selected repositories on GitHub, there were 10.5% who used more than one email address. A smaller number (2.9% of total) used both a firm email address and a non-firm email address to contribute. These accounts contributed 21.8% of the total number of commits.

Table 2.8. Contributions of developers committing with both firm and personal email addresses

	Total No. developers	Total No. commits	Mean No. individual commits	Median No. individual commits	Total SLOC	Mean individual SLOC	Median individual SLOC
All developers	113,614 (100%)	2,824,690	24.9	2.0	305,211,890	3,524	8
Developers committing with both firm and non-firm email addresses	3,729 (2.9%)	614,746 (21.8%)	183.9	26.0	74,536,051 (24.4%)	24,535	487
Top-1000 developers	1000 (0.88%)	1,599,236 (56.6%)	1642	953	145,005,973 (47.5%)	85,600	4,493
Top-1000 developers committing with both firm and non-firm email addresses	44 (0.04%)	186,118 (6.58%)	4,286	2,622	9,395,992 (3.08%)	204,917	118,096

These 3,279 developers contributed 24.4% of all terms of Source lines of code (SLOC). More than half (369,529) the commits contributed by the 3,279 individuals who both used firm and non-firm email addresses were made using a firm email address, whilst 233,429 were made using a non-firm or personal email address: volunteer contributions are significant.

Multiple email address use within the top-1,000 contributors shows that 44 of these GitHub account holders used both firm and non-firm email addresses (89 in total). These 44 GitHub accounts, less than 5% of top-1,000 (and 0.04% of all developers), contributed 11.6% of the commits by top-1,000 contributors. In terms of SLOC, their median and mean contributions were by far the highest of all contributors.

CONCLUSION

Firm employees massively contribute to the creation of digital infrastructure in top GitHub repositories but volunteers also play a significant role.

Firm investments vary between projects

INVITED COMMENT:
HERVE LE CROSNIER

THE ETHICS OF FOSS MUST BE ENACTED ACROSS THE ONLINE SERVICE ECONOMY

The emergence of the free software movement was a key moment for the world of computing as well as for the commons movement. Everyone knows the history that gave rise to free software as an alternative to the power of hardware manufacturers, and subsequently to that of large software companies. This cycle started in 1984, with the meeting of San Francisco hackers and Richard Stallman's creation of the Free Software Foundation on the one hand, but also with the desire to influence the rules of the global market being established at the same time by 'infogopolies:' the alliance between Hollywood, Big Pharma and the emerging software industry, soon joined by biotech firms.

This alliance strongly influenced the creation of the World Trade Organisation, which made intellectual property a central tenet of membership. These contradictory forces propelled ownership and sharing issues into the public sphere, generating new academic debates, which were taken up by activists ranging from public health specialists (the issue of drug patents) to people advocating for open creation (Creative Commons).

In this clash of ideas, the free software movement played a role that was both practical – without free software, the Internet would be very different than what it is today – and intellectually foundational: what does it mean to be 'free and open' and why is it necessary to protect freedom and openness from capture by industry and states? Yet the free software movement faces a dual challenge: where it used be emancipatory, it has become a 'development model;' and whereas it was user-centric, software has been included in online services, shifting the balance of power over ideas and data.

At the World Summit on the Information Society (WSIS) organised by the United Nations in 2005, Microsoft representatives suggested that free software was a specific 'development model' and therefore should not be included in the resolution... but above all, that no public market should mention the use of free software as a prerequisite. Naturally, civil society participants protested – for them the 'freedom' of coders and users was the key element, in this pre-social media period. Unfortunately, over time, Microsoft won: free software indeed became diluted into a development choice. By using free code, and by sharing code, each firm can expand its business and customer relationships, and Microsoft, IBM and Google are major contributors: who can afford to contribute because of the belief in sharing? It is difficult to maintain computer code over time, on the basis of volunteering alone, in the midst of a constantly changing universe. Commentators (myself included) were blind to the basis of any kind of work: how will it be paid for in the long run?

The opportunistic attitude of decision-makers, at all levels, also played a role: they only saw in free software a way to save money, by not paying for development. This was a strategic error: if open creation is not financed during tenders or public procurement processes, the disproportion between the private solution (benefiting from public subsidies via tenders) and the open solution (abandoned to its solely voluntary trajectory) quickly becomes glaring, mainly in terms of design and ergonomics.

Free software produces functional code, from the perspective of the person who developed it. The contribution of designers remains outside of this development, even though it plays a fundamental role in the usability of computer systems. This strategic error was also shared by large firms in the IT sector as demonstrated by the 'Heartbleed' case: this was a bug in the heart of the openSSL system, which was used by firms worldwide to protect digital exchanges. This free software had never been funded, but was widely used, and the discovery of vulnerabilities stemming from the slight amount of time devoted to maintenance endangered the majority of the world's websites.

Another issue is the transition from software as a tool for accessing information, to 'services,' directly

to the public via the web or via ‘cloud-based’ firmware. The collection and analysis of digital trace data is at the heart of the business models of search engines and social media, in which the ‘freedom’ of users not to be traced, monitored and influenced never comes into play.

We are therefore at a pivotal time in the struggle for digital liberties: the freedom to use, to understand, to share and to cooperate. Free software remains the necessary core, so that we, the ‘many eyes’ of developers (which explain why large companies are so interested in free software, or at least open source software) can verify that the services do not contain hidden elements to manipulate users.

But we must look beyond this: how can we build services, both for users and between professionals, that guarantee the autonomy of users, the absence of tracing, and an ethos of sharing that does not represent the means to attract the content and traces of users?

This ethical approach to IT services emerges in parallel to the movement of the commons that question the ethics of choices made in various communities. But the guarantee of a sharing ethos is never a given. It is an ongoing political struggle. ‘Public-common partnerships,’ which direct public investment towards operators which rely on the commons (in the field that concerns us today, the software and digital services commons) are major tools for the reorientation of the immaterial economy.

For this purpose, things being free as in ‘free beer,’ to return to the basic notions of free software, is no longer the issue. The question is how public money will be invested so that the ‘third sector’ of digital software and services becomes a major player for the freedom of Internet users... and therefore offers services with the same ergonomic appeal, the same accessibility and the same speed as those of current private actors who make their living off the data of users, and the sale of these profiles to the influence industry.

The free software movement showed us what digital politics could be. It is now necessary to rethink the political and ethical objectives that animated this movement in a situation where platforms which absorb private data are in a dominant position, and where the public economy does not know how to choose between immediate results and submission to private operators, and development projects financed to defend the freedom of users and the absence of manipulation in digital services. The aim is to develop a ‘free policy’ in the digital domain, beyond software itself.

3

COPRODUCTION AND VOLUNTEER LABOUR IN THE IT MEDIA

- Analysis in three online platforms (ZDNet, PCWorld and Slashdot) of articles featuring firms and projects.
- There are no mentions of volunteer or free labour: the twenty co-located articles with the highest proportion of labour-related terms discussed work in terms of career development and of a firm-project 'community.'
- Chinese firms such as Huawei and Alibaba are under-represented.
- The issue of cloud computing is central.

PROJECT AND FIRM TERMS FOR MEDIA ANALYSIS

We identified 50 projects and 50 firms to search for in media articles (Tables 3.1 and 3.2).

Table 3.1. Project-terms for article search in IT media (N:50)

	Literature review	GitHub	Wikipedia
Source	Survey of FOSS academic literature	Most active projects on GitHub by number of commits (lifelong till 30.04.2019)	Projects with most viewed Wikipedia pages
No.	10	20	20
Terms	Open source, Debian, GNU/Linux, Ubuntu, LibreOffice, Apache HTTP Server, MySQL, Free software, FOSS, F/OSS	React, TensorFlow, Angular, npm, Atom, Azure, IntelliJ, Visual Studio Code, Kubernetes, swift, kotlin, ASP.NET, Elasticsearch, Ansible, Docker, MariaDB, Mono, mongo dB, Google Go, Node.js	OpenCV , RStudio, rust, Kodi, Ping, Popcorn Time, Firefox, Mantis, Eclipse, Git, Blender, Octave, GIMP, 7-Zip, Rsync, X Window System, PuTTY, LaTex, FFmpeg, Drupal

Table 3.2. Firm-terms for co-location search (N:50)

	Forbes[10]	GitHub	LinkedIn[11]
Description	Top-100 technology firms by profit	Most active firms contributing to open source projects on GitHub by commits (lifelong till 30.04.2019)	Most appreciated by employees in 10 countries
No.	20	20	10
Terms	NVIDIA, Salesforce, ServiceNow, Square, Analog Devices, Palo Alto Networks, Splunk, Adobe Systems, Broadcom, ON Semiconductor Corp., Match Group, Tech Mahindra, Workday, Tencent Holdings, Micron Technology, SK Hynix, Twitter, Arista Networks, Baidu, Catcher Technology	Google, Red hat, Apple, Intel, Facebook, Pivotal, GitHub, AMD, Huawei, Mellanox, Oracle, Elastic, Arm, Databricks, Intellij, Gradle, MariaDB, Meteor, Chef, Crytek	Deloitte, Alibaba, SAP, Orange, Banco Santander, Safra, Bell, Capgemini, Zalando, Softtek

[10] https://www.forbes.com/top-digital-companies/list/#tab:rank

[11] https://www.linkedin.com/feed/news/linkedins-top-companies-of-2019-4392643/

IT MEDIA PORTRAYAL OF COPRODUCTION

To restrict our dataset to work-related articles, we created a glossary of labour-related terms, based on the ILO's *Thesaurus of labor and employment.*[12] We reduced them to their stems, as listed in Table 3.3.

We searched for labour terms in articles featuring both firms and projects, and collected articles featuring a minimum of five occurrences. The steps used to create our media article dataset are summarised in Table 3.4.

Labour and work were seldom addressed: Table 3.5 shows that only three out of twenty articles deal with paid work within the community-firm collaboration context. When they do so it is under the guise of the professionalisation of open source development, with an emphasis on 'job opportunities' and 'careers.' The role of volunteers and the value of volunteer contributions are not addressed.

[12] *Glossary of labour law and industrial relations-ILO*

Table 3.3.
Stems of labour-related terms

arbitr/; appoint/; backfill/; bargain/; breach/; bonus; boycott; career/; casual/; compensat/; conflict; demotion; dismiss/; dispute; earn/; emolument; employ/; fixed term; flexibility/; fund/; freelanc/; grievance; human resources; injunction; job; hir/; labour; labor; livelihood; maternit/; misconduct; out-sourc/; overtime; part-time; picket/; redundan/; salar/ self-employ/ subcontract/ termination; trade union; unfair; wage/; work/

Table 3.4.
Summary of media dataset

	Number of articles
Total unique articles[a]	13,174
Firm-project co-located articles[b]	1,424
Labour intensive firm-project co-located articles[c]	86

[a] Articles contain at least 1 project term.

[b] Articles contain at least 3 project terms and 1 firm term.

[c] Firm-project co-located articles featuring more than five occurrences of labour-related terms.

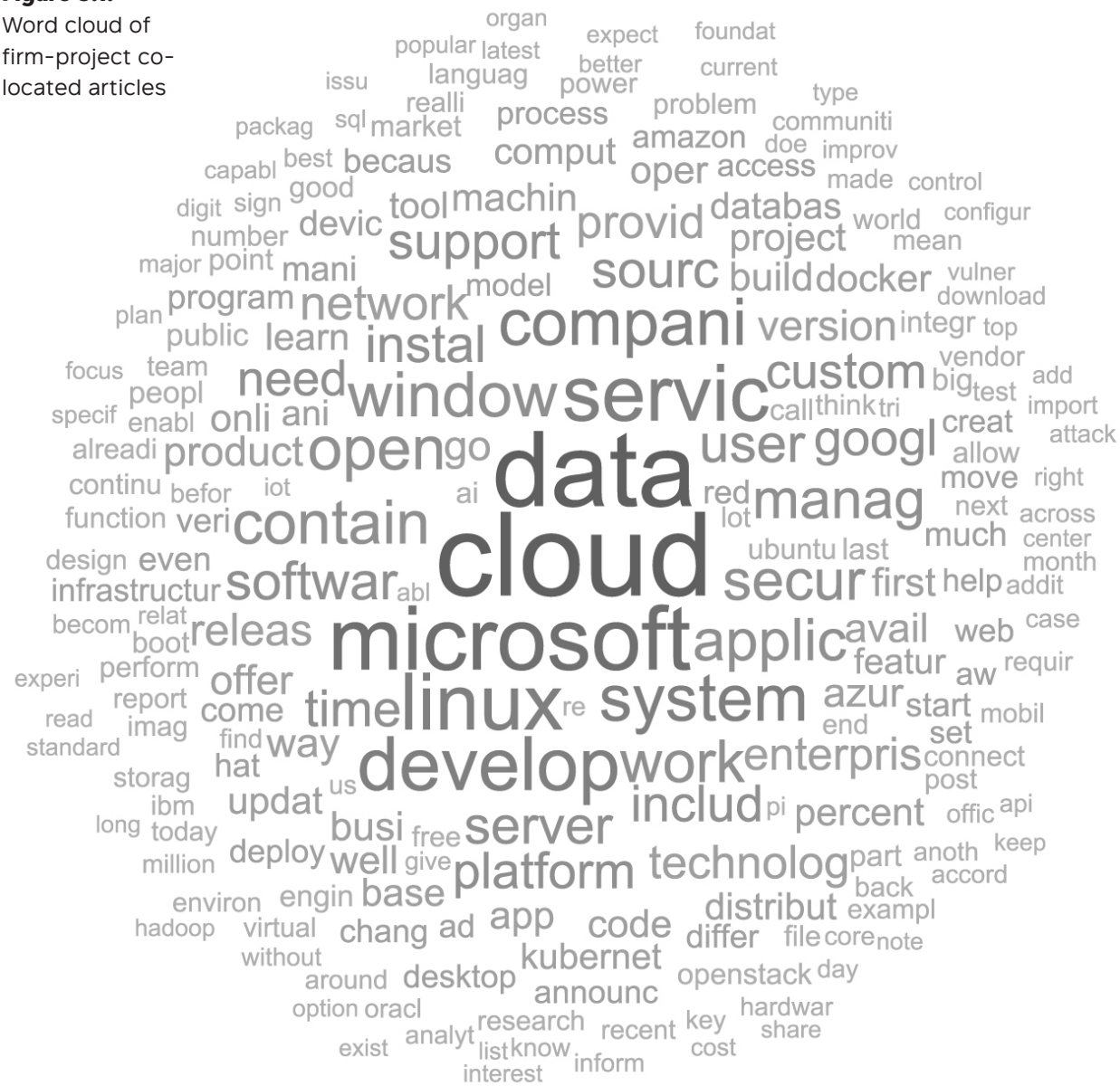
Table 3.5.
Top-20 co-located articles with the highest number of labour-related terms, 01.01.2015 to 31.05.2019

