U.PORTO
FLUP FACULDADE DE LETRAS
UNIVERSIDADE DO PORTO

MESTRADO EM SOCIOLOGIA

The open technologies in the global era – a Sociology of Open Access

João Pedro Bernardes P. F. Dias

M

2018

João Pedro Bernardes P. F. Dias

The open technologies in the global era – a Sociology of Open Access

Dissertação realizada no âmbito do Mestrado em Sociologia da FLUP orientada pela Professora Doutora Alexandra Lopes e coorientada pelo Professor Doutor Luke Martell

Faculdade de Letras da Universidade do Porto
Setembro de 2018

The open technologies in the global era – a Sociology of Open Access

João Pedro Bernardes P. F. Dias

Dissertação realizada no âmbito do Mestrado em Sociologia, orientada Professora Doutora Alexandra Lopes e coorientada pelo Professor Doutor Luke Martell

Membros do Júri

Professor Doutor Natália Maria Azevedo Casqueira
Faculdade de Letras – Universidade do Porto

Professor Doutora Maria Elisa Ramos Morais Cerveira.
Faculdade de Letras – Universidade do Porto

Professor Doutora Alexandra Cristina Ramos da Silva Lopes Gunes.
Faculdade de Letras – Universidade do Porto

Classificação obtida: 17 valores

Table of Contents

Declaration

I declare that the present dissertation to the best of my knowledge, is original and has not been submitted in whole or part for a course or a degree in this or any university. References to other authors (statements, ideas, thoughts) follow scrupulously attribution rules and are appropriately indicated throughout the text and references, according to citation norms. I am conscious that plagiarism and auto-plagiarism constitutes an unlawful academic practice.

Porto, November 6, 2018
João Pedro Bernardes P. F. Dias

Acknowledgements

I would first like to thank my thesis supervisors Prof. Luke Martell (University of Sussex) and Prof. Alexandra Lopes (University of Porto), their supervision and advice broadening the horizons of my research. Professor Alexandra provided me with the technical tools and always deposited great confidence in my work, her judgement helped me to keep critical and focused. Professor Luke was always positive, enlightening and guided me through most of the processes and obstacles and I cannot thank him enough for agreeing to work with me in the first place, it was a great pleasure knowing him and working with him and expanding my experience with such a notorious professional. To both I thank the knowledge, time and professionalism.

I would also like to thank the experts that kindly responded to my questions and interviews: Prof. Ginsparg (Cornell University), Prof. Berry (University of Sussex), Prof. Eve (Birkbeck, University of London), Prof. Marsden (University of Sussex), Prof. Townend (University of Sussex), and Dr. Webb (University of Sussex). Without their participation and input, this analysis and investigation would be lacking in awareness and touch with reality – I cannot thank them enough for the values they added to my work.

I would also like to acknowledge Dr. Nuno Cardoso (University of Porto) – and his wife Dr. Lígia Massena (University of Porto) – for the friendship, patience and selfless advise. Without his steadfast wisdom and amazing powers the road would not be so clear.

Finally, I must express my very profound gratitude to my parents, to my sister, to my partner and son for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Resumo

Palavras-chave: Open Access; Sci-Hub; Publicações Académicas; Globalização; Sociologia;

A sociedade contemporânea é cada vez mais influenciada pelas tecnologias da comunicação. Informação é hoje tanto a matéria prima quanto o produto da produção industrial. No contexto de produção e inovação, um série de movimentos abertos emergem e ganham importância na cena digital. A esfera das publicações académicas não é exceção. Enquanto o mundo da produção de conhecimentos e da sua publicação se encontra atado a um ciclo vicioso de mecanismos simbólicos e concentração de poder, o movimento Open Access está a tornar-se no cânon das publicações digitais, prometendo novas ferramentas legais, regulamentação, e melhorias gerais no aproveitamento e performance da investigação académica e da sua disseminação.

Esta investigação oferece uma introdução crítica ao Open Access e aborda-o duma perspectiva global: questionando se o Open Access é capaz de oferecer uma plataforma livre, aberta e diversificada de produção e disseminação de conhecimentos. Fá-lo através da interpretação da análise dos dados estatísticos do uso do serviço online Sci-Hub durante o ano de 2017 e aplica a mesma a um conceito mais geral de consumo de artigos ao nível global. Os resultados revelam diferentes padrões de utilização em diferentes regiões, sobretudo relacionadas com a existência de diferentes conjeturas estruturais que influenciam o equilíbrio entre oportunidade e recompensa.

Abstract

Keywords: Open Access; Sci-Hub; Scholarly Publishing; Globalization; Sociology;

Contemporary societies are ever more influenced by the communication technologies. Information is now both the raw material and the product of industrial production. In the context of production and innovation there are several open movements emerging and gaining importance in the scene. The world of scholarly publications is not exception to this. While the world of knowledge production and publication is tied in a vicious loop of symbolical mechanisms and concentration of power, the Open Access movement is becoming academia's publication standard for digital publications, promising new regulatory tools that will help improve research and its dissemination.

This investigation offers a critic introduction to Open Access and approaches it from a global perspective, questioning whether Open Access is being able to deliver an open and diverse platform for knowledge dissemination. It does so by interpreting the usage statistics of the Sci-Hub shadow service for the year of 2017 and applying them to a broader notion of worldwide article consumption. The results reveal different appropriations and behaviours towards Open Access, due to existence of different structural contexts and logics influencing opportunities and payoffs.

List of Figures

List of Tables

Abbreviations

APC Article Processing Charges

ASC American Society of Chemistry

ASN Academic Social Networks

BBB Budapest Bethesda and Berlin Declarations

BOAI Budapest Open Access Initiative

BPC Book Processing Charges

BRICS Brazil Russia India China and South Africa

BSD Berkeley Software Distribution

CC Creative Commons

DOAJ Directory of Open Access Journals

DOI Digital Object Identifier

EC European Commission

ERA European Research Era

ERC European Research Council

EU European Union

FAIR Findable Accessible Interoperable and Reusable

FDI Foreign Direct Investment

FOSS Free and Open Source Software

FP7 EU's Seventh Framework Programme for Research and Technological Development

GDP Gross Domestic Product

GNOME GNU Network Object Model Environment

GNU GNU is Not Unix

GPL GNU Public License

GVC Global Value Chain

HASS Humanities Arts and the Social Sciences

ICT Information and Communication Technology

IP Intellectual Property

IPR Intellectual Property Rights

ISBN International Standard Book Number

JIF Journal Impact Factor

LA Latin America

MIT Massachusetts Institute of Technology

NGO Non-Governmental Organization

OA Open Access

OAI Open Access Initiative

OAI-PMH Open Archives Initiative Protocol for Metadata Harvesting

OECD Organisation for Economic Co-operation and Development

OLH Open Library of Humanities

OUP Oxford University Press

PDF Portable Document Format

PIPA Protect IP Act

Plos Public Library of Science

RCUK Research Councils UK

R&D Research & Development

REF Research Excellence Framework

SCI-ISI Science Citation Index produced by the Institute for Scientific Information

SciELO Scientific Electronic Library Online

SOPA Stop Online Piracy Act

STEM Science Technology Engineering and Medicine

TA Toll Access

UN United Nations

UNESCO The United Nations Educational Scientific and Cultural Organization

US University of Sussex

VPN Virtual Private Network

WoS Web of Science

Introduction

This thesis addresses the emergence of open movements in contemporary societies and seeks to provide sociological levels of knowledge on these recent phenomena. Open movements, such as the notoriously successful Open Sorce movement, are movements born and nested in the Internet and are helping to re-shape the landscape of industrial production through innovation – reshuffling inequalities and opportunities in a highly unequal global world.

This research is built around the specific case of the Open Access (OA) movement and seeks to provide answer into whether it can or not provide or even enhance an alternative channel for 'the commons' (Berry, 2008) to produce and share informational products.

This work is becoming more relevant by the day because while it witnesses the early stages of a sphere of social construction (namely internet regulation) it offers the possiblity to give back to the community by expanding knowledge, information and discourses. It is releveant to Sociology as a discipline because it brings the critic of public discussion into the realm of the social sciences. It is also pertinent in the context of academia and of the sciences because it readresses scientific analysis from a more sceptical perspective on a matter that is mostly self-reflective of scientific practices.

This thesis involved the analysis of the Sci-Hub's statistics log for the year of 2017, a log of downloads of scholarly articles through a major online shadow libray. The data is interrelated with other relevant data extracted from several data handlers, most notoriously, the World Bank and the Organisation for Economic Co-operation and Development (OECD).

The thesis is organized in the follwoing manners, Chapter 1 – *The Chessboard of Open Access* – gives a historical contextualization and the state of the art review of the theme of the dissertation, providing the reader with the needful set of tools to follow the study's main arguments. Since the OA movement is very much nested inside academia (the main sector of knowledge production in a global industrial production line), the chapter also provides some insight on the mechanisms of scholarly publishing, namely those that interact the most with the tensions between the new (based on OA) and the traditional models (based on subscription models, or, pay to read) of article publication.

Chapter 2 clearly defines and limits the research question and in a first section provides the method design for the research, made up of concepts that are already present in social theory as well as some concepts that we propose and develop throughout the research. A second section is dedicated to the clarification of the methodological steps that were taken throughout the analysis,

as well as the caveats and drawbacks related with those options.

Third chapter lays down the results of the analysis and is followed by a conclsuion section dedicated to explore and relate the findings with our research question. This section will also provide insight into future work.

This dissertation explores the developments in digital media and the Internet, namely and specifically OA, by bringing to light key aspects of contemporary societies such as the transition into a digital and economic, scientific and social globalization. Although economic factors are important to understand globalization, the diaspora of billions of people into the Internet, caused deep change in social behaviour that cannot be ignored. Distance was bended and control and coordination was enabled and unprecedented forms of controlling the industry were created. That is why contemporary societies are partly shaped by the idea of globalization and more so every day. Either a single society, or a diverse singularity, globalization is often envisioned as one or the other, either giving place to a homogeneous society, through the spread of unrivalled rule of capitalism, or differently by an ecosystem of *mélange* where technologies play an important role flatting out inequality of opportunities among different players. These two hypothesis of globalization shall be studied throughout this dissertation in close relation with the digital media and internet developments, namely the OA transition.

The Information and Communication Technology (ICT) revolution affected everyday life for a large amount of people that suddenly became more interconnected. It was no longer an issue of military and academic, the internet started entering households by the turn of the millennium. The diaspora of millions of people into the internet caused deep aberrations in social behaviour, state power, trade, and above all else, industry, one of the early offspring of technological development. The global value chain is the term used to described the new logics and intuition of the industry regarding the division of labour and social stratification assuming an overcoming of national border's obstacles, and characterized by horizontal or vertically integration of enterprises in a big common bazaar. This gave rise to recent studies of global inequalities – these studies highlight the fact that contemporary social order is better explained through an international scope, i.e. by comparing inequalities among countries. Also, the new big digital bazaar that emerged from this appropriation of the technology by the industry, gave rise to new opportunities, solutions, ideas, etc.

One of the major ideological restructuring of globalization besides race and gender, are (Intellectual Property Rights (IPR)) related issues. Mainly copyright, Not only did copyright as it was conceived since the XIXth century endured grave pressure to the level of collapse in industries such as the music industry, also the internet and the digital media are now facing a new wave of regulation after internet regulation started to collide with sociopolitical behaviour, such as the Facebook case. Although change happens fast, there is the request for it to be even faster, sometimes it is hard to take a step back and monitor our own behaviour since it is more rewarding pushing on the wave of technological innovation.

OA to scientific literature became an issue around the turn of the millennium when the digital technologies started being used globally by the academic community. Scholarly documents could be passed on through digital media and there was a whole set of new rules to apply to the traditional publishing system, based on printed versions of journals. The digital media played a revolutionary part in scholarly publishing, since it enabled new forms of enhanced sharing, access and visibility for authors. OA is an important issue today, because there is clearly the need to reinvent the way we publish our work especially in the fast paced sphere of scholarly publishing. It was not only the music industry and the film industry who suffered from the rise of internet sharing,*sharism*, also the publishing industry, the news media and the broader book publishing industries are being forced to shift into the realm of the digital at an accelerated pace. OA plays a huge role in this scenario seen that it provides a new modern model for publishing in the scholarly sphere leaning on the technological capabilities made possible by an intense innovative industry of software and policy development. OA is even promising outside academic publishing, and vast infrastructures were developed to OA to older books, and also newly published books. As a business friendly and economic sustainable model, it is still taking its first steps, In the case of academic books, such as monographs this raises the problem of how to implement OA in monographs related discipline.

While copyright was created with the goal of managing book publication and such, protecting publishers and authors in a way that would grant the possibility to continue publishing new material, it became deprecated with the rise of the digital. Since then new forms of copyright have been needed to deal with a publishing environment of non-scarcity. The fact that publishing does not require the same material costs with printing that traditional publishing used to require, makes it that copyright is now acting against its putative premise - enabling publishing. Old logics become obsolete and new logics emerge, and the need for policy revue and a broader sense of change is at hand.

Chapter 1

The chessboard of open access

1.1 Historical Contextualization of Open Access

It has been almost 30 years since the first signs of an open, digital movement started to appear in the sphere of academia. After the dawn of the digital and the rise of a new network society, industry, at a global scale would engage in acute change. Its genesis would take place in the regions of the planet where the services industry would be more developed and integrated with corporate activity. Places such as countries' capitals would nest technological innovation and apply it to the global chain. The result was an integrated circuit of manufacturing, where different countries would specialize in different parts of the production process. The only logical outcome of the new world economic order is reflected in many aspects of human activity today, from financial activity to technological innovation and knowledge production. While the economy of the old west (from now on global North) shifted into services and finance (Sassen, 1991), different parts of the globe took new functions in the process of production. While China became the factory of the world, London became the centre of world finance. Meanwhile, corporate power grows embedded in a net that bends traditional rules such as time and space, reshaping productivity, market, rule and politics.

Evolving from a tradition with more than 350 years (Larivière, Haustein, & Mongeon, 2015), scholarly publishing was deeply struck by the ICT revolution and by the globalized economy – printing would no longer be done on paper but simply on Portable Document Format (PDF). This dissertation takes on the analysis of OA from a sociological perspective, strongly leaning on the political-economic contexts and motivations of the global industry that shape local activities into delegation and specialization, affecting the people, work, sociability, culture, geography, environment and peace. While all the world is affected by it, it is so differently from place to place, time to time, and the prospects for a cosmopolitan, post-racial, post-gender, post-imperialist and equal society, a society united rather than divided by borders, seems to become a recent-past or a distant future. The role of technologies in the process of the globalized economy is central for its understanding, for technology affects supply and demand, creating new spaces for profit and

growth. It creates new opportunities, profit, autonomy, freedom, mobility, etc depending on the logics of each party – each draw their cards and try to maximize their outcome perspectives.

Industry changed rapidly since the Second World War. In 1947 the transistor was created at the Bell Laboratories, a simple mechanism of turning on and off an amplified electrical signal – the principle of electronics. By the year of 1971 , a microprocessor was capable of processing the equal amount of 2300 transistors, by 1991 the value would amount to almost 5 million. The exponential growth of processing power (see Moore's Law) meant that a powerful tool was being developed and that new utilities had to be written for it: namely in the form of software.