Articles titles	Term freq.	Media source
Is a coding boot camp right for you?	31	ComputerWorld
Don't just code: Career advice from the programming masters	16	ComputerWorld
Risk vs. Opportunity: Data use and availability in Australia	15	ZDNet
Google revamps Jobs search to streamline vacancy hunting	13	ZDNet
What is Kubernetes? How orchestration redefines the data center	13	ZDNet
AutoML is democratizing and improving AI	11	ZDNet
Linux at 25: Linus Torvalds on the evolution and future of Linux	10	ZDNet
Is Chrome OS right for you? A 3-question quiz to find out	10	ComputerWorld
Fedora, Manjaro, and Ubuntu MATE on the Raspberry Pi 2 & 3	10	ComputerWorld
How Salesforce got its developer conference right, while Microsoft, Apple, Facebook, and Google lost their way	10	ZDNet
Robotics adoption: The SMB guide to industrial automation	10	ZDNet
Open source professionals are more in demand than ever	9	ZDNet
Generation Z cheerfully welcomes our new robotic overlords	9	ZDNet
Protecting Your Clouds - Research Report	9	ZDNet
Cybersecurity predictions for 2016: How are they doing?	9	ZDNet
How Mark Shuttleworth became the first African in space and launched a software revolution	9	ZDNet
GraphQL for databases: A layer for universal database access?	9	ZDNet
Backhand slice: 5G and the surprise for the wireless cloud at the edge	9	ZDNet
Java vs. Node.js: An epic battle for developer mind share	8	ZDNet

DOMINANT THEMES

We created a word cloud of articles featuring at least three project terms and one firm term (Figure 3.5). The most central terms are ‘data’ and ‘cloud,’ followed by ‘Microsoft’ and ‘Linux,’ confirming that these are the key concerns and entities in this semantic space.

Figure 3.1.
Word cloud of firm-project co-located articles



SOME FIRMS ARE PORTRAYED NEGATIVELY BY THE IT MEDIA

We measured whether firms are portrayed positively or negatively by collecting the semantic environment (words preceding and succeeding a term) of firm terms. The firm with the highest number of negative words (Huawei) is Chinese, followed by Apple and Microsoft.

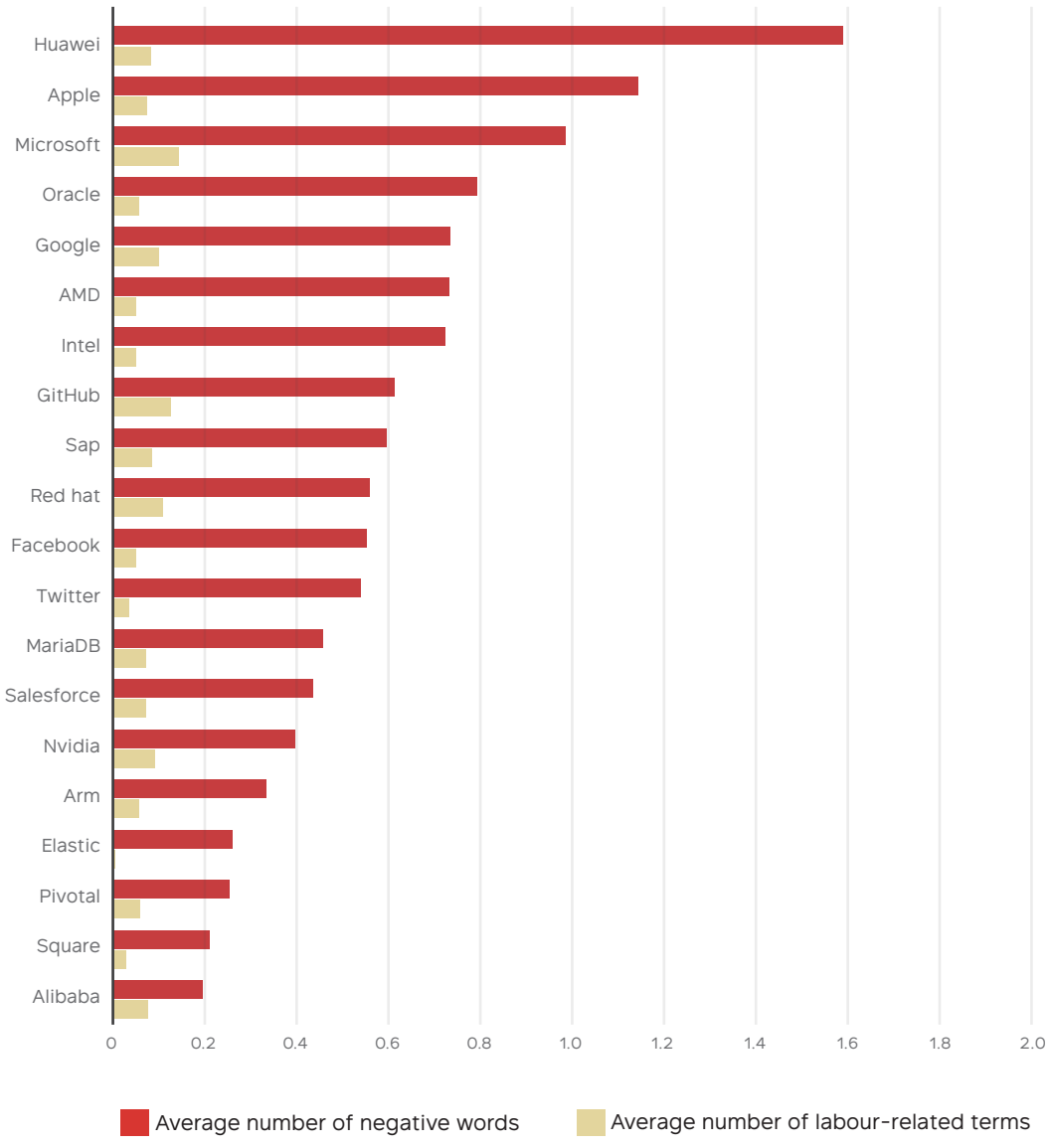


Figure 3.2. Average number of negative terms and of labour-related terms in semantic environments of firm terms, 01.01.2015 to 31.05.2019.

IT MEDIA COVERAGE OF FIRMS

The number of times firms were mentioned in our IT media article sample is summarised in Figure 3.3. Microsoft was covered extensively, which may account for its high negative word density (see Figure 3.2) whereas Apple and Huawei both received low coverage, but still elicited high negative sentiment.

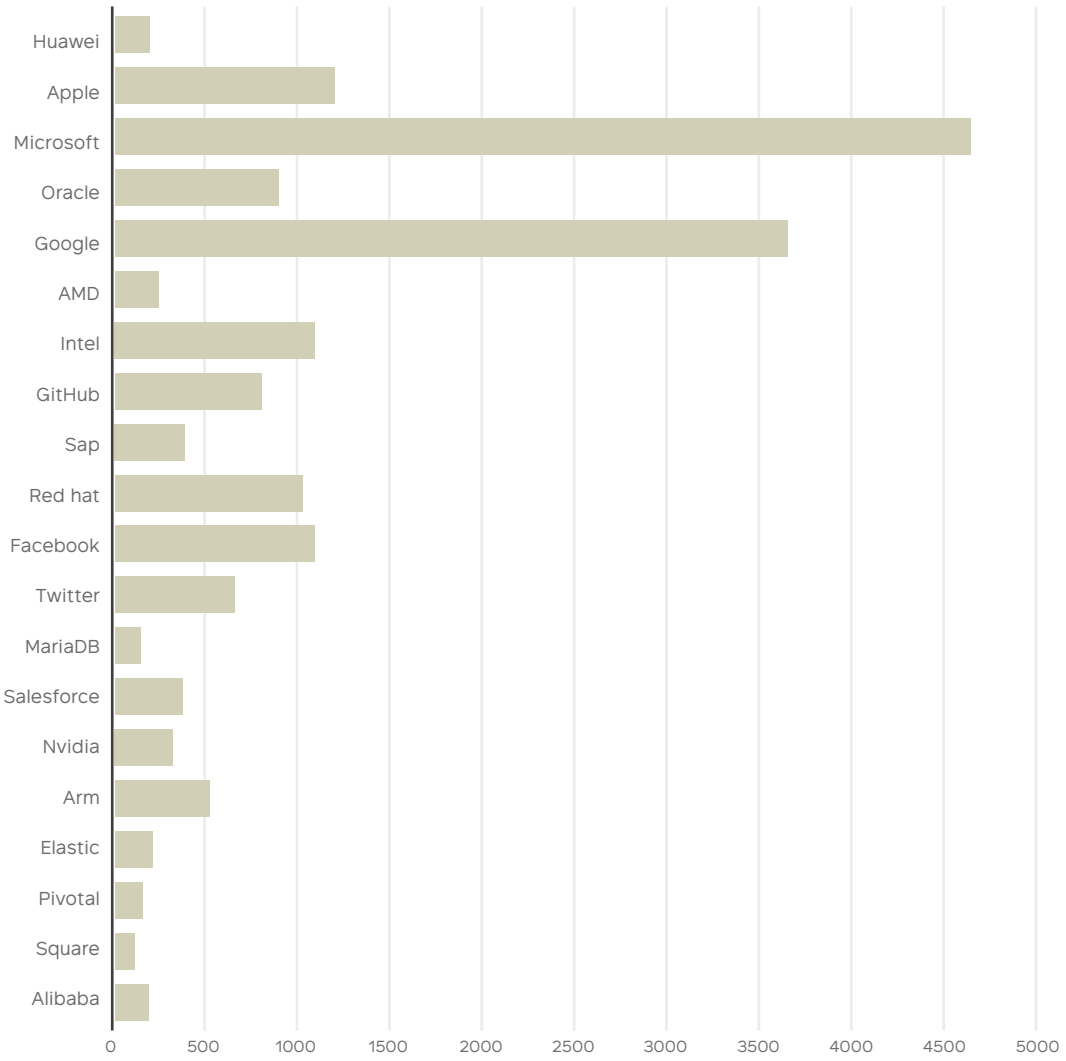
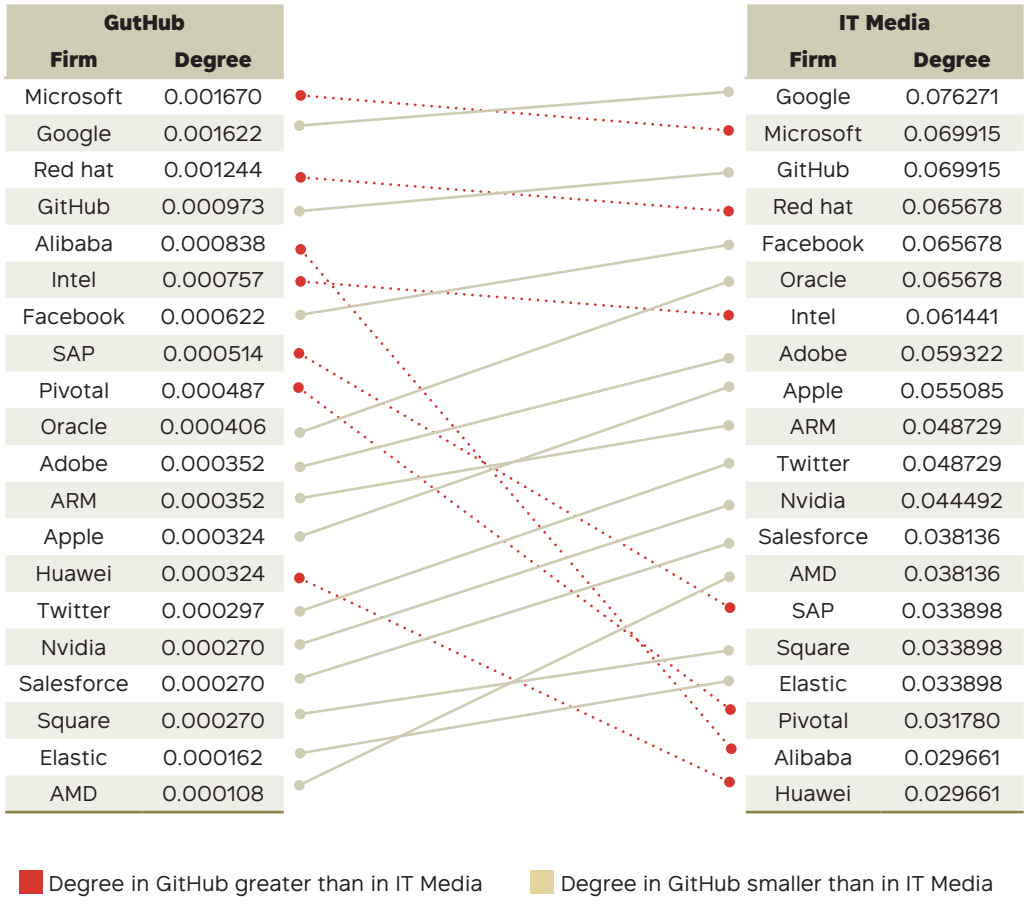


Figure 3.3. Firm media coverage, 01.01.2015 to 31.05.2019.

COMPARISON OF CO-PRODUCTION AND IT MEDIA NETWORKS

Much like on GitHub, a minority of firms and projects also concentrate the majority of mentions in the IT media. Our comparison of the networks (Figure 3.4) shows that some firms which are very active on GitHub are under-represented in the media sources we collected. These include Chinese firms such as Alibaba (e-commerce) and Huawei (telecommunications equipment and consumer electronics) as well as SAP, a German enterprise software firm, indicating a US-centric orientation in our IT media sources. Pivotal, a US firm, is also comparatively under-represented in the media, perhaps because its activities – supporting code development by testing quality – is not found to be interesting or 'exciting.' In contrast AMD and ARM (microprocessor manufacturers and Linux supporters) as well as Oracle (databases) are more prominently featured in the media.

Figure 3.4. Degrees of top-20 firms in project-firm networks on GitHub and IT Media



CONCLUSION

IT news media portrayals of coproduction do not mention free labour or the role of volunteers. Data and cloud are the central issues.

INVITED COMMENT:
THIERRY CARREZ

PREDATION IS NOT SUSTAINABLE, SO GIVE BIG TECH A CHANCE

With the success of free and open source software (FOSS), traditional software companies have transitioned from fighting it to adapting to it. This pivot manifested itself in two ways. The first is the adoption of methods of production that respect the letter of FOSS licenses but not their spirit, such as open core (using free software as a form of trialware) or the combination of copyleft licenses with exclusive Copyright License Agreements (CLAs) allowing a single party to retain full control. The second is the massive adoption of software service delivery over a network: Software as a Service (SaaS) or cloud delivery, allowing firms to capitalise on FOSS without necessarily contributing to its development.

In both cases an existential threat for FOSS is created - in the first instance, through the dilution of the value of open source (including the latest attacks on the meaning of open source and introduction of 'fauxpen' licenses), and in the second case, by eroding the sustainability of the commons. So what can be done about it?

In my opinion it is an error to frame this situation in absolute terms, by saying for example that 501(c)(3) foundations are universally good whilst 501(c)(6) foundations are intrinsically bad, or that firms are naturally opposed to individual contributors, or that larger firms are pure predators. While it is true that 501(c)(6) foundations behave in the best interest of their members (rather than 'in the public interest'), for the past 12 years the United States' IRS has stopped granting 501(c)(3) status to foundations contributing to the production of open source software – for example the recently-established Rust Foundation, a foundation created around a programming language, is a 501(c)(6). Even foundations that have individual members and a mission aligned to the public interest are assigned 501(c)(6) status, because their mission is focused on sustaining the production of software. So it is important to look at the details: consider the mission, governance and past behaviour of organisations, rather than which article of the US tax code they are attached to.[13]

I work every day on making sure computing infrastructure technologies are accessible to everyone, rather than just to a couple of giant Internet companies, by enabling open collaboration around open source projects focused on infrastructure. OpenStack was built as an open source alternative to Amazon Web Services (AWS), allowing anyone around the world to set up their own modern programmable infrastructure, and enabling the federation of smaller actors. But despite its success, OpenStack did not really make a dent into Amazon or Microsoft's success. A better approach might be to try to engage with those firms and to incite them to understand the value of sharing their own technologies: consider them as potential partners in the spread of innovation, rather than enemies.

Predation on the commons is, by definition, not sustainable. The very open source projects that those firms depend on will die in case too much predation occurs. There is an equilibrium to be found, and I am hopeful that we can get those companies to participate in the commons, and share more of their own technology in openly-governed projects. There is also an equilibrium to be found in the various end goals of FOSS: between creating a sustainable commons and maximising innovation, between the liberation of software and ensuring key end users benefits.

It will not be easy. But what is sure is that we cannot just apply the old recipes from the 1980s or from the 2000s, back when the attacks on FOSS were much simpler. Debunking 'Linux is a cancer' was easy. Proving that we can build better software with open source than with proprietary methods was easy. In the last ten years our environment evolved and firms adapted, while we rested on our laurels. Today, 'open source' is not enough. The FOSS community needs to come together and create new tools, terminologies, and labels to drive sustainable production of the commons in the 2020s.

[13] While this is a valid point, we stand by our chapter 4 findings: the two types of foundations are connected to clearly distinct discourses.

4

DISCOURSES ABOUT OPEN SOURCE IN THE IT SECTOR

- The discourses of firm employees and foundation representatives are divided into two camps:
  - Large IT firms such as Microsoft prioritise the collection of user data, so these firms promote cloud computing and Software as a Service. Their concerns were that projects adopt consistent performance, safety, documentation and technical standards, as well as diversify their contributor base. They do not address the cost of labour. For-profit 501(c)(6) foundations are aligned with this discourse.
  - Small IT firm such as Matrix or Nextcloud were absent from our GitHub sample. They produce open source software, following the 'open core' model. They are more likely to address economic and ethical concerns such as the sustainability of open source business models, the costs of labour, free riding, and control over open source. Non-profit 501(c)(3) foundations are in alignment.
- We mapped semantic networks, finding significant clustering which reflects the above division.
- We searched for the occurrence of firm keywords in two open source project email discussion lists, finding a clear contrast between a project controlled by a firm (Firefox) and one where a multiplicity of firms take part (Linux).

THREE OPEN SOURCE CONFERENCES IN 2019

FOSDEM19:
Free and Open Source Developers' European Meeting in Brussels, Belgium, 2-3 February. An international community event held on the campus of the Université Libre de Bruxelles, clearly developer-centric.

OSS19:
Open Source Summit Europe in Lyon, France, 28-30 October. An international corporate event held in Lyon's largest conference centre, organised by the Linux Foundation to promote commercially-focused 501(c)(6) foundations (CNCF, AWSF, etc.) and corporate discourses.

POSS19:
Paris Open Source Summit in Paris, France, 10-11 December. European/French business event, held in a moderately-size conference centre outside Paris, mainly geared towards smaller European IT firms.

ETHNOGRAPHIC METHODS

During the ethnographic phase in Lyon and Paris we selected keynotes, talks and workshops according to our main interest - the participation of firms in open source projects and the firm-volunteer community relationship – leading to a focus on management and organisational, rather than technical, presentations. Table 1 summarises the collected data. We were unable to be physically present in Brussels so watched videos of presentations, as they were all recorded. We also collected all the presentation summaries and author professional affiliations from the two conferences which included them on their websites.

Table 4.1.
Ethnographic and content data collection at three open source conferences.

	OSS19 Lyon	POSS19 Paris	FOSDEM19 Brussels
nbr of keynotes/talks/workshops	350	400	740
speakers/audience interactions	yes	yes	yes
collection method	on-site recording	on-site recording	online archives
presentations collected by	research team	research team	event staff
collection format	audio	audio	video
annotations + verbatims	54	51	57
additional interviews	booths + f2f	Q&A (author talk)	no
summaries scraped from website	yes	not available	yes

DISTRIBUTION OF EMPLOYER CATEGORIES

After comparing the results of the ethnography and content analysis of summaries, we created employer categories and matched presenters with these categories. Table 4.2 presents the employer categories and Table 4.3 their distribution in FOSDEM19 (Brussels) and OSS19 (Lyon).

The combined number of large and small firm employees is by far the largest component, in roughly similar proportions across events: 61% in FOSDEM19 and 71.6% in OSS19. In contrast, when it comes to the distribution of large and small firms within each event, there are striking variations, which correspond to each event's identity and purpose.

In FOSDEM19, an event organised in a public institution with a strong FOSS community focus, there was an equal number of large and small IT firm employees. In contrast attendees at the Linux Foundation's OSS19 were predominantly large IT firm employees (49%) whilst small IT firm employees only represented 22.6% of attendees. This accords with the event's location - Lyon's largest conference centre - and with the Linux Foundation's mission of acting as an interface between industrial and communal worlds.

The distributions of other types of speakers also reveal contrasts between the conferences. In the case of DEVs, at FOSDEM19 12.9% of speakers identified as independent developers while OSS19 comprised only 4.6%. There were twice as many attendees at FOSDEM19 (N: 738) than OSS19 (N: 349) so the difference in the number of DEVs attending is significant: 95 individuals at FOSDEM19, against 16 at OSS19.

Similar differences concern employees of large firms who consume open source products (bigUSR) - 1.9% at FOSDEM19, 6.3% at OSS19 – as well as representatives of foundations. FOSDEM19 mainly featured presentations by employees or representatives of 501(c)(3) (non-profit) foundations: 6% (N:44) with only 0.9% (N:7) representatives of 501(c)(6) (industrial consortia) foundations. Instead OSS19 mainly hosted representatives of 501(c)(6) foundations: 6.9% (N:24) of the total with only 1.7% (N:6) for 501(c)(3) foundations.

Table 4.2. Employer categories

IT-big	mainly big tech giants e.g. Microsoft, Google, Amazon
bigUSR	firms using open source software in their products e.g. car manufacturers
DEV	independent developers (or developers who chose not to identify as firm employees)
IT-small	less than 500 employees
EXP	experts, academics, lawyers, and hacktivists
FND6	501(c)(6) foundations (industrial consortia) e.g. Linux Foundation
FND3	501(c)(3) foundations (non-profit) e.g. GNOME Foundation

Table 4.3. Presenter employer categories by event

	N. FOS	N. OSS	% FOS	% OSS
IT-big	219	171	29.7	49.0
bigUSR	14	22	1.9	6.3
DEV	95	16	12.9	4.6
IT-small	231	79	31.3	22.6
EXP	78	28	10.6	8.0
FND3	44	6	6.0	1.7
FND6	7	24	0.9	6.9
oth	50	3	6.8	0.9
total	738	349	100.1	100.0

OPPOSING DISCOURSES IN THE DATA MARKET

Conference talks at open source conferences by large and small IT firm employee and 501(c)(6) and 501(c)(3) foundation representatives describe contrasting understandings of open source software. We present in Table 4.4 examples of these discourses, as well as of 'big user' firm employee discourses. We separate these understandings, linked to these firms' respective positions in the data/cloud market, into four discursive areas, which respond to the following questions:

- What is digital infrastructure?
- What are foundations for?
- What is our business model?
- What is the nature of the firm-community relationship?

Table 4.4. Firm and foundation discourses about open source

	IT-big/FND6 discourses	IT-small/FND3 discourses	bigUSR discourses
Digital infrastructure			
Large players control the data market today through platforms and systems created by web giants such as Facebook, Google and Amazon, or by historically dominant IT firms such as Microsoft. Data management and services depend on digital infrastructure whose development is based on open source code. Cloud computing enables 'Software as a Service' (SaaS), which outsources the deployment of software applications. In a 'traditional' mode, a software program is downloaded and executed by customers on their own hardware. In a SaaS mode, the program is never transferred onto the customers' machines, but is executed remotely on the provider's hardware and used online (e.g, within a Web browser). This creates a SaaS 'loophole' in the FOSS principle, as the service provider is no longer obliged to offer access to the code: copylefted software used as SaaS is not 'distributed' - since only a service is being provided - and therefore fails to trigger the reciprocal character of the GPL.	Digital infrastructure is a free resource to be used/exploited. Need for the centralised control of data, need to agree on open technical standards, which are less risky and costly. The key aim is to develop technically neutral infrastructure. How to assist in making open source technologies acceptable products for commercial users. The questions of data, the market it represents, and Software as a Service are absent.	Digital infrastructure (Free and open source software) are common goods produced by projects, small firms and communities of volunteer contributors. Proposals for decentralised or federated alternatives, implemented in open source products and promoted by activist collectives. The economics and politics of the data market, and the concept of 'data capitalism,' are frequently mentioned. A critical discourse questions the monopolistic management of data on centralised infrastructures controlled by large IT firms. Ethical issues include equality of access, data protection, the sustainability of open source, and control. Technology must be efficient, but it must serve independence, so there must be interoperability between platforms, servers, languages and clients to enable the coexistence of platforms, servers, and languages.	Big Users in the public administration sector report a change in business model: activity is no longer based on data collection and exploitation, but on open data and the supply of data to other actors, with a new activity and revenue model that might move from data exploitation to data quality certification.