In 1980, the World Wide Web was born, and with it a whole new way of developing the very technology that fuelled it. Computers became cheaper and many universities, mostly in Europe and the United States started hosting servers, terminals and a connection. As computers became cheaper, so broadband, and the speed rose. By this time the financial market was deeply embedded in a networking system, the Latin America faced a severe financial debt crisis after a decade of oil prices rising, the US had abandoned the gold standard and stock-market became a widely popular way for companies to seek capitalization. The economic order would change during this decade. The ICT were enablers. Great conglomerates of capital and huge corporate power were created, this time, disregarding state and frontiers, but nesting in cities and specific regions of the planet (Sassen, 1991). Stock market, mergers and acquisitions and debt seem to be the great boosters of the world economy near the end of the century.

By the end of the XXth century several markets were already beginning to feel the effects of the digital age undermining traditional manufacturing and production. Both the music and the film industry had to adapt their activities to the new set of logics that succeeded the rise of the digital and of the Internet in order to survive the upcoming of the digital revolution. On the other hand, since 1970s entrance of China in the global market, manufacturing was quickly abandoning regions on the North. The press and publishing were only later affected by the ICTs and achieved a more subtle transition. In a way OA is a direct consequence of this adaptation process. On one side, publishing costs are going down, while subscription prices in academia tend to go up quicker than inflation. The key aspect to retain seems to be that the revolution resides in the fact that publishing is slowly becoming exclusively digital, while abandoning printed formats. So the substance of the publishing industry is changing from printed versions to digital copies. This means that the publishing industry is going through a period of adaptation and there will be possibly a reshuffling of capital and power among stakeholders. OA it seems, will be the new standard for scholarly publications, and thus whoever achieves a most complete form of appropriation (research community, universities, funding agencies, libraries or publishers) will likely be in control of scholarly communications.

Prior to the Second World War most scholarly journals were still published by scientific societies Larivière et al. (2015), ie, mostly not-for-profit societies. By 1960 the world of scientific publishing was being populated, if not steadily invaded, by commercial publishers, accompanying a broader trend of fierce corporate activity and world economic reshaping that characterized

the decades to come. The number of mergers and acquisitions grew gradually during this period Larivière et al. (2015), Munroe (2007), a trend that could be traced back to several areas of industry. By 1980, librarians and universities were facing alarming signs of predatory behaviour among publishers of scientific journals. According to Munroe (2000) cited by Lipscomb (2001), between the years of 1998 and 1999 alone, 60 mergers and acquisitions had happen among publishing house. As diversity faded, so prices went up and what some identified as oligopoly was shaping scholarly publishing.

By 1990, the Internet was starting to become a common feature in some universities and researchers would be able to access it on a more or less stable fashion. In 1991, in Los Alamos National Laboratory, the arXiv was put online[1], and it quickly became a repository where articles would linger indefinitely, accessible to any researcher provided with Internet connection. The arXiv became thus the first successful article repository in the history of OA, and since its inception it reduced some inequalities of access among researchers in the area of high energy physics, and at the same time giving an opportunity for distant researchers from other countries to have access to that same literature.

The change of priorities was acute. For one, research material was getting out of traditional controlled channels, and on the other hand, research was no longer being printed, but replicated digitally. The logics sustaining the arXiv into existence soon became a worldwide phenomenon, or an Internet phenomenon, again, disregarding national borders or bringing them down, yet centralized in poles of technological development. It was only until the leap of the century, in 2001, that librarians united in Budapest in what would be the historically accepted event that marked the beginning of the OA movement, or initiative, the Budapest Open Access Initiative (BOAI), which stated the logics of OA and the way to achieve it. It was no longer a matter of technological fitness, but of moral and ethical imperatives (Swartz, 2008). From tradition to new technology and ultimately to 'unprecedented public good', the BOAI was complemented by the Bethesda Statement on OA Publishing (2003), and later on by the Berlin declaration on OA to Knowledge in the Sciences and Humanities (2003), what later, Suber (2012) would call the Budapest Bethesda and Berlin Declarations (BBB).

On the overall, digital publishing certainly gave strength to corporate power, since the dawn of the digital does historically overlap with the formation of bigger multinational corporations. Although correlation does not signify causation, and many other reasons could be pushing this market into a global unity (the '5 million dollar journal' as caricatured by Munroe (2007)), it has been suggested by Sassen (1991, 2014) that a commodity such as information tends to agglomerate and that ownership tends to be in the hands of an increasingly restricted elite. For academia this poses a problem of leveraging the dependence of research on economic factors such as commercial popularity or viability of the subject of the study, rising prices of journal subscription, or a more general access to research. Nevertheless, more than giving priority to corporate power, the digital technologies seem to be reshuffling the economics of scholarly

[1] by then called www.lanl.gov (Ginsparg, 2011a)

publishing and while from one side oligopoly and corporate control is growing, new methods, infrastructures, joints and practices seem to be developing on top of the digital technologies, namely, OA is growing exponentially. Its consequences on the world economy are the core of our analysis throughout this dissertation, we will assess the current situation of scholarly publishing and try to draw some trends for the future, but first, we are going to introduce OA in its major characteristics and its major players.

1.2 The metrics of Scholarly publishing

The metrics of scholarly publishing can be considered as a part of what some authors call the economics of OA (Eve, 2014; Suber, 2012). In this study we give a larger emphasis to the macroeconomics of OA rather than to the microeconomics of it. This section will be dedicated to a revision of the literature on the latter.

As we have seen, scholarly publishing was invaded by the digital revolution, especially in the form of digital publishing but also in the form of more efficient ways to deal with large amounts of data at a broader level. OA was developed in a sense that would benefit the academic community by applying existing technology to the benefit of knowledge and society at large. But, while this seems clear within advocate discourse, in practice, the shift from subscription based, to OA publishing is proving to be less straightforward than it would seem likely at first. So what are exactly the logics of scholarly publishing that meddle with OA's purely logical development?

Scholarly publishing is more than merely a channel of communication among academics. Paper publication is also widely used as a measure to differentiate academics among their peers and to assess their work, in fact many tools have been developed in order to measure the quality of articles without having to actually read them. This is vital for employers to select among candidates, and for employees to keep their grants – in a post-doc labour market, this is particularly relevant. In an increasingly precarious labour market, where researchers are constantly in need to renew short contracts, this mechanism is revisited yet more often every day. The quality of the work published by a researcher is, nevertheless, evaluated according to the prestige of the venue or journal he publishes in, meaning that employers will look mainly into the *quality* or *prestige* (Eve, 2014) of the journal to assess on the researcher's work, rather than by reading his or her work. While this may sound strange, the explanation lies somewhat in the fact that the quantity of articles published yearly makes it impossible for all of them to be individually assessed. This means that the publishing venue's prestige acts as a surrogate and a proxy to the quality of the general work that is found within its pages. This idea was introduced clearly by Eve (2014) when analysing OA's economics. While prestige (in all its bespoken and symbolical contours) is certainly an insightful way to put it, academia relies increasingly on the rigid routine of metrics to measure this institutional lather.

Before the year 2000, there was the Science Citation Index produced by the Institute for Sci-

entific Information (SCI-ISI) (circa 1960 (Garfield, 2006)). In 2002, Thomson Reuters launched the Web of Science (WoS), still widely used today, while other competing indexes were created in the upcoming years, such as Elsevier's Scopus and Google Scholar (Hicks, Wouters, Waltman, De Rijcke, & Rafols, 2015). Many of these are not openly available to the public (with maybe the exception of Google Scholar) hardly compelling with OA ideals. Annex to these indexes, bibliometrics were created, a form of hierarchization, a proxy to the quality of a researcher's performance and quality of articles. In 2005 the *h*-index, a citation scoring system, was proposed. The Journal Impact Factor (JIF) a relationship between the citations a journals recieves in a year and the total amount of publications was introduce and published regularly since 1975 in the Journal Citations Reports. As the interest in these areas gain manpower (scientometrics, journalology, see (Cooper, 2015; Garfield, 2006)) altmetrics also start to develop. On a rather sceptical approach, Hicks et al. (2015) describes the evaluation of academic's work as a shifting from judgement into a data led analysis. As Hicks et al. (2015) continues to explain: 'Metrics have proliferated: usually well intentioned, not always well informed, often ill applied' – a concern often detected in other areas of data analysis and statistics. This explains the emergence of altmetrics – attempts to measure researchers through alternative means other than citation number, such as the amount an articles is downloaded, or references in social and academic social media. As Eve (2014) points out though, these are often easily 'gamed' in the sense that fake social media accounts can easily be created, or that torrents of downloads can be requested by bots instead of actual readers.

Assessment methods end up influencing researchers publishing behaviour. On the other hand, mandatory OA by forcing authors into publishing OA versions of their finalized work or by publishing in OA journals, forces publishers into adopting new behaviours. Now we shall see into more detail the different forms that OA can take at mechanical level.

1.3 The Technicalities of Open Access

What is OA exactly and why did it emerge in academia? Briefly, Harnad (2014) referes to OA as 'free online access to peer-reviewed research journal articles'. We argue that this definition fails by medling peer-review in OA, which is not at all necessary. In fact OA may also refer to monographs, widely used in the humanities, and not only articles (discussed ahead). We will use this definition for now, because it highlights the relationship between OA and scholarly literature. Which means that OA does not relate with other types of literature other than academic.

So OA is a term radically related with scholarly publishing and the way this niche of the publishing industry takes action. OA is a concept that broke into mainstream media right after the year 2000 associated with the idea that publicly funded scholarly literature such as research and scientific papers as well as precedings and conference papers, book reviews and even monographs should be kept under an open licensing in order to enhance access, communication, interlay among researchers and institutions, and scientific development at large. On a political and

ethical stand the movement towards OA was also seen as an initiative to cut off the middle man – publishers – and claim scientific discovery back to the public domain. The primordial arguments for OA are very simple and robust and were directed against the mainstream publishing mechanisms, namely aimed against subscription fees, or access fees and paywalls.

The main argument for OA publishing states that publicly funded research should remain in the public domain. This not only would benefit information accessibility and speed up scientific discovery since research output would be available to everyone and speed up scientific discovery, it would reduce universities' spending by avoiding publications from being payed for twice. Scholarly publishing is seen as being payed for twice, in the sense that universities pay for the research in the first place and end up paying publishers to get access to their research. This takes place when millions are spent yearly by universities' libraries in order to subscribe or buy scholarly journals. This way, universities pay for researchers to produce the knowledge in a first stage, and afterwards they pay publishers to get access to that same research. Once published, scholarly research, in the form of articles, is copyrighted, rights reserved to publishers. Scholarly publishing has become a very lucrative business[2] for a very low number of big publishers. A second argument questions the legitimacy of traditional publishers by stating that the value added by the most central step in quality control, peer-review, is done *pro bono* by researchers at universities (Larivière et al., 2015). Finally, with the rise of the digital in determent of traditional paper and ink, copying and storing costs are going down severely, which contributes to add doubts on the active mechanisms dictating scholarly publishing. In this dissertation we shall clarify these underground mechanisms dictated by unstated socio-cultural and symbolical intricacies that stakeholders in the game of OA share among each other. First we shall go through the discriminated commonly accepted norms ruling OA development.

Traditionally, scholarly communications use journals and article publication in order to keep recent academic works reachable for peers and other researchers. Scholarly publishing is thus usually done by either scientific societies[3], university presses[4], which are commonly non-profit organizations, or by private for-profit publishers[5]. The latter usually own a very large amount of journals, have headquarters in several countries, and provide more services than publishing alone. Although scholarly publishing dates back to the XVIIth century, it was not until later that it would become the standard for communication among scientists and academics. After the turn of World War II. private publishers were becoming the norm in scholarly publishing. Around the 1970s, these publishers started acquiring key journals, and merging into bigger publishing houses.

During the last years of the XXth century and the transition to the digital, the economics

[2]Monbiot and Buranyi track it down to Robert Maxwell and unearth the term "a perpetual financing machine", see `https://www.theguardian.com/commentisfree/2018/sep/13/scientific-publishing-rip-off-taxpayers-fund-research` and `https://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science`

[3]e.g. the American Society of Chemistry (ASC);

[4]e.g. Oxford University Press (OUP);

[5]e.g. Nature Publishing;

of scientific academic publishing[6] was already showing concerning signs of dependence on external factors. By external factors we mean factors related with knowledge production that are unrelated with knowledge itself, factors from a sociological, economic, cultural or political nature. By the turn of the millennium, most of scholarly publishing was done by a few private publishers, and more recently, by only a handful.

Currently journals are the *de fato* platform for scholarly communications. As seen above, scholarly journals have been around for centuries, but only recently have they become so commonly privatized and a source of income for its owners, as well as a strategic point of control in a broader information and innovation economy. The rationale behind scholarly publishing has been changing since the last decades of the XXth century. The dependence of academics on publishers has been rising as the number of publishers is going down; consequently academic research has become more and more dependent on a smaller number of increasingly bigger publishing organizations. As researchers develop the need to publish in specific journals, these become key gateways through which publications need to get through in order to become relevant.

The idea of OA gains power and visibility in a context where scholarly publishing becomes more important in a broader political-economic scenario. In a way academics, universities and librarians are claiming back control over their own product. Also it coincides with a growing incapacity of libraries to keep up with the rising prices of journal subscriptions.

The concept derives from other similar models such as (Free and Open Source Software (FOSS)) and share the characteristic of not being proprietary or owned by any particular individual or organization with exclusive rights to distribute, copy or develop. (Heylighen, 2007). The proposition is that such models enhance productivity and performance of the individuals and working groups involved in these projects by granting a fluid use of the Internet's technology capabilities, freeing data and communication and promoting what some authors call 'stigmergy' (see (Heylighen, 2007)). FOSS projects such as Debian (see (Coleman, 2013), or GNU is Not Unix (GNU) (see (Raymond, 1999; Stallman, 2002)), are early, well known examples of successful 'open' projects. Similarly, in scholarly publishing, the arXiv (Ginsparg, 2011b) is also an early example of a successful open strategy for knowledge management, as well as the *open access citation advantage* (Harnad et al., 2008) which we will describe in more detail ahead in the text.

Since OA emerges in parallel with an 'access crisis' (Suber, 2012) in the academic world[7], the actors involved in this revolution quickly became quite noticeable, and accordingly to what a global interpretation of the world economy might suggest, those actors involved are specialized institutions or groups of people integrated in a worldwide distributed chain of production, thus

[6]Scientific in its broader terms, i.e. including Science, Technology, Engineering and Medicine (Science Technology Engineering and Medicine (STEM)), as well as the Humanities, Arts and the Social Sciences (Humanities Arts and the Social Sciences (HASS));

[7]Characterized by the rise of subscription prices and the decrease of publishers numbers;

involving several people and institutions in apparently unrelated areas of economic, cultural and social activity.

When librarians and universities became aware of the capabilities of applying OA strategies (such as institutional repositoring) they quickly invested in implementing them due to the power that such strategy would grant their institutions. Basically, university's libraries would be able to regain control over their own production, which was going directly into the hands of private multinational publishers (and still does) whose interests are increasingly deviant from those of academia. Intimately involved with these two central players of OA, are funding bodies or institutions, that can be either private or state sponsored, which brings in to the for big governmental organizations such as the Research Councils UK (RCUK) , or the 'Wellcome Trust' in the UK, that tend to centralise most governmental funding within their structure. Research funders are key players in today's OA developments, which by granting funds to research might enforce a top-bottom, mandatory, OA licensing and publishing to be coined in their sponsored research.