	IT-big/FND6 discourses	IT-small/FND3 discourses	bigUSR discourses
Foundations			
501(c)(3) are non-profit e.g., the GNOME or Apache Foundations. 501(c)(6) are commercial, representing the interests of industrial consortia, e.g., the Linux Foundation.	Foundations are open, technically agnostic cocoons governing collaboration between end-users and suppliers so as to create standards and foster innovation and growth by pooling the costs of developing the underlying infrastructure.	Foundations are entities which have a key role to play: they protect projects and prevent some firms from imposing models that will reinforce their position in the data market. Foundations are advocates for open source values.	501(c)(6) foundations provide a safe haven for a large-scale deployment plan within a transparent framework.
Business models			
Business models are how firms define profit-making.	There is no open source business model, i.e., no way to generate revenue from the development of open source software Open source serves to reduce the development costs of digital infrastructure. The issue of payment does not deserve mention, it is non-problematic.	It is difficult to find a sustainable source of revenue, as their firm's activity is mainly centred (unlike large IT firms) on the production of open source software and services. While different types of business models are mentioned, from Open core, to service, to dual licensing, the challenge of charging users remains the key issue. The issue of software developer payment is explicitly raised.	The financial benefits of using open source solutions are difficult to quantify and less decisive than other types of benefits.
The firm-community relationship			
A project is an initiative to collaboratively develop a software solution, which may originate with volunteers, a foundation, or be sponsored by a firm. Projects usually have membership and governance rules. A community is a group of developers that share the project's goals and values and therefore contribute to its achievements.	There is an open source community which firms are part of, and relationships within this community are productive and beneficial for all parties involved. Issues of cost and remuneration associated with managing the community, or developing code, are not discussed. 'Project' and 'community' are synonymous. Professionalisation means improvements in software performance, quality and safety (requiring documentation) as well as the need for projects to respect ethical standards such as diversity when recruiting and managing a volunteer workforce. Non-coding activities (documentation, translation, communication…) are valuable.	'Project' 'and community' are clearly distinguished. There are costs involved in managing a project's community. One does not refer to 'the community,' but rather to distinct communities associated to specific projects, which are differentiated by the values they share and can dissolve or migrate to another project. The management of such communities need to respect 'open source principles,' in particular through the use of licenses strictly compatible with the Open Source Definition	Employees are willing to convert their company to open source. However, overcoming cultural and managerial resistance related to the fear of losing control, or to IP, quality or security issues remains a challenge. Not all Big User firms are at the same stage in this process. For some, evangelisation consists in encouraging developers to create an open source community limited to the firm itself, starting with small and non-mission-critical projects so as to learn how to 'do open source.'

THE COMMUNITY MYTH

Small IT firm employee discourses hinge around the issue of sustainable models. The projects upon which these firms rely can only survive by finding resources to pay core software developers before a community of volunteer developers joins and enriches this work.

In contrast cost is not an issue for big tech, since large IT firms have the resources to invest at a loss in open source digital infrastructure. Their goal is therefore to show that they are an integral part of this community of volunteers. The discourses of bigIT employees and foundation(c)(6) representatives are identical in this respect with IT news media discourse: firms and projects are part of a unified software development 'community.'

It is remarkable that the question of data and the market it represents is wholly absent from the discourses of large IT firm employees, almost as if it were a taboo topic. In the context of open source conferences, where these employees strive to present their firms as members of the wider open source community, it conceivably does not seem appropriate to dwell on the fact that digital infrastructure, when it is delivered via SaaS, is proprietary, thereby contradicting open source ethical principles – sharing, transparency, openness.

FIRM AND FOUNDATION DISCOURSES AND KEYWORDS

Presentation summaries conveyed five major themes: business, infrastructure, labour, neutrality/independence, and professionalisation. We attributed specific keywords to each theme. A sixth theme (diversity) was qualitatively distinct from the others but was not quantitatively significant, so is not included in the following table. Table 4.5 provides a sample of a simple count of keyword occurrences within presentation summaries, distinguished according to the speaker's employer type.

Employee categories predominantly use uniquely specific keywords which agree with their employer's business model and/or ethical values, in contrast with other employer categories: small and large IT firm employees were dominant users of keywords such as 'data,' 'platform' and 'infrastructure.'[16] However only small IT firm employees referred to 'funding,' and large IT firm employees were much more likely to refer to the 'cloud.' Foundation 501(c)(3) representatives were more likely to refer to 'privacy' and 'community' whereas Foundation(c)(6) spokespersons were the top users of 'documentation' and 'safety.'

[16] The ethnography found that that the data theme was overlooked by big IT firms. This can be explained by the fact that their references to 'data' took place in technical sessions which we did not attend; its regulatory dimensions were seldom discussed.

Table 4.5. Keyword frequencies (% of all presentation summaries) by speaker category in OSS19 and FOSDEM19 presentation summaries. Figures in brackets indicate total number of summaries per employer type.

	IT-b (390)	bigUSR (36)	DEV (111)	IT-s (310)	EXP (106)	FND6 (31)	FND3 (50)	Oth. (53)	Total (1087)
Business									
privacy	1.8	0.0	5.4	3.2	4.7	0.0	24	5.7	4.0
data	23.1	19.4	13.5	24.5	32.1	9.7	30	20.8	23.1
ecosystem	6.4	2.8	6.3	7.7	3.8	9.7	6	3.8	6.3
Infrastructure									
cloud	14.6	5.6	3.6	6.1	1.9	9.7	6	5.7	8.6
platform	17.4	13.9	15.3	14.8	8.5	0.0	8	13.2	14.4
infrastructure	10.5	8.3	4.5	10.6	5.7	0.0	12	1.9	8.7
decentralised	0.5	0.0	5.4	1.0	0.9	0.0	4	1.9	1.4
centralised	0.8	0.0	0.9	1.0	0.9	3.2	2	1.9	1.0
Labour									
community	17.4	19.4	14.4	18.1	19.8	19.4	38	7.5	18.1
freedom	0.3	0.0	1.8	1.6	0.9	0.0	16	1.9	1.7
user	20.3	11.1	25.2	19.4	29.2	9.7	42	26.4	22.1
Neutrality/Ind.									
license	2.8	5.6	2.7	2.3	0.9	3.2	16	5.7	3.3
control	7.9	8.3	14.4	9.7	14.2	6.5	14	11.3	10.1
ethical	0.3	0.0	2.7	0.3	0.9	0.0	2	0.0	0.6
open access	0.0	0.0	0.9	0.3	0.9	0.0	0	3.8	0.5
regulation	0.0	0.0	0.9	0.3	1.9	0.0	6	0.0	0.6
funding	0.0	0.0	0.0	2.6	0.0	0.0	2	0.0	0.8
transparency	0.3	0.0	1.8	1.0	0.0	0.0	6	0.0	0.8
Professionaliz.									
safety	2.3	8.3	1.8	1.9	4.7	9.7	2	3.8	2.9
documentation	2.1	0.0	1.8	2.3	4.7	6.5	4	3.8	2.6
open standard	0.3	0.0	0.0	0.6	0.0	0.0	4	0.0	0.5

MAPPING LEXICAL WORLDS

Conference presentation summaries were analysed using the iRamuteQ software, based on the Reinert method, which counts word co-occurrence and builds a series of lexical clusters.

Definition and social regulation of open source projects (31% of the total corpus). The first lexicon in Figure 4.1 comprises terms used to discuss the specificity of the social and practical world of open source such as principles of organisation and regulation: open source, project, software, license, as well as ethical values: collaboration, inclusion, contribution, diversity, commitment, etc.

Mixed lexicon relating to technical objects, services, technical operations and issues: cloud and infrastructure, hardware and kernel, languages, databases and code (69% of the total corpus). Cloud computing is an entity that is deployed and works, but has no objectives as such. This technical lexicon also describes embedded computing: the use of free code in objects that cannot operate without computing, such as air conditioners and motor vehicles. It also contains terms relating to higher levels in the open source software stack such as databases and languages, as well as algorithms, search engines and the types of operations or analysis that are performed on numerical data (query, collect, pattern, semantics, graph, visualisation, etc.).

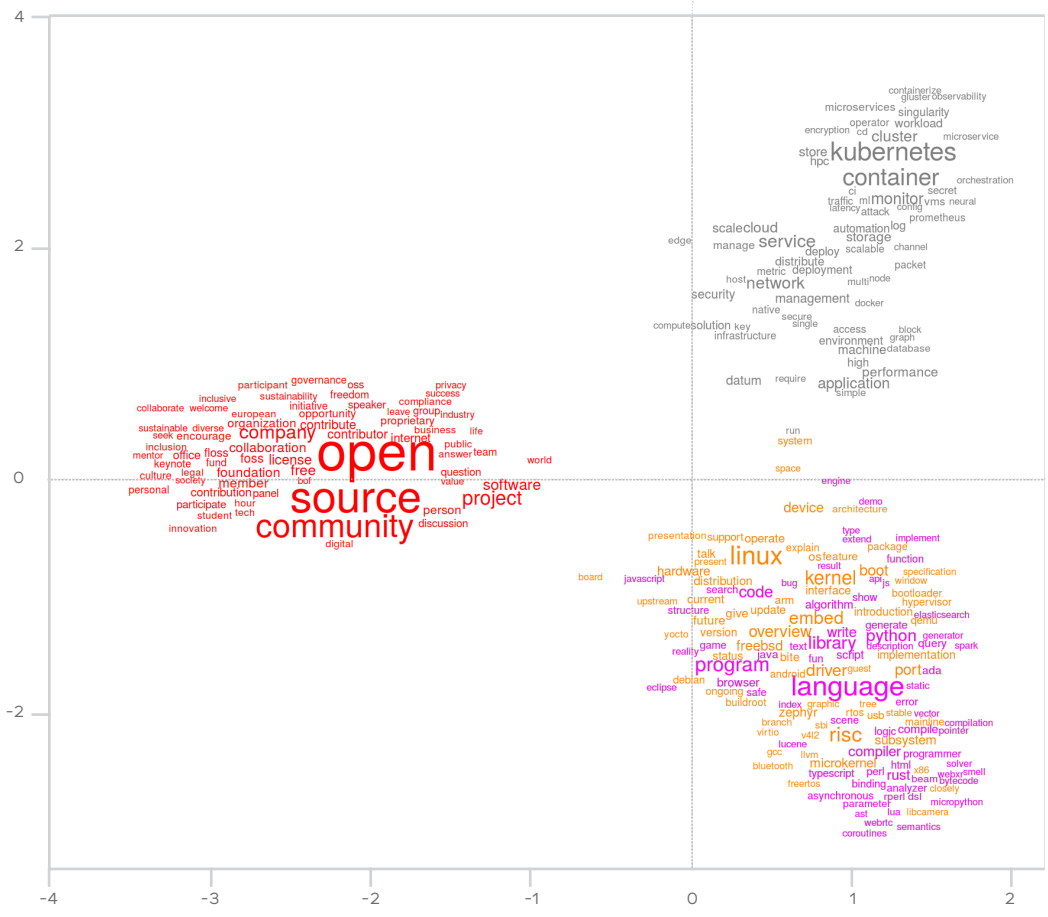


Figure 4.1.
Two major lexicons: regulation of open source (red, left of vertical axis), and technical objects and operations (grey, orange, purple, right of vertical axis)