Other big players in OA are publishers and journals, with their editing boards normally composed by members of academia and others, granting them a relative level of influence within academic think tanks. Indeed as we have seen in the historical contextualization section, there was a particular transition in scholarly publishing where most publishing went from mostly published by scientific societies to commercial publishers. Publishing became an important market and private companies became multinational corporations through several years of mergers, acquisitions and predatory behaviour. Nowadays publishers are put at the centre of an exponential rise of subscription prices – for instance, Elsevier presents a profit margin above 35% (913 million £) for the year of 2017. Historical observation tends to reinforce the idea that digital publications are favouring these publishers by lowering production costs – abandoning paper printing for instance – and widening their market to a more globalized market, with an increased number of consumers and producers. In fact publishers are getting increasingly pro-active into aligning their practices with OA policies, and their total revenue in OA publications is steadily growing by the day (Aspesi & Luong, 2014; Pinfield, Salter, & Bath, 2016).

Another important stakeholder is the researcher. The researcher's logics are particular to him or her, since researchers publish articles for two main reasons. First of all, they publish and read in order to execute and transmit their investigations, that is the main, direct use of scholarly communications. But researchers also obey quite selfish logics. Logics that emanate from the labour market, such as going up the institutional lather, keeping grants, publishing work accomplished. These intermingle with personal motivations non-related to work – researchers can be seen both from the community side and from the individual side. So in a sense, for most researchers, those who do not see themselves as OA advocates, or who are not even aware of such a concept or its intricacies, OA is not the priority – as we will see in the analysis section, this may well be the case for countries in the global South. The priority is to publish in high impact journals and to have strong bibliometric indicators such as a high *h*-index. Concepts such as the *h*-index or the JIF are constantly put under scrutiny by many (Stephan, Veugelers, & Wang,

2017), and several studies tend to reveal that these indexers can be more deceiving than helpful at times. But in practice it is widely claimed that most institutions use those indexers when it comes to hiring researchers, because the number of researchers and publications seems to render impossible a finer method of staff selection (Stephan et al., 2017). This makes it that researchers are most often worried about publishing in those prestigious venues, most of them owned by private companies than they are worried about keeping their work publicly available. It makes it also the case that interest in a few journals is essential, catapulting prices of subscription or publication charges (see Article Processing Charges (APC)s ahead in the text).

The fact that globalization spreads worldwide does not mean it is not extremely territorial. The case of OA is not an exception. While there maybe a theorisation of the cyberspace (Barlow, 1996; Lessig, 2006), as space populated by a community ruled by specific contexts that go beyond the usual physical borders, this does not exclude that different regions in the world will assimilate differently similar mechanisms and develop different methods of appropriation of that same space. In a global economy different parts of the world are usually involved in different parts of the production process; this translates in academic publishing that different regions of the world will exert different kinds of OA . Another way to put it is saying publishers, researchers, universities, and all stakeholders will develop different relations between each other according to their inherent logics. This, and I shall argue for this further throughout the analysis, means that different regions of the planet will have different uptakes and interpretations of OA . Some will produce more, some will consume more, others will build new channels of scholarly communication, seeks independence, others will attempt at entering into direct competition with the dominant production sector. Finally, as we shall see, there are also those who are excluded from the process of producing knowledge. As this will be discussed thoroughly during the analysis section, I will not enter into detail in this introductory section. The core idea one should retain at this point is that while global there are many types of OA, arguably, as many as there are different position in the global political economy.

1.4 Open Access, from Movement to Institutionalization

> *'But sharing isn't immoral — it's a moral imperative.' (Swartz, 2008)*

1.4.1 Licenses and (H)activism

OA, is described as a movement, or an initiative by many authors, but it can also be analysed as a managing toolkit for digital content – a bundle of policies, recommendations and laws in a process of high institutionalization. OA literature, as described by Suber (2012) 'is digital, online, free of charge, and free of most copyright and licensing restrictions', and this applies to

anyone able to access the Net. So OA is also a very global concept in the sense it unites groups of people with matching interest, despite location, country, age, religion, etc. In this section we will start by discussing some of the concepts encrusted around the legal aspects of data managing with a very strong focus on scholarly publishing and OA. Copyright, Copyleft, Open Licenses, and particularly the Creative Commons License will be deeply discussed in the first half of the section, while the second part will delve into the more political and ethical aspects of OA, commonly described in the literature as *advocacy*, while drawing special attention to some tensions points between the OA movement and its high levels of institutionalization.

Copyleft: a spoof on the word copyright. Copyright is a form of Intellectual Property (IP), a legal right common in most countries, that grants its owner exclusive rights for several years after its creation. Copyright is becoming a key mechanism of control over information in contemporary digital society as we will see. The condition of exclusivity grants the author (or the owner of the IP – often publishers) the right to control distribution or copy (copy-right). Copyright dates back to the beginning of the printing press in the XVIIIthcentury and was envisioned as a legal enabler that would grant publishers and authors a legal protection through exclusivity rights that would ensure there was a sustainable way for authors to be published. Berry (2008) argues that while it was created at the concurrently with the printing revolution, only by the turn of the century copyright began being reinterpreted in the context of a digital economy. In this context, 'copyright constricts the flow of immaterial goods and thereby increases the scarcity', Berry (2008) adds, 'it is an artificial scarcity', 'only held in place through the operation of copyright law, which was not intended to operate as a restriction on the flow of knowledge indefinitely', thus being 'transformed from its original intention (...) artificially create new markets' (Berry, 2008).

Copyleft can be understood as the practice of licensing work freely (in the sense of freedom, not "zero price"[8]). Rather than publishing it uncopyrighted, the idea is that a 'free' work should be licensed as such, ensuring that down the line it will not be privatized (copyrighted) down its life cycle. In the process of describing the difference between uncopyrighted work and 'copyleft' the GNU project explains that 'the simplest way to make a program free software is to put it in the public domain, uncopyrighted', still, this would allow 'uncooperative people to convert the program into proprietary software' (see `https://www.gnu.org/copyleft/`). In this case, what applies to software can be applied largely to all sorts of data commodities or artefacts, such as video clips, sound tracks, videos, documents, databases, etc. So the sharing principle is that uncopyrighted material is vulnerable to copyrighting thus the need for a robust licensing that would grant the public domain the full right to remain so.

In the context on software developing, since mid 1980's until today several licenses were made. Such are for instance the Massachusetts Institute of Technology (MIT) license, Berkeley Software Distribution (BSD) license, as well as the GNU Public License (GPL) license – no-

[8] `https://www.gnu.org/copyleft/`

toriously used in the flagship project Linux[9]. These early examples of software licensing were important in order to provide antecedents and proof of success of the model, which incited other areas of knowledge to adopt open licenses. If copyleft licenses were so successful in developing the Linux, GNU Network Object Model Environment (GNOME), Android and most of Google's infrastructure, what are the capabilities of such a model applied to other areas of production? Where can we apply copyleft that is going to benefit productivity and innovation? As the digital economy grows, and more and more content is put online, the answer gets clearer everyday.

Most data could and probably should be made free, so the naïf 'information wants to be free' saying goes[10]. Most digital content can be put online and its increasingly harder to keep data private. There are two main reasons for this to be true. First because of what many authors call the principle of non-scarcity (Heylighen, 2007). Innovation material is commonly not hard to replicate. This has been true since European monks quit writing replicas by hand and Gutenberg invented the moveable type. This scenario became especially more acute with the rise of the digital. It is not a coincidence that data storing is getting cheaper by the day, data processing as well at the same time as Open concepts gain energy. In the digital realm we have reached a point of a virtual non-scarcity, which means that replicating an article would have the cost of near 0 cents a copy, while production costs are all concentrated in the making of the original. The second reason that makes data increasingly hard to keep private is the fact that in a networking system, data is only valuable when available. That is why, as we have seen above, one of OA 's requisites for articles is for them not only to be freely accessible, but also to be findable[11].

Consequently, privatizing, or blocking access to information might bring some competition difficulties for corporations in a market of information. Traditional copyrighting can become problematic, since copying and transferring is becoming quick, cheap and global[12]. For a big corporation like Elsevier it is becoming increasingly expensive to try to take down copyright infringement, since most of it is located in several countries with several different laws, although efforts to standardize copyright laws are constalty being lobbied[14]. This seems to suggest that, seeking to maximise their outcome, publishers will possibly flip into OA models. In fact OA models, as explained in the above section, seem to be designed in such a way to relocate income. Corporations in this line of business seek income in several forms of information management and not only in scholarly publishing, which also means that, while scholarly publishing is not a whole industry of its own, these multinational conglomerates gain from inflating their ecosystem with, good quality and high mobility data.

The Creative Commons (CC) License is the accepted standard for OA publishing, referred to in policy, widely used in academia and also backed up by a community of experts represented by

[9] Actually GPL2.0;

[10] See (Doctorow, Gaiman, & Palmer, 2015)

[11] The same applies to the concept of Findable Accessible Interoperable and Reusable (FAIR) science in which 'F' stands precisely for 'findable' (Commission, 2016);

[12] See for instance EU's Pirate Party movement 'Kopimi'[13], or Sharism (Lindtner, 2015; Lindtner, Anderson, & Dourish, 2012);

[14] The effects of 'piracy' will be explained in more detail in the last section of this chapter;

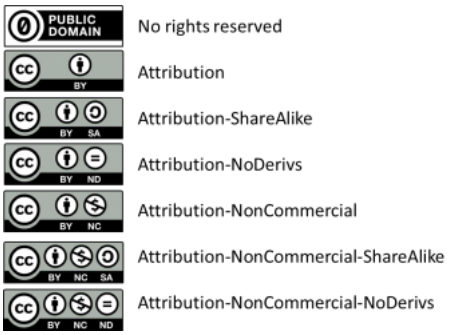


Figure 1.1: Creative Commons Combinations

the CC Foundation which makes it particularly flexible and reliable. The CC License is a legal instrument that applies a principle of 'some rights reserved' urlhttps://creativecommons.org/faq/ to any type of media product such as a video, track, picture, book or a scholarly paper. Essentially it is a shield that ensures that whatever is under this license is kept in the public domain while on the other hand granting also that the author of the work gets its recognition for it. On a practical level, the CC License is a very simple tool from the author's perspective, while at the same time providing robust back-end legal protection. This back-end, while maintained by a strong community of academics, most of them lawyers an intellectuals (Hetman-Krajewska, 2016) is 'specific for each country to comply with national general copyright regulations' (Hetman-Krajewska, 2016). The author is given a choice between a rank of permissions or 'attributions' that will enable more or less degrees of freedom to the licensed artefact. These freedoms will dictate how a third party can re-use the licensed product.

The CC License is made up by 4 elements: CC BY[15] - when present, this element will require the third party to give credit to the creator of the original piece. This is thus the primordial, basic element; CC NC[16] - No one else but the creator can make money out if it; CC ND[17] - when such an element is present there is no permission to change the original artefact, which means one cannot alter, but simply use as is; CC SA[18] - By keeping this element present in its works license, an author legally binds a third party using the original piece to keep the same License Elements as the original in future versions of this (See Fig. 1.1[19].

By combining these four permissions, the author can make up to 6 different combinations, all of them different among each other, some are more permissive, others are less so: Some studies suggest that the CC license more commonly allowed among the bigger publishers are the most restrictive ones (A. Björk, Paavola, Ropponen, Laakso, & Lahti, 2018)[20]. The CC license thus

[15]Attribution;
[16]Non-Commercial;
[17]No Derivatives;
[18]Share Alike;
[19]`https://library.osu.edu/copyright/creative-commons`
[20]Of course that will depend on the way the study defines a restrictive CC license. In the case of A. Björk et al. (2018), the authors define restrictive any CC license that is not the CC-BY;

acts similarly to the GPL and other software specific licenses by maintaining its subject under the public domain, but has a wider flexibility in several aspects. Contrarily to many copyright licenses, it also designed for the Internet, or with it in mind, so its suitability is naturally higher in contexts of digital publishing.

Since the initiative in Budapest, acoa started being refered to as a social movement (Guédon, 2004). From a sociological perspective it is interesting to question this position and ask whether OA is indeed a social movement rather than simply a shift towards an OA model of publication. Much of the research on the matter seems to address issues intimately related to social movements: awareness, advocacy and public good are the concepts most often used on studies and even reports on OA. In fact, research in OA, is done, not surprisingly by many researchers from different spectra of academia, from biology and computer science, to media and humanities, and many different approaches are taken by different researchers. Some use techniques of data analysis traditional of their disciplines, others focus their research on giving their own research community a feedback on the developments of OA in their area of study. Usually an advocate for OA, see for instance (Allington, 2013), will self proclaim himself as such since early days of OA, see for instance (Stevan Harnad, 2006). OA is seen as moral stand, where technology and science are put side to side with moral integrity against power centralization and injustice. It is conceived by its advocates as a tool to change the world, an active position to redefine society in its less self-aware definitions.

This takes us into the idea of *awareness*, another widely talked about concept inside the movement. Many reports and studies emphasise the importance of awareness on OA or lack of it. During the famous Stop Online Piracy Act (SOPA)/Protect IP Act (PIPA) fight for Internet freedom, it was many times said by various actors, how a lack of awareness on Internet play a huge deal in protecting accumulation of capital and power. Researchers awareness for instance is central for a pro-active dissemination and application of OA policies and general practices. Curiously, O'Carroll et al. (2017) states that awareness is less common among younger generation researchers than older ones. Awareness is important in the sense it represents the action of several groups of professional academic workers, from librarians, to researchers and university staff, to bring to the fore ideas that might be relevant to all spheres of society. Without the awareness, the potential for change, in say, countries of the South is reduced and the opportunities offered by the ICT and the digital are attenuated.

This brings us to the idea of public good. It would not be the first time that the Internet would host a social movement within itself, but in a way it is an innovation the way that OA breathes inside and outside of it – the implications of OA, while very material are still very much ideological in the sense they carry the benefits of the commons, of the public domain, as if a sense of one single commons would bring together the public, disregarding national borders and politics. This discourse is not completely alien to the multinational global corporation, and as we will see ahead in this study, OA and especially in the context of an open digital publishing market, is going to be very well incorporated into corporate logics. This is not say

that the OA movement is flawed or a fraud, but simply that there are practical implications into licensing that make it so that OA in the academic context is vulnerable to commercial logics. There may be an OA movement, but OA is more than just that movement. OA is reaching high levels of institutionalization, and involves, for instance, an investment several million of pounds in the UK alone, nevertheless a fervent community of advocates is constantly active in its dissemination.

1.4.2 Institutionalization

Since approximately the last 10 years recommendations, policy[21] and several reports[22] have come out of national, international, political and governmental organizations – for instance the European Commission (EC), mostly in the context of the European Single Market and the European Research Era (ERA). It is generaly stated in these documents that an OA approach on scientific knowledge would generate great benefit to the economy of information and innovation. Not long after the BBB foster2007, Europe already started deploying several infrastructural measures such as policy and data repositories, and were referring to the OA movement (Commission, 2007) in official communications. Europe would invest a sum of 50 billion euros during the 7 years of EU's Seventh Framework Programme for Research and Technological Development (FP7). By the end of the program Europe's infrastructure of institutional repositories and related policy was already quite developed.

Numerous organizations and initiatives were created within countries of the global North with the goal of promoting Open Access, Open Data, FAIR Data, Open Science, etc., with the common characteristics of being European Union (EU) based organizations advocating for OA measures implementation, either giving support, reporting, surveying, etc. One good example of this is the openAIRE project, created to 'provide technical infrastructure and support' for the 'identification, deposition, access and monitoring of results from FP7 and European Research Council (ERC); projects' (O'Carroll et al., 2017). Another project, FOSTER, would be put up to provide researchers the training required to make good practices in an 'open science' environment. In this context, OA and Open Science not only seeks to make scholarly literature free to use, it also seeks to make scholarly literature findeable and accessible, which in practical terms involves the development and deployment of data mining and scraping tools, i.e. not only building repositories but the tools to manage information (including datasets) and build new indexers.