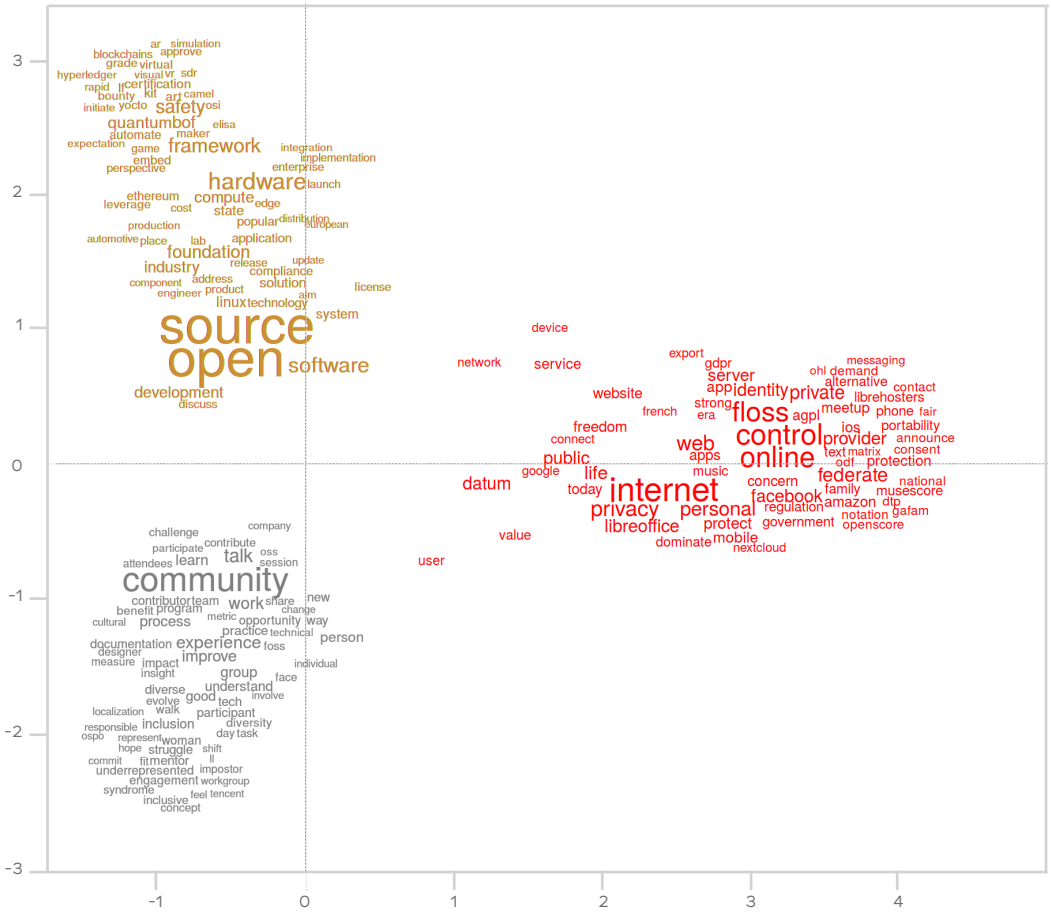


Figure 4.2.
The open source/ community lexical universe is composed of 3 sub-lexicons: open source (light brown, top left), community/ diversity (grey, bottom left), control/privacy (red, right of vertical axis)

Open source sub-lexicon, light brown (34% of the regulation lexicon). Top left in Figure 4.2 is a lexicon of industry, business, and economic strategy around open source. Sets the guidelines for the professionalisation of this field, both as a hegemonic mode of open source deployment through the digital industry, and as a 'corporate' discourse now integrated by the FOSS world, which had preserved its autonomy during the 1990s/2000s. Features terms that define, in the open source world: main players (platform, engineer, license, industry, foundation, distribution, consortium, enterprise, client, project, etc.), objects (software, hardware, license, linux, system, certification, application, technology, etc.) and actions (development, compliance, address, compute, implementation, automate, embed, integration, simulation, initiate, etc.).

Community/diversity sub-lexicon, grey (44% of the regulation lexicon). Elements to portray work experiences and firm and project organisation; a lexicon of personal fulfillment both in terms of success (successful, opportunity, individual, hope) and challenges (struggle, impostor syndrome, under-represented, barrier, etc.). This sub-lexicon also features elements of relationship management (participate, challenge, program, team, measure, foster, organisational, OSPO, etc.) that accord with a social justice vocabulary which characterises diversity policies and the transformation of systemic power relations between social groups (inclusive, represent, group, woman, mentor, shift, etc.). 'Community' operates as a catch-all term, designating corporate teams, project teams, communities of users, etc.

Control/privacy sub-lexicon, red (22% of the regulation lexicon). This sub-lexicon addresses the regulation of technological relations of power, with discourses addressing Internet control and privacy issues. It includes brand/service names, with items located near the centre of the graph (google, facebook, libreoffice), while items situated at the right edge of the graph are specifically related to privacy and internet control issues (except for 'GAFAM' or 'amazon' that probably appear because they are under-represented in the summaries corpus). Legal-political terminology (dominate, protect, regulation, protection, consent, enforce, gdpr, freedom, fair, concern, government, etc.). Relatively common technical terminology (near the centre) such as data, network, device, service, user, document, as well as more specific vocabulary about personal data and their exploitation (federate, mobile, messaging, phone, server, etc.).

DISCOURSE AND COUNTER-DISCOURSE

In sum, Figure 4.2 shows that the 'community' and 'open source' sub-lexicons are opposed around the horizontal axis of the graph: speakers use one or the other of these lexicons, rarely both at the same time. However, these sub-lexicons dialogue to some extent. They are aligned in parallel to the vertical axis and belong to the same formal plane. In contrast, the 'control and privacy' sub-lexicon is relatively autonomous, showing that the 'community' and 'open source' sub-lexicons effectively mirror each other as corporate discourse and counter-discourse, defining work relationships in terms of access and therefore never as 'labour' issues.

FIRM DISCOURSES IN PROJECTS

We collected data from the email discussion lists of Linux, the historic centre of the FOSS world, and Firefox, the most widely used open source browser. Table 4.6 presents the details.

We performed computational content analyses to track the presence of key firm discourse terms. We wanted to understand in what way these discourses are being used by volunteers and firm employees in these two projects. Initial results showed there was little discussion of diversity, and more generally that many keywords were not frequently used. For this reason, we reduced the number of keywords to a few core instances, listed in Table 4.7. Table 4.8 lists how these keywords were used in the Firefox and Linux email lists.

Table 4.6.
Email discussion lists corpus

Project	Source of data	Type	Collected messages	Duration
Firefox	https://mail.mozilla.org/pipermail/firefox-dev/	Mail archive	13,700	Jan 1, 2015 – Apr 30, 2019.
Linux kernel	https://marc.info/?l=linux-kernel	Mail archive	228,000	Jan 01, 2018 - April 30, 2019

Table 4.7.
Restricted list of keywords

Theme	Keyword
neutrality	accessibility
neutrality	compliance
infrastructure	database
infrastructure	infrastructure
business	ecosystem
business	privacy
business	security
labour	employer
labour	value

Table 4.8 shows that firm engagement with keywords is high in Firefox (approx. 80-90%). In contrast Linux has a more diverse contribution mix, with firm domains representing between 60% and 70% of keyword users. In terms of content, neutrality is the least engaged with theme.

Table 4.8.
Sum of restricted keyword category frequencies in Firefox and Linux kernel email list archive

Theme	Business	Infrast.	Labour	Neutrality
Total	24381	5361	91549	640
Firefox	**922**	**238**	**413**	**101**
Firm-email domains use of keywords	727	211	337	86
Percentage in total email lists	3.0%	3.9%	0.4%	13.4%
Percentage in Firefox list	78.9%	88.7%	81.6%	85.1%
Linux	**23459**	**5123**	**91136**	**539**
Firm-email domains use of keywords	14005	3498	47221	326
Percentage in total email lists	57.4%	65.2%	51.6%	50.9%
Percentage in Linux list	59.7%	68.3%	51.8%	60.5%

USE OF FIRM KEYWORDS IN PROJECT DELIBERATIONS

Table 4.9. Most frequently mentioned restricted keywords in top 20 email domains, minimum sum of keywords frequency in an email: 20 (number in brackets: total no. restricted keywords)

Project	Business	Infrastructure	Labour	Neutrality
Linux	schaufler-ca.com(4254)	intel.com(1625)	linuxfoundation.org(13596)	gmail.com(55)
	linuxfoundation.org(2904)	linuxfoundation.org(378)	kernel.org(11219)	linuxfoundation.org(48)
	redhat.com(2140)	kernel.org(315)	gmail.com(8104)	gateworks.com(46)
	chromium.org(1738)	arm.com(249)	redhat.com(4527)	kernel.org(40)
	linux.vnet.ibm.com(1404)	gmail.com(196)	arm.com(3624)	collabora.com(26)
	linux.ibm.com(1110)	redhat.com(139)	intel.com(3304)	li0ro.org(23)
	gmail.com(946)	schaufler-ca.com(117)	li0ro.org(3173)	cadence.com(20)
	intel.com(844)	linux.intel.com(113)	linux.intel.com(2579)	
	kernel.org(806)	codeaurora.org(109)	codeaurora.org(1730)	
	sargun.me(512)	li0ro.org(108)	mediatek.com(1463)	
Total Linux	23,459	5,123	91,136	539
Firefox	mozilla.com(519)	mozilla.com(151)	mozilla.com(219)	mozilla.com(73)
	gmail.com(134)	gmail.com(31)	gmail.com(98)	
	softvisioninc.eu(32)			
	rtfm.com(25)			
	gavinsharp.com(21)			
Total Firefox	922	238	413	101
Total	24,381	5,361	91,549	640

CONCLUSION

Large firms can afford unprofitable digital infrastructure development activities as these costs are offset by their lucrative data-storage and processing activities. BigIT and Foundation(c)(6) discourses are identical with IT news media discourse: firms and projects are part of a unified software development 'community.' The IT firm sector is not monolithic: there are contradictions and conflicts, stemming from large and small firms' status as users, or primary producers of open source software, which lead to alternative proposals. These alternatives are explored in the remainder of this report.

INVITED COMMENT:
DANIÈLE BOURCIER

BUILDING COMMON KNOWLEDGE: THE SCIENCE COMMONS

In the age of open science, the use of free licenses has become necessary to encourage the sharing of research as well as to protect the scientific community.

But how do these scientific knowledge commons stand in the face of policies implemented by the state or public institutions such as universities? Scientific knowledge includes both private assets, i.e. information produced by R&D labs which is patented or (exceptionally) placed in the public domain, and public goods, i.e. information produced by universities that apply internal policies (patents or open archives). In addition, there are more and more common goods, that is to say information produced by researchers which is useful to researchers and the public, regardless of their nationality and status, thereby evading state regulation.

The impacts of digital commons on the world of science have been multiple both in terms of property and sovereignty. The traditional right of ownership in terms of scientific data has been exploded into different kinds of rights: access, direct use, differed use, management, exclusion, alienation (resale). International sharing projects thus respond better to the requirements that results be rapidly circulated. Examples include GENBANK,[17] which uses mirror sites in three different countries that are updated every night, or GBIF,[18] a global cluster of local biodiversity databases.
Finally, the Science Commons project, launched in 2005, was meant to create a community of scientists aiming to define strategies and tools for the dissemination of rapid and effective scientific research, as enabled by network connectivity. States were not involved in the assembly. These commons included templates for open legal model agreements and web systems that made scientific results readily available.

These data sharing communities bring together researchers and institutions who consider that any impediment to the flow of results is not only ineffective, but contrary to the fundamental principles of generalised and open pooling of knowledge. A good example of this type of cooperation amongst the world's scientific community is the Global Alliance for Genomics and Health (GA4GH), which is developing an International Code of Conduct for Genomic and Health-Related Data Sharing.[19]

Research consortia share the same spirit. Starting from the 1990s, they combined private and public laboratories to promote the sharing of techniques, tools and data. They were often created to compete with private biotech companies so as to circumvent their nascent monopolies and form new commons. Private consortium agreements demonstrate a good deal of inventiveness in the area of regulation and management of commons: the category of common property, for example, which they had invented, was adopted by the European Community.

The sharing of data, a voluntary process emerging from a community, differs from *open data*, which is a scientific policy of open public data put in place by states for the public and businesses. Scientific commons are not only open data: like other commons, they depend on contractual rules that are rules of governance developed by the commoners themselves who are, in this case, scientists.
These agreements are presented as a solution to find a balance between various public and private interests, that is, between the market and the state. Sometimes firms may want to access common data without having to produce them themselves, sometimes firms (such as Merck) have participated in the creation of commons and placed them in the public domain.

Extract from: D. Bourcier (2021) Interroger les communs numériques face à l'État et au Marché, in D. Bourcier, J. Chevallier, G. Hériard Dubreuil, S. Lavelle and E. Picavet (Eds), *Dynamiques du commun - Entre État, Marché et Société*. Paris: Editions de la Sorbonne. Translation : DCPC.

---

[17] https://www.ddbj.nig.ac.jp/

[18] https://www.gbif.org/

[19] https://www.ga4gh.org/

5

STRATEGIC RESPONSES TO APPROPRIATION AND FREE RIDING

- The structural imbalance between what big tech obtains from the digital commons and what it gives back should be challenged.
- Possible strategies include: Tracking GAFAM predatory behaviour and making it better known to the public; Fostering a debate in FOSS about predation and the recognition of volunteer labour; Evaluating the viability, cost-effectiveness and benefits of federated technological services; Identifying free-riding firms.
- Beyond FOSS, it is necessary to consider the issue of digital services; and beyond digital services, it is necessary to discuss future labour relations.
- In short, new connections are needed: the contributions of researchers to digital infrastructure should be identified; there should be more recognition of the voluntary production of digital commons by the state.

GAFAM APPROPRIATION

If big tech firm employees and the IT news media obscure predatory behaviour and highlight the 'open source community,' it could be because the position of the GAFAM is morally untenable. Large tech firms are negating the stated ethical values of free and open source software: reciprocity, transparency, openness. However not all GAFAM engage in the same behaviour, so it is necessary to track precisely who does what. Once this has been clearly documented, further action could include making these facts as widely known as possible.