Despite the huge infrastructural attempts to make scholarly literature freely available, by the year of 2017, levels of dissemination among researchers in western countries were still in an early stage. The reason why researchers still do not seem to embrace Open Science or OA

[21](Commission, 2007, 2009, 2010a, 2012a, 2012b, 2012c; Commission & Commission, 2012; Commission et al., 2012; Commisson & Commisson, 2011), etc;

[22](Commission, 2010b; Finch, 2012; Hodson et al., 2018; Lynch & Lippincott, 2005; O'Carroll et al., 2017; RCUK, 2005; Research Councils UK, 2012; Tickell, 2016) to name a few;

practices is commonly referred to as lack of awareness[23]. Another recently talked about factor seems to give emphasis to the relation between researchers' publishing behaviour and career management (Brienza, 2011). Eve (2014) and Suber (2008), for instance suggests a form of prestige that some journals enjoy, a symbolical capital that forces authors to be published in specific journals disregarding their openness status and ultimately affecting the larger countours of OA.

The OA or Open Science project is a global project in the sense it does not regard traditional borders and involves a holistic integration of all stakeholders involved in its process. Currently there seem to be two major approaches by institutional OA to be disseminated radically. Boycotting and OA mandates.

Currently several funding institutions are adopting an enforcement on OA, commonly referred to in the literature as Mandatory OA. Funding institutions adopting these measures are increasingly common; they require that the work they fund gets published under an OA license. In practice researchers are obliged to ensure that their work gets published in a way that it ensures it is kept under the public domain. It might also contribute to make the process of attributing an open license a simpler project, since everyone needs to do it. Accordingly, the CC license is commonly the accepted standard.

In a sense what mandatory OA does is it delegates awareness to a second level of importance and to many it feels like the fact that it is being enforced into researchers makes them less free and more conditioned to external obligations: ' I feel like a man with a beard in a country where shaving has just been banned.' writes Allington (2013), a self-proclaimed OA advocate in a blog-post. Mandatory OA can be sometimes controversial not only because of its top-bottom approach, but also because OA models do not seems to be well tested enough, or even developed yet. This is felt for instance in the context of monograph publishing (Finn & Fisher, 2018) within the UK universities. Researchers in these universities are required, in the context of the Research Excellence Framework (REF), to publish their work in OA in order for it to get evaluated in a process that will later dictate the university' funding amount. In this case, researchers, mostly in the humanities – where research is most often published in the form of *circa* 200 pages monographs – are going to be obligated to publish in OA, while the models to do so still seem too untested, too expensive and are worrying some authors into thinking this might not be at all beneficial for their study area, and that it might even block them from getting their work published, since Book Processing Charges (BPC) might scale up to 11000£, and since the humanities are going through a 'dry climate' this might undermine authors' freedom to make research in the subjects they find more suited and force the into following popularity trends directly (and merely) related with the publishing business.

Mandatory OA surely seems to be an efficient way to disseminate OA, though: see (Harnad, 1999) in (Ezema & Okafor, 2015), that conclude that within two years of the adoption of the mandatory archiving policies the depositing rates were 'close to 100% (Ezema & Okafor, 2015).

[23] see the end of last section

So, even if some researchers think of it as top-down interference, its efficiency seems to be in a general approach to be extremely high.

On another move by institutions to foster OA, the boycott to Elsevier is a good example of how the movement is both supported on an institutional level as well as in social level. In this boycott, several universities stopped or did not renew their contracts with the multinational publisher, due to the fact that '1) they charge exorbitantly high prices for subscriptions to individual journals. 2) In the light of these high prices, the only realistic option for many libraries is to agree to buy very large "bundles", which will include many journals that those libraries do not actually want. Elsevier thus makes huge profits by exploiting the fact that some of their journals are essential. 3) They support measures such as SOPA, PIPA and the Research Works Act, that aim to restrict the free exchange of information.'[24][25].

A similar movement sparked in Munich's city council, regarding FOSS. The project, named LiMux, was an attempt from Munich's City Council to migrate local government IT services to run on FOSS. Since this move was highly political – backed by political parties and ideologies – the political outcome of local elections inside the local government dictated the end of the project, still similar cases sprout in other regions, especially in big cities such as Barcelona[26]. This can be compared with a boycott to Microsoft. FOSS does not normally directly and actively oppose proprietary software in the way OA in the west opposes big publishers. This might help understand some of the obstacles that the OA community and movement are facing.

1.5 Open Access Models:

There are two main models introduced in early literature to modulate OA – the Green and the Gold route. These two models for OA publishing are already exhaustively described in the literature since they were some of the primordial concepts to be developed in the matter. The consequences of OA development, though, evolved in different forms of OA. These have been designated as Bronze and Black, but should not be confused with the first two in the sense they are of two different natures. While Gold and Green are accepted publishing models – whose terms were coined during the BOAI – Bronze and Black are derivations of OA that can only be accepted as such in a context where markets are populated by lack of regulation and piracy. These terms were coined several years into the implementation of OA, they derive mostly from the interpretation of what OA should be versus what it actually is. We shall begin by describing Green and Gold OA, as well as Hybrid OA – which we assess here as a derivation of the two – and their relation with the historical development of OA, institutional repositories and OA journals. Further down we will dedicate a subsection to introduce the concepts of Black and

[24] See thecostofknowledge.com;

[25] https://twitter.com/petersuber/status/1030887107848556549

[26] See for instance `https://itsfoss.com/barcelona-open-source/`

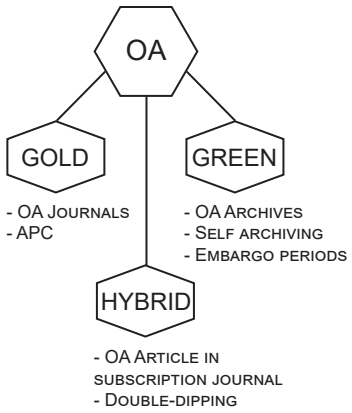


Figure 1.2: Open Access Routes

Bronze OA, and their relation with piracy, social medial, unregulation and copyright, we will see how in some ways, these cannot be considered OA models even though they should be taken into account.

Green Open Access

Repositories are the core infrastructure for OA and were among the first moves libraries, along with universities, executed in order to keep research on their side of the wall. Repositories 'centralize, preserve, and make accessible an institution's intellectual capital' (Ellen Finnie Duranceau, 2008). Steadily and early in the history of OA, western countries implemented institutional repositories (Lynch & Lippincott, 2005; Stevan Harnad, 2006) – and reaped the benefits of being in greater control, with the organizational capability to store and manage their own research. The notion of escalation was present in most of these initiatives, so the technological infrastructures in which this was done, was always aiming at findability and accessibility, two main topics of FAIR data today (Hodson et al., 2018). Thus also the protocols regarding metadata management such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), were implemented since its inception.

Institutional repositories are characterized by being owned by a big research institution such as any university or funder (this is common place in pharmaceuticals and medical disciplines) although recently we are also able to find institutional repositories owned by private companies such as Elsevier. Be it because they built their own repository or because they bought an already existing one. In fact, the implementation of institutional repositories in these countries was quite acute, both because there was the need to implement this technology, and accordingly because of the need of universities to regain managing control over their work and their relationship with the publishing industry. This and the implications of any technological innovation (excitement,

opportunity, etc) incited librarians to invest in modernising infrastructures and procedures, thus these repositories became very useful to safely store works such as Masters and Doctorate thesis. Recently, as storage prices go down, repositories are also the place to store datasets, or databases (Kircz, 2005), and because they were built with scalability in mind, they very certainly will be very important in tomorrow's Open Science, a strategical pillar of ERA and of the European Single Market.