A GAFAM-wiki could be used to document instances of predatory big tech behaviour. Persons or organisations interested in contributing to or funding this initiative should contact the DCPC.

DEBATE IN FOSS

A debate about GAFAM appropriation, the recognition of volunteer labour, and other salient issues should be initiated and led by the FOSS community itself. But where could this happen?

We will release our 2016 survey of the Debian project as a DCPC report in the second half of 2021, and use this release as an opportunity to sound out the Debian project and community about these issues in a follow-up 2021 survey. See Chapter 6, p. 65.

EFFECTIVENESS OF ALTERNATIVE DIGITAL SERVICES

Federated solutions provided by Framasoft (see Gosset, this report; Gruson et al., this report) rely on connections with trusted others. This model may constitute a viable alternative for video-conferencing, collaborative pads, and other functional tools. There could be less potential for constituting viable alternatives when it comes to entertainment, as a small network is unlikely to be able to compete with the quasi-infinite content available on YouTube, for example.

We need objective and precise evaluation of federated tools and services: how user-friendly and effective are they? If they are found to be effective, how can public opinion and policy be oriented towards increased adoption of these technological services? One possibility would be to return to Framasoft's origins in schools. Persons or organisations interested in contributing to or funding this initiative should contact the DCPC.

DEVELOPING RECOGNITION OF VOLUNTEER LABOUR BY THE STATE

Widespread automation and job losses, combined with catastrophic environmental degradation, mean it is time to consider alternatives to unlimited growth and full-time paid work. A rebalancing of paid and volunteer/community activities offers a sensible path towards ensuring more people have access to work, and energy needs decrease.

The recognition of volunteer labour by the state, or increased involvement by the state in volunteer-run collectives, is not without risks (e.g., associating a monetary value to voluntary activities could weaken some positive aspects of community work) or challenges (e.g., cultural differences between communities and public authorities).

More public debates about the recognition of volunteer labour are needed. See chapter 6 'Debates' for arguments for and against money in FOSS, software licences, and universal basic incomes.

IDENTIFYING CONTRIBUTIONS OF RESEARCHERS

It is important to measure accurately the contributions of research institutions and researchers to code development as they contradict the dominant orthodoxy which pretends that innovation can only be generated by private 'start-ups.'

We are investigating this issue and expect to release findings in 2022.

'FREE RIDING' FIRMS

This report shows that though contributions made by non-firm email accounts to top GitHub repositories are less numerous than those made using firm email accounts, they are still significant. That FOSS is being co-produced by commercial firms and communal projects raises the question of the extent to which firms are benefiting from unpaid labour.

Nadia Eghbal concluded her *Roads and Bridges* report by asserting that 'fundamentally, digital infrastructure has a free rider problem' (2016: 106). But when it comes to free riding, or using a resource without contributing in return, some distinctions need to be made.

Big tech firms. GAFAM appropriate open source code, and their appropriation should be documented and publicised. Free riding by big tech is epitomised by the use that Amazon made of the Redis open source database (re-branded as Amazon Web Services Elasticache), without giving back to the community of developers (Moody, 2018). Yet, as we have documented in this report with our analysis of GitHub commits (see chapter 2), other big tech firms such as Microsoft and Google do contribute to some FOSS projects 'in ways that may not always be apparent from public sources, such as employing core project developers, making donations, and joining project steering committees in order to advance strategic interests' (Butler et al., 2019). Such firms should then properly be characterised as 'non-free riding appropriators.'

Big user firms. The near-ubiquity of FOSS in a growing range of end-user applications means most firms routinely deploy FOSS components in products and services. All firms today, not just big tech firms, are benefiting from the volunteer labour of FOSS project contributors who are not firm employees. But do firms who use digital infrastructure 'give back' by contributing? If they do not, they could be described as 'free-riding firms.'

A PRACTICAL METHOD TO IDENTIFY FREE-RIDING FIRMS

We selected firms from three sources (see Table 3.2, p. 26) and analysed to what extent they contributed to our sample of selected repositories and to Linux. As expected, most firms contribute very little, or not at all.

Table 6.1.
50 firms: Commits to GitHub repositories, to Linux, and mentions in IT media.

Firm	N. GitHub commits	N. Linux commits	Most commits to	Mentions in IT media	Core business [Source: Wikipedia]
Source: Firms with highest contributions to selected GitHub repositories, 2015-2019 (N: 20)					
Microsoft	438220	970	Microsoft/ vscode	22045	Develops, manufactures, licenses, supports, and sells computer software, consumer electronics, personal computers, and related services.
JetBrains	126485	0	JetBrains/ intellij-community	53	Software development company targeting developers and managers.
Google	72387	5927	tensorflow/ tensorflow	12302	Internet-related services and products, which include online advertising technologies, search engine, cloud computing, software, and hardware.
Red hat	54788	18090	openshift/ origin	3993	Providing open-source software products to the enterprise community.
Apple	37520	7	apple/swift	4602	Designs, develops, and sells consumer electronics, computer software, and online services.
Gradle	33934	0	gradle/gradle	18	Providing automation system service.
Intel	27613	22530	torvalds/linux	3144	Semiconductor chip manufacturer.
Facebook	23752	2186	facebook/ react-native	2370	Social media and social networking service company.
Pivotal	15984	0	spring-projects/ spring-boot	368	Cloud platform hosting, business transformation. Services, enterprise application design environments, containers, microservices, and consulting services.
GitHub	15324	0	atom/atom	2247	Hosting for software development version control using Git.
IBM	12543	7937	torvalds/linux	3800	Computer hardware, middleware and software, hosting and consulting services in areas ranging from mainframe computers to nanotechnology.
AMD	10610	10512	torvalds/linux	580	American multinational semiconductor company.
Huawei	7980	5121	torvalds/linux	849	Provides telecommunications equipment and sells consumer electronics, including smartphones.
Mellanox	7045	6993	torvalds/linux	34	Israeli-American multinational supplier of computer networking products.
Crytek	6349	0	CRYTEK/ CRYENGINE	0	German video game developer and software developer.
Elastic	5917	0	elastic/ elasticsearch	339	Creators of the Elastic (ELK) Stack.
Oracle	5851	5226	torvalds/linux	2769	Sells database software and technology, cloud engineered systems, and enterprise software products.
MariaDB	5206	0	MariaDB/ server	503	Enterprise open source database solutions.
Arm	5131	4928	torvalds/linux	1082	Semiconductor and software design company, design of ARM processors (CPUs), software development tools, infrastructure and software.
Databricks	4529	0	apache/ spark	399	Cloud platform for massive scale data engineering and collaborative data science.

Firm	N. GitHub commits	N. Linux commits	Most commits to	Mentions in IT media	Core business [Source: Wikipedia]
Source: Forbes ranking of top-100 technology firms by profit (N: 20)					
NVIDIA	3291	2260	torvalds/ linux	969	Designs graphics processing units for the gaming and professional markets, as well as system on a chip unit for the mobile computing and automotive market.
Broadcom	2169	2169	torvalds/ linux	152	American designer, developer, manufacturer and global supplier of a wide range of semiconductor and infrastructure software products.
Analog Devices	157	157	torvalds/ linux	0	American multinational semiconductor company specializing in data conversion, signal processing and power management technology.
Salesforce	124	0	kubernetes/ kubernetes	1062	American cloud-based software company.
Baidu	122	54	torvalds/ linux	123	Chinese multinational technology company specializing in Internet-related services and products and artificial intelligence.
Tencent Holdings	120	38	torvalds/ linux	62	Internet-related services and products, entertainment, artificial intelligence and technology.
Twitter	83	3	pytorch/ pytorch	1042	Provides microblogging and social networking service.
Adobe Systems	43	0	apache/ spark	670	creation of multimedia and creativity software products, with a more recent foray towards digital marketing software.
Micron Technology	24	24	torvalds/ linux	0	American producer of computer memory and computer data storage including dynamic random-access memory, flash memory, and USB flash drives.
Palo Alto Networks	13	0	ansible/ ansible	59	American multinational cybersecurity company.
Workday	8	0	apache/ spark	137	American on demand financial management and human capital management software vendor.
Splunk	4	0	moby/moby	92	American public multinational corporation produces software for searching, monitoring, and analyzing machine-generated big data via a Web-style interface.
Match Group	4	0	jlord/ patchwork	0	American Internet company that owns and operates several online dating web sites.
Arista Networks	3	3	torvalds/ linux	5	American computer networking company.
Tech Mahindra	1	0	home-assistant/ home-assistant	13	Indian multinational subsidiary of the Mahindra Group, providing information technology services and business process outsourcing.
ServiceNow	0	0	NA	100	US software company that develops a cloud computing platform to help companies manage digital workflows for enterprise operations.
Square	0	0	NA	153	American financial services, merchant services aggregator, and mobile payment company.
ON Semiconductor Corp.	0	0	NA	0	Semiconductor supplier company.
SK Hynix	0	0	NA	0	South Korean memory semiconductor supplier of dynamic random-access memory chips and flash memory chips.
Catcher Technology	0	0	NA	0	Manufactures aluminium and magnesium die casting products.

Firm	N. GitHub commits	N. Linux commits	Most commits to	Mentions in IT media	Core business [Source: Wikipedia]
Source: LinkedIn survey identifying firms which most attract interest from job applicants (N: 10)					
Alibaba	739	226	ant-design/ant-design	441	Chinese e-commerce, retail, Internet, and technology corporation.
SAP	127	0	DefinitelyTyped/ DefinitelyTyped	1766	Makes enterprise software to manage business operations and customer relations.
Orange	65	6	ansible/ansible	107	French telecommunications corporation.
Bell	41	41	torvalds/linux	84	Leading US communications company.
Zalando	36	0	kubernetes/ kubernetes	8	E-commerce company.
Capgemini	9	0	jhipster/ generator-jhipster	26	French multinational corporation that provides consulting, technology, professional, and outsourcing services.
Deloitte	2	0	DefinitelyTyped/ DefinitelyTyped	62	A multinational professional services network.
Softtek	2	0	jlord/patchwork	0	A global provider of IT services and business process solutions.
Banco Santander	0	0	NA	2	Spanish multinational commercial bank and financial services company.
Safra	0	0	NA	3	Banking and financial institutions and industrial operations.

REFERENCES

Butler, S. et al. (2019). On company contributions to community open source software projects. IEEE Transactions on Software Engineering. https://doi.org/10.1109/TSE.2019.29193052016: 106

Eghbal, N. (2016) Roads and Bridges: The Unseen Labor Behind our Digital Infrastructure. Ford Foundation: NYC.

Moody, G. (2018). Time for Net giants to pay fairly for the Open Source on which they depend. *Linux Journal*, 5 Nov

INVITED COMMENT:
PIERRE-YVES GOSSET

FRAMASOFT, A PLURALIST ALTERNATIVE TO BIG TECH

In recent years, Internet giants have entered into our everyday lives extremely quickly. Google, Apple, Facebook, Amazon, Microsoft (GAFAM) and a few others have colonised the Internet for some time now. Their privileged position allows them to collect data on a scale never seen before in the history of humanity through search engines, social networks, the management and storage of our personal photos or working documents. They also have access to our emails, our music, our videos, our Christmas shopping, our geographic locations and our professional, friendly or romantic relationships. To collect this amount of data, they have intruded into our computers, but also in our telephones, watches, printers, televisions, refrigerators and even our cars.

The technical domination of these companies is incredible. YouTube accounts for nearly 40% of the world's mobile Internet traffic. Every day, Facebook welcomes 1.8 billion people.[21] Alongside this technical domination, these five companies also exercise financial domination. In just a few years, they have become the world's five largest stock market capitalisations. At the end of 2020, the market capitalization of these companies was staggering: US$ 2,300 billion for Apple, 1,695 billion for Microsoft, 1,661 billion for Amazon, 1,185 billion for Alphabet/Google. More importantly, their equity is extremely high. They hold a fifth of the cash of all American companies. Apple alone has $192 billion in cash, representing more than Algeria's GDP. These sums allow them to 'enclose innovation' – they have the power to buy out almost any company on the planet.

But the greatest danger posed by these companies is, without a doubt, political and cultural. The five head offices are located in two North American cities. These companies are run by people with a specific vision of the world, a specific vision of morality, a specific understanding of what it means to be a human being. And this is not a good thing for diversity, for innovation, for the plurality of points of view. By realising their dream of a 'global digital village,' they standardise our consumption, they shape our relationships, they control our means of expression, they influence our vote and, more broadly, aim to organise our lives.

So, what can we do about it? Pending effective regulation of their behaviour, in which Europe plays its role as much as it can, we believe that it is possible to act and react effectively in an extremely simple way: by abandoning their products as much as possible, and instead use other ethical tools that are respectful of our personal data. This is the subject of the 'Degooglisons Internet'[22] [*De-Googlify the Internet*] campaign which we launched at the end of 2014. The objectives were ambitious, but simple: to raise public awareness of this technical, economic and cultural domination; and to implement software alternatives that are free, ethical, decentralised and solidarity-based. The result is a real success.

In three years, by installing, using, or developing free and open source software, Framasoft has implemented 34 alternative solutions to Google, Apple, or Facebook products. Today, about one million people use these tools every month. Free of charge. Without advertising. Without personal data exploitation. We have shown that a very small, highly motivated organisation with only ten employees can provide a daily service to hundreds of thousands of people. There are open source alternatives to services such as Google Doc, Google Calendar, Google Maps, WeTransfer, Twitter, DropBox, Doodle, and many others. We are now working to produce alternatives to YouTube[23] and Facebook Events.[24] But that is not enough.

That is not enough, because we firmly believe that building a single alternative actor would replicate the toxic GAFAM model. On the contrary, a network of several hundred tools and people would be much more resistant to Google or Facebook. That is why we initiated a year ago the CHATONS[25] (Collectif des Hébergeurs Alternatifs, Transparents, Ouverts Neutres et Solidaires) collective project, which can be translated as 'KITTENS are Keen Internet Talented Teams Engaged in Network Services.'

To date, this collective numbers about ninety entities, which can be non-profit associations, firms, cooperatives, individuals. These entities, based on the Framasoft model, offer online services respecting a charter. This charter obviously features technical obligations – for example, the entity must commit itself to setting up user data backups in case of problems. But above all it comprises ethical obligations, such as commitments to not use Google products, or to not use the personal data of users for commercial purposes.

Google, Apple, Facebook, Amazon and Microsoft are successful firms that are worth billions, host the best minds and produce a huge amount of innovation. But these companies are toxic. They lock us into patterns of actions, thoughts and relationships. They make us dependent. Our experience has shown that it is possible, with free software and very few resources, to no longer use the services of these companies. But free software, whose operational model is that of a common good, does not receive the necessary public attention. We strongly believe that it is possible to break out of data silos, out of the attention economy and of surveillance capitalism, by returning to a decentralised Internet built around reasonable and interoperable technologies that allow for chosen – not imposed – federation, such as the Solid[26] or ActivityPub[27] protocols.

This represents a clear political choice that is poorly understood by our governments, who prefer to support industrial models or classic start-ups, favouring the enrichment of the owners and shareholders of these companies. We doubt the capacity of states or of the European Union to question these public-private partnerships. However, if this were to happen, we believe that it is time for states to actively support commons initiatives, particularly those of civil society, through public-commons partnerships.

Moving from big tech to small tech, from a consumer society to a contributory society, will not happen overnight. But it is the only option available for states to regain a form of digital sovereignty, for users to emancipate themselves from addictive systems, and for society to find its way back to democratic debates, away from polarised conflicts.

Exponential growth is not sustainable, whether ecologically, economically, or psychologically. If we want a fairer, more caring world, we must politically affirm our rejection of the model of surveillance capitalism, and therefore of big tech. Their participation in open source has no other purpose than to optimise their toxic model. Open source is only a means for them to achieve their goal of monetising human behaviour. For us, it is a means to produce technology for the common good that is owned, controlled and funded by the commons. The essential challenge for us remains to provide not only resistance to the extraordinary firepower of the GAFAMs and their toxicity, but also to enable people to no longer suffer from the renunciations of the neoliberal state.

---

[20] https://www.statista.com/chart/17321/global-downstream-mobile-traffic-by-app/

[21] https://www.businessofapps.com/data/facebook-statistics/

[22] https://degooglisons-internet.org/en/

[23] https://joinpeertube.org

[24] https://joinmobilizon.org

[25] https://chatons.org/en

[26] https://en.wikipedia.org/wiki/Solid_(web_decentralization_project)

[27] https://en.wikipedia.org/wiki/ActivityPub

6

DEBATES AROUND THE RECOGNITION OF UNPAID VOLUNTEER LABOUR: UBI, MONEY, LICENCES[28]

- Free riding, predation and the fair sharing of the benefits of the production of digital commons occur while automation rises, employment is projected to decline and the environment is in crisis. There is a need to rethink the amount of time spent working in a paid job vs other types of useful activities which could give access to 'social drawing rights.' This chapter summarises three current debates.
- A Universal Basic Income (UBI) could enable people to have the freedom to engage in more creative work, but it could also have negative outcomes.
- Financially supporting non-intrinsically rewarding activities could help make FOSS projects more sustainable, but it also contradicts traditional understandings of computer development.
- Making licenses distinguish between users or activities and charging accordingly also contradicts traditional FOSS precepts. Yet this absence of 'moral' or 'political' distinction appears as one of the key obstacles preventing the development of a society based on contribution.

[28] Sébastien Broca also contributed to this chapter.

WHY RECOGNISE?

DECLINE IN EMPLOYMENT

Whether caused by automation, or by the slowdown of economic growth since the 1970s as industrial overcapacity spread around the world (Frey & Osborne, 2013; Benanav, 2019), employment opportunities have been decreasing, and are likely to continue decreasing. Inclusive and progressive remedies include the expansion of the cooperative sector, the reduction of working hours, and degrowth: the downscaling of over-production and over-consumption. Reducing energy consumption could also be aided by reducing the length of the working week, or by measures allowing people to work more slowly, and with less pressure (Mair et al., 2020).

NO SOCIAL RECOGNITION

The IT industry and the global digital economy's reliance on software partly produced by volunteers raises the question of the lack of societal recognition of these self-organised contributions. Should key innovation work such as this be more widely valued? Why is this work, which benefits so many organisations, not being compensated? What form would this compensation take? Beyond FOSS, what of other types of peer produced digital goods which benefit everyone, such as Wikipedia articles, OpenStreetMap maps, or Open Data datasets?

FUREAI KIPPU

In terms of state policy, contributions to non-rival common goods are not well recognised. A relevant example of a state recognising and valuing (rival) contributions to social care is Japan's Fureai Kippu or 'ticket for a caring relationship,' an alternative currency system where an hour of labour helping an elderly person is converted into a credit held in an online clearing house. This credit can then be drawn upon when needed, for example to pay for insurance premiums, or passed on to a relative (see Hayashi, 2012 for an overview of the scheme).

THE CONTRIBUTION QUESTION

How could the production of digital commons be similarly linked to social rights and social welfare? In 2014, a report on the 'Digital transformation of the French economy' produced by Philippe Lemoine called for the creation of an 'Individual Right to Contribution.' In France still, radical economists such as the Economistes Atterrés (Appalled Economists) and philosophers such as Bernard Stiegler have proposed variants of 'social drawing rights' and 'common labour rights' which would enable people who contribute to the commons to then earn points, or access to social services (Maurel, 2019).

CRISIS OF MEASURE

These proposals are being put forward because waged labour is not necessarily the best way to deal with the production of digital commons which involves thousands of contributions. The basis of wages is expropriation from the fruits of labour, but this labour needs to be measured before the expropriation occurs: with digital commons, some individual contributions are massive, whilst others are minuscule, yet all play a part. The symbolic or reputational rewards earned by participants to FOSS projects effectively remedy the failure of capital to measure this kind of labour. This crisis of measure can be summed up with a question: what is the impact of one line of code on the whole of Red Hat? Beyond Red Hat and private enterprise, this reminds us that there are ways for the state to support the production of digital commons and other volunteer work, through measures such as a universal basic income (UBI) for example.

DEBATE: UBI

A REDUCTION IN SOCIAL PROTECTION?

Reports of a new wave of automation have prompted approving parliamentary reports on a universal basic income in France and Australia; UBIs have been tried out in Canada, Finland, the Netherlands, and Scotland. The principle of a UBI is simple: everyone gets a certain amount of money, irrespective of their income level. There are concerns that a UBI would ultimately reduce the breadth of social protection. The proposal's embrace by some conservative politicians and high-profile techno-entrepreneurs could suggest that UBIs are a plot to remove social benefits for the most vulnerable, by replacing them with a single income (Alaluf & Zamora, 2016).

UBI AND GENDER INEQUALITY

UBIs might also have contradictory impacts on reproductive labour. On the one hand, they could operate as a feminist advance since 'having children markedly intensifies gender inequities in time allocation by increasing specialisation and women's workload' (Craig, 2006): a UBI would be particularly useful for single mothers, whose income is the most adversely impacted by childbirth. It would address a longstanding concern of feminist scholars such as Mariarosa Dalla Costa and Selma James (1972) who identified the vast amount of monetarily unacknowledged but economically essential household labour done for free. Without the invisible unpaid or reproductive process of caring, cooking, and so on, paid labour power would not be ready for work in the morning. But on the other hand, UBIs might encourage women to give up employment and return to traditional housework.

SHOULD SOCIAL SERVICES BE EXPANDED INSTEAD?

The question of whether a UBI is preferable to an expansion of free public services, which would lessen the need for money, is very much in debate: should provisioning be socialised (by free public services), or should demand be made solvent (by a UBI)? Spain's introduction of a UBI in 2020 will enable large-scale data to be collected and may generate some answers (O'Neil et al. 2021).

DEBATE: MONEY

A TABOO TOPIC

The issue of free labour is rarely discussed in FOSS projects. Part of the computer hacker philosophy is to view one's activity as a calling, not a profession. This could explain why Eghbal (2016) suggests that money is a 'taboo' topic in projects. Ethics of reciprocity, transparency and openness are meant to animate FOSS. These ethical values have traditionally been understood so that the defence of the 'four freedoms' (to use, copy, change, and redistribute modified versions of software) matters more than the fair distribution of profits stemming from software development (Broca, 2018). This normative stance was made very early in the history of free software.

THE STALLMAN DOCTRINE

In the GNU Manifesto, Free Software Foundation founder Richard Stallman contends that 'there is nothing wrong with wanting pay for work, or seeking to maximise one's income,' but only 'as long as one does not use means that are destructive.' Indeed 'extracting money from users of a program by restricting their use of it is destructive' (Stallman, 1985). Over the years, when asked to comment on the valuation of open source firms and the fact that they benefit from unpaid voluntary labour, Stallman has given the same answer: these issues are secondary. They are mainly 'a distraction from what really matters: that these programs are available for everyone to use in freedom and community' (Stallman, 2018).[29]

'CLUELESS' REFORMERS

In other words, the free software movement should consider software as resources upon which users have certain *rights*, not as products of a *labour* that deserves monetary retribution. Hence, free riding that is not based on enclosing code but on free labour is not a major concern (O'Neil & Broca, 2021). An illustration of this view's pervasiveness occurred when open source firm Redis attempted to introduce a new 'Commons Clause' licence to limit Amazon's free riding and was denounced as 'clueless' by prominent FOSS community members (Vaughan-Nichols, 2019).

A DEBATE IN DEBIAN

In contrast, others in the FOSS community have claimed that 'open source has a working-for-free problem,' with some launching firms with the explicit aim of providing independent open source maintainers with a reliable income (Pennington, 2019). The sustainability of volunteer labour is an important question. Our 2016 survey of work in the Debian project and subsequent interviews with Debian Developers (de Blanc et al., 2017; O'Neil et al., 2020) uncovered new work arrangements inside Debian. They contradicted the absolute freedom of traditional FOSS by introducing new organisational mechanisms between the commercial and communal worlds, such as work groups within the project and associated norms of professional behaviour. In parallel, a candidate for the 2019 Debian Project Leader election proposed in his election platform that Developers should be able to make Debian their careers if they chose to, thanks to increased firm involvement and grants (Michlmayr, 2019).

[29] Whilst Stallman's continued advocacy for the right to freely access and modify code to improve it has proved an inspiration for many, he sadly demonstrated in 2019 a lamentable insensitivity to sexism (see Musil, 2019).

DEBATE: LICENSES

SEPARATING ACTORS AND USES

The use of open source licenses risks enabling firms to delocalise contributor communities, and to prevent the generation of resources, much in the same way that countries with low environmental and social regulation drive labour costs down. This is because open source licenses refuse to translate their values into operational criteria that would enable the classification of actors either according to their nature, or according to their behaviour, and thus potentially restrict the authorised uses of the software (Broca, 2018). Such criteria would be in stark contradiction with the commitment to open access professed by Stallman and by most actors in the open source community. However, the absence of this 'moral' or 'political' distinction appears as one of the key obstacles preventing the development of a society based on contribution (O'Neil & Broca, 2021).

THE PEER PRODUCTION LICENCE OR 'COPYFARLEFT'

Kleiner (2007) argues that the General Public License ('copyleft') does not sufficiently address ownership. His Peer Production License ('copyfarleft') model distinguishes between commercial usages enacted by communal organisations where profits are equally distributed amongst workers, and those of capitalist enterprises based on the exploitation of wage labour. In contrast to noncommercial licenses, Copyfarleft attempts to favor communal organisations by allowing the cooperative economy to commercially exploit the commons, whilst the wage-labour based one cannot. Copyfarleft excludes entities from freely using non-rival goods, therefore going against the universal public good, so Saïd Vieira and De Filippi (2014) propose instead a commons-based licensing model that restricts commercial usage according to how much the user has contributed to the common pool. Their Commons Reciprocity License attributes commercial rights according to contribution, based on four criteria. This type of approval process raises issues such as the measurement of heterogeneous contributions to a common, the conversion of these contributions into different rights of use, and the control of the rate of exchange (Broca, 2018). The advantages of a general license are lost in favor of case-by-case decisions. Kleiner's Peer Production License risks treating massive transnational firms with limitless resources and small commercial organisations in the same way; but its ontological distinction (to be or not to be a cooperative, that is the question) has the merit of clarity.

TRADEMARK-BASED CONTRACTS

Another direction, proposed by Benjamin Jean of the Inno[3] consultancy firm, is to introduce trademark-based contracts for open source software maintainers. The motivation behind trademark-based contracts is to ensure a fairer sharing of value in favour of maintainers, while respecting the integrity of the framework set for free and open source software, by focusing on trademarks rather than copyright. This type of contract thus aims to be complementary to any free and open source license used on a program. It allows to control the way the trademark of the program is used and to request sharing part of the value generated by the commercialisation of services based on the exploitation of this trademark.[30]

[30] See https://inno3.frama.io/tm-contract-for-oss-maintainers/

REFERENCES

Alaluf, M. & Zamora, D. (2016). *Contre l'allocation universelle*. Montréal: Lux Éditeur.

Benanav, A. (2019). Automation and the future of work. *New Left Review*, 119.

Broca, S. (2018). Du modèle du logiciel libre au modèle productif des communs. Les licences pair à pair contre le free software? Working Paper 9, projet EnCommuns.

Craig, L. (2006). Children and the revolution. A time-diary analysis of the impact of motherhood on daily workload. *Journal of Sociology*, 42(2): 125-143.

de Blanc, M., O'Neil, M., Raissi, M. & Zacchiroli, S. (2017) Preliminary report on the influence of capital in an ethical project: Quantitative data from the 2016 Debian Survey, *Journal of Peer Production*, 10.

Dalla Costa, M. & James, S. (1972). *The Power of Women and the Subversion of the Community.* Bristol: Falling Wall Press.

debian-project (2019). Realizing Good Ideas with Debian Money. May-June. https://lists.debian.org/debian-project/2019/05/threads.html

Eghbal, N. (2016). Roads and bridges: The unseen labor behind our digital infrastructure. NYC: Ford Foundation.

Frey, C.B. & Osborne, M.A. (2013). The future of employment: How susceptible are jobs to computerisation? Working Paper. Oxford Martin School.

Hayashi, M. (2012). Japan's Fureai Kippu time-banking in elderly care: Origins, development, challenges and impact. *International Journal of Community Currency Research*, 16(A): 30-44

Mair, S., Druckman, A. Jackson, T. (2020) A tale of two utopias: Work in a post-growth world. *Ecological Economics*, 173.

Maurel, L. (2019). Common labour rights and right to work in the commons, trans. M. O'Neil & S. Collins. *Journal of Peer Production*, 13.

Michlmayr, M. (2019) Platform. https://www.debian.org/vote/2019/platforms/tbm

Musil, S. (2019) Computer scientist Richard Stallman resigns from MIT after Jeffrey Epstein comments. CNET.

O'Neil, M., Muselli, L., Raissi, M. & Zacchiroli, S. (2020) 'Open Source has won and lost the war': Justifying commercial-communal hybridisation in a FOSS project. *New Media & Society.* doi. org/10.1177/1461444820907022

O'Neil M. & Broca S. (2021) Peer production and social change. Ch. 21, *Handbook of Peer Production*. Hoboken, NJ: Wiley.

O'Neil M., Toupin S. & Pentzold C. (2021) What's next? Be Your Own Peer! Ch. 30, *Handbook of Peer Production*. Hoboken, NJ: Wiley.

Pennington, H. (2019). Open source has a working-for-free problem. https://blog.tidelift.com/open-source-has-a-working-for-free-problem

Saïd Vieira, M. & de Filippi, P. (2014). Between Copyleft and Copyfarleft: Advance reciprocity for the commons, *Journal of Peer Production*, 4.

Stallman, R.M. (1985). The GNU Manifesto. https://www.gnu.org/gnu/manifesto.en.html

Stallman, R.M. (2018). Private conversation with Rob Lucas. Posted on the Nettime mailing list by permission. https://nettime.org/Lists-Archives/nettime-l-1811/msg00049.html

Vaughan-Nichols, S. (2019). The battle between real open source vs. faux open source heats up. ZDNet, February 20. https://www.zdnet.com/article/the-battle-between-real-open-source-vs-faux-open-source-heats-up/.

INVITED COMMENT:

CÉLYA GRUSON-DANIEL, BENJAMIN JEAN AND CAMILLE MOULIN (INNO³)

FEDERATED RESPONSES TO BIG TECH MONOPOLIES

This report documents the manner in which powerful firms are co-opting open source software. Taking Microsoft as an example: not only did it purchase the GitHub collaborative development platform, it also acquired other strategic assets such as LinkedIn, and now has a strong grip on key levers of the digital economy. Microsoft now occupies a hegemonic position as an indispensable interconnector and mediator, raising privacy issues in terms of the management of individual data which rival those posed by Facebook and Google.

In response, the principles of federation offer the possibility of bypassing strategic conflicts between platforms, which always conjure the risk of recreating a 'Google 2.0' situation (the capture of value by monopolies). Federation enables users to be independent from a platform to communicate with other users through interoperable mechanisms (protocols, APIs, etc.). The principle is similar to email: anyone can send emails to anyone without having to register with an application. Imagine if it was absolutely necessary to have a Google mailbox to send emails to a Gmail address![31] Examples of federation are Framasoft projects such as PeerTube (YouTube-like) or Mobilizon (Facebook page-like), where each self-hosted instance (or server) can indicate the other instances to which they want to be connected by subscribing to them. Service federations are in flux and users receive updates from other instances as time goes by, without being tied to a single entity. The central principle of federated models is to observe a similar protocol (just like email, or the blockchain). This protocol can itself evolve just as much as the interface and software that connect to it.

In addition to these technical and structural elements, these alternatives cannot work without questioning, and also intervening on, logics of use. Individuals are meant to recover sovereignty over their data. This certainly involves interoperability mechanisms, the creation of data recovery and reversibility functionalities, but the levers of action are also social and communicational.

Framasoft rose to this challenge. This association proposes a set of open source tools (e.g. pad, gitlab, questionnaire, videoconference, etc.) on their servers. Their initial campaign 'De-google-ify the Internet' evolved after seeing that many users were seduced by this 'non-Google' solution, but without understanding the underlying mechanisms. Hence their recent communication advocates 'Deframatising' the Internet and developing archipelagos and CHATONS[32] (collective of independent, transparent, open, neutral and ethical hosts) so that local members inhabit the territory by offering to provide hosting. Open source solutions provide applications (such as the jitsi videoconferencing tool or the gitlab code repository) and archipelagos of hosts provide the material means to run these applications and make them available to local users.

The parameters to be taken into consideration for the adoption of these models thus stretch beyond the strictly technological dimension; they also include considerations relating to the maturity of the end user. For example, it is not always straightforward for someone using a federated solution to understand that shared information can also be displayed on other instances. This is what occurs, for example, when using the Mobilizon event sharing platform. Publishing an event via one Mobilizon instance will allow other instances to broadly disseminate this information. It is necessary to assist users to understand these mechanisms and to shift from a vision of data security as controlling and securing your data (security = closed + control of detention). Brand and trust markers are key for these networks to be appropriated, by explaining that open solutions allow people to have control over their data while facilitating their circulation in an ecosystem of trusted others (security = openness + control of circulation).

In this sense, truly 'open' models today require thinking in terms of trust ecosystem aimed at re-empowering users.

[31] Nevertheless, this is what is happening more and more when using instant messengers, which require other users to use the application.

DATA COLLECTION

GITHUB DATASET

GitHub repositories were selected based on their size and popularity (Table 2.1). Data including email addresses of committers, user IDs, dates committing, and source lines of code of commits were collected between 1 January 2015 and 31 May 2019 with the GitHub REST API v3 using R packages such as 'httr,' 'httpuv' and 'jsonlite.'

IT MEDIA DATASET

Articles published between 1 January 2015 and 30 April 2019 in ComputerWorld, Slashdot, and ZDNet featuring at least one mention of one of 50 FOSS projects (Table 3.1) were collected with the R 'rvest' package. Table 3.4 summarises how we reduced our dataset from 13,174 unique articles to 1,424 articles featuring co-locations of firms and projects, and finally to 86 articles featuring at least five labour-related terms.

PRESENTATION SUMMARIES DATASET

The 'rvest' R package was used to scrape data from the FOSDEM19 and OSS19 open source conference websites. We collected presentation titles and abstracts, session names, presenter names and biographies, including professional affiliations.

EMAIL LIST DATASET

We used the 'rvest' R package to collect the content and authors of messages posted between January 1, 2015 and April 30, 2019 to the Linux and Firefox email discussion lists. Table 4.6 details the sources and number of messages.

ETHNOGRAPHIC DATASET

We selected keynotes, talks and workshops according to our main interest, the participation of firms in open source projects and the firm-volunteer community relationship. Notes were collected according to a predetermined template. On-site in Lyon and Paris, in addition to recording presentations, we counted the number of attendees; we evaluated the size of rooms; and we recorded significant interactions between speakers and audiences, such as interjections. We spent time with developers in projects booths, and a team member gave a presentation in Paris, enabling us to collect professional insights about firms and infrastructure during the Q&A. Room capacity was a good indication of which streams or speakers were deemed important by organisers whilst speaker/audience interactions were of use in identifying overt and latent conflicts. Our original research design was that teams of three researchers would attend OSS19, POSS19 and FOSDEM20 in person, recording presentations and taking field notes. Our field work was adversely affected by the national strikes against pension reform that paralysed France between December 2019 and February 2020. We were unable to attend FOSDEM2020 (held 1-2 February 2020) due to the unsettled transport situation. Instead we collected and analysed presentations from the FOSDEM2019 presentations video archive. FOSDEM is traditionally designed to be followed both remotely and asynchronously, so all 740 speeches and talks from the 2019 edition were downloadable from the conference website. Table 4.1 summarises the collected ethnographic data.

SENTIMENT ANALYSIS

Sentiment analysis was conducted with the 'tidytext' R package on the semantic environment (words preceding and succeeding a term) of firm terms in IT media articles. The 'bing' sentiment lexicon which classifies 6,786 English words into binary categories (negative/positive) was used to calculate the densities of negative words of firms. The density of negative words is the frequency of negative words in a firm semantic environment divided by the frequency of the firm's name.

REINERT METHOD

We used the IramuteQ application, based on the Alceste software to map semantic clusters in conference presentations. This software performs a primary detailed analysis of a text corpus' vocabulary, constituting a dictionary of word frequencies. By successive splits, it then divides the text into homogeneous segments and classifies these segments by locating the strongest oppositions. This method enables the extraction of lexicons and sub-lexicons, made of the statistically most significant words and sentences. These lexicons represent the dominant ideas and themes of the corpus.

RESEARCH ETHICS

All the stages of this project were reviewed and approved by the University of Canberra's Human Research Ethics Committee (Project No: 1865).

RECOMMENDATIONS

GOVERNMENTS SHOULD GOVERN

(See Broca, p. 17; Le Crosnier, p. 30)

To the regulators: please regulate Big Tech. Taxation is a start. Curtailing of surveillance must come next. Beyond that, it is the hegemonic position of Big Tech in society at large that must be questioned. Public authorities and regulatory bodies must cease to consider FOSS and digital commons only as the means to save money, by integrating them into public tenders for example.

PUBLIC AWARENESS OF APPROPRIATION AND FREE RIDING

(See chapter 5, p.52)

Large technology firms seek to obscure knowledge about their predatory behaviour by referring to a unified 'community' (see chapter 4). Precisely documenting to what extent they are appropriating common resources may generate wider social understanding of how much the digital economy depends on open source. Every firm, NGO and government agency is now reliant on digital infrastructure, but not many support it. We provide a method for identifying 'free-riding' firms.

VIABILITY OF FEDERATED ALTERNATIVES

(See Gosset, p.58; Gruson et al., p.68)

This report features invited comments by members of Framasoft and Inno3 who advocate for the use of federated online services as alternatives to big tech. Evaluating and publicising to what extent these solutions are viable and cost-effective is an important step towards their wider social dissemination.

ROLE OF RESEARCHERS, DEBATE IN FOSS

(See chapter 5, p. 52)

The prevailing rhetoric is that FOSS sustainability requires for-profit solutions: this hides the contributions of researchers to FOSS development, which should also be publicised. A debate about GAFAM appropriation, the recognition of volunteer labour, and other issues should be initiated and led by the FOSS community.

BRIDGES BETWEEN THE DIGITAL COMMONS SECTOR AND INSTITUTIONS

(See chapter 6, p.60)

IT sector proposals (firm support, bounties, paid licenses) do not recognise that digital infrastructure is a public good which is linked to broader social issues and requires broader solutions. Despite cultural differences and possible conflicts, the rise of automation and predicted job losses require creating more connections between the volunteer sector and state institutions. Proposals such as social drawing rights and UBIs, and mechanisms such as Fureai Kippu show the way.

DCPC21