If infrastructural implementation of institutional repositories was successful in leading research countries it was also in satellite or semi-peripheral countries (Wallerstein, 2004) such as Portugal (Santos, 1985) and also in countries around the world such as Brazil (Costa & Leite, 2008) and India (Mukherjee, 2012). As we will see ahead in the text, institutional repository adoption plays an important role in defining what type of OA is applied in a certain location of the planet, and will have a deep correlation with the type of research that a certain region of the world produces. For instance, China, as we can see in `http://v2.sherpa.ac.uk/view/repository_visualisations/1.html`, has a rather low number of OA repositories while having a higher number of journals, which takes into our next subject – OA journals – that differently from repositoring, provide researchers with the Gold OA route. We will discuss that next.

Gold Open Access

OA Journals are different from repositories in the sense that they only host published material. While repositories host mostly preprints, post-prints, dissertations, etc in an archiving fashion (the so-called Green route), OA journals are characterized by publishing as any other traditional publishing venue would, under an OA license. The Directory of Open Access Journals (DOAJ) is currently the main indexer for OA Journals, and researchers are usually advised to check with the directory before they contact a publisher. So for a journal, being indexed in the DOAJ is as a credential of quality and trust. This reflects a very particular situation relating to Gold OA journals having to do with the fact that many new OA journals flooded the publishing market when the transition to OA began, giving rise to publicly alarming issues such as bogus, or scam journals - inconveniently named 'predatory journals' in the literature (Beall, 2012).

So what is Gold OA? By definition Gold OA is an OA model that automatically grants OA licencing to an article as soon as it gets published. It is best described as a business model where the editor's income is shifted from the 'selling magazines' and 'subscriptions' point, to the 'publishing works' point. For a publisher to cover its expenses, survive and eventually to make its profit margin, instead of selling magazines and articles, it charges the author to publish in its venues. Most of the times, what this means is that the author's funder is going to pay the processing fees (commonly refereed to as APC). Athough the 'gold' allusion of the term, not all Gold OA requires the payment of an APC (Morrison, Salhab, Calvé-Genest, & Horava, 2015). APCs average price is hard to track down. 'Among the established OA publishers with journals listed in Scopus, (...) The current APC averaged about 1,418 USD.' (B.-C. Björk & Solomon,

2014), Morrison et al. (2015) finds a mean of $964 among DOAJ indexed journals charging APCs (only *circa* 26% of all DOAJ indexed journals would indeed charge these). According to Nassi-Calò (2013), the average cost could be around US$ 5,000 per article.

In a APC business model it also becomes profitable for journals to simply publish articles in higher numbers. In fact the APC has proven to be quite problematic in the sense it is predatory by nature, i.e. publishers are always incited to publish in quantity; it is a business prospect for many to publish articles that were not accepted by other more prestigious venues. These may be published without proper quality control and the journal may cease to exist in just years or months. This raises concern on the fact that unrevised articles are being published and that raises questions about the best way to preserve scientific information in what many call the data's life cycle.

Predatory journals can be defined as 'publications taking large fees without providing robust editorial or publishing service' (Clark & Smith, 2015). Beall (2012) defines predatory publishers as those publishing 'counterfeit journals to exploit the open-access model in which the author pays.' *'I first noticed them in 2008 and 2009, when I received spam emails soliciting me to submit to broad-scoped, newly-launched library science journals I had never heard of before.'* (Beall, 2017). The amount of such journals is particularly worrying: 'Beall counted 59 predatory OA publishers in March 2012. That figure had doubled 3 months later, and the rate has continued to far outstrip DOAJ's growth' (Bohannon, 2013).

Hybrid Open Access

Hybrid OA can be defined as a publishing model where the author has 'the option to pay an APC in order to make a particular article open within an otherwise subscription-based journal' (Pinfield et al., 2016). The rate of adoption of this particular model is quite low (below 2%) (B. C. Björk, 2012) from the side of authors – mostly because of the model's typical high prices (around 3000$). As B. C. Björk (2012) explains, *'in hybrid journals the Open Access is an extra luxury, since the article is published in any case. In the case of most hybrid journals the author can also achieve almost the same OA effect for free by uploading a legal manuscript copy to an institutional or subject-based repository.'*. On the other hand, a large of publishers operate a hybrid OA option (Pinfield et al., 2016).

Hybrid OA is also problematic because of the double cost it represents for universities. Since APCs are by universities and funding bodies, these in most cases would be required to pay the subscription to the journal, while at the same time paying an extra charge to publish one specific article with an open license inside it. Many acuse publishers of 'double-dipping' (charging twice for the same article, both in subscriptions and APCs). Although Eve (2014) refers that 'publishers do not wish to be seen charging twice for their work', and that many 'have implemented arrangements whereby the amounts paid in processing charges are deducted from the costs paid by subscribers or purchasers.' (Eve, 2014), it seems to be quite hard, if not impossible, to keep track and control of whether publishers are indeed or are not 'double dipping'.

Hybrid OA represents, either way, a good analogy of how OA is being incorporated into comercial publishing as a second and extra source of income, and that licensing an article with an open license might end up being extremely expensive.

1.5.1 The Importance of Unregulated Online Publishing in Access to Literature

> *'First note that the gratis/libre distinction is not the same as the green/gold distinction. The gratis/libre distinction is about user rights or freedoms, while the green/gold distinction is about venues or vehicles. Gratis/libre answers the question,* how open is it? *Green/gold answers the question,* how is it delivered?' *(Suber, 2012)*

Even though OA's core concept is about regulating ownership rights and permissions so that an open dissemination of artefacts is feasible in a lawful and consistent way, recent studies on the subject, not only suggest that a large amount of academic literature is accessed through unregulated channels, they also reflect a very deep necessity to incorporate these unregulated routes into OA's theoretic suite of colors. These are what some authors have referred to as Bronze and Black OA, a subset of Gratis OA, which we will discuss in detail throughout this next subsection.

Gratis OA, unlike OA, is not *libre* in the sense it does not offer legal rights of freedom or permissions to the use of the artefact. Gratis, as the name suggests, simply means that the artefact is available online without any payment restrictions. Although most of the literature refers to it as OA, strictly speaking, gratis is not open. In the software world the analogy of comparing *gratis* and *libre* would be to compare *free software* to *freeware*. Gratis does not grant the user the freedom nor the control, resulting in something essentially different and inconsistent with the proposition of the Open Access Initiative (OAI)[27]. Since the literature uses the term, we shall use it also, in order to keep consistency with the review of the terms Bronze and Black OA, which are, in theory, something other than OA.

While the ecosystem of Gratis OA is most important in academic papers accessibility, the type of access it grants is marginalized by the lack of regulation that characterizes it. The necessity to theorize new OA flavours comes as a consequence of recent empirical studies on the subject that attempt to estimate the percentage of OA articles in the total of scholarly publishing. These end up revealing that a significant amount of access is being offered not by OA Journals or repositories but by Academic Social Networks (ASN)s, Pirate Libraries, or simply *offered* by journals without the consistency of a proper open license.

[27] See `https://sparcopen.org/wp-content/uploads/2016/01/hoii_guide_rev4_web.pdf`

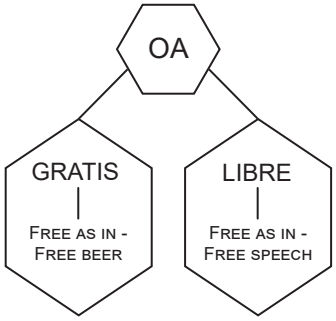


Figure 1.3: Gratis Vs Libre

Piwowar et al. (2018)'s study revealed that the majority of unpaywalled articles were accessible via the publishers' website while lacking a proper license. This kind of Gratis OA they dubbed Bronze OA [28]. It is a phenomenon that seems to be explained by a certain lack of awareness and consistency from the part of the publishers and journals regarding OA (Chen & Olijhoek, 2016). This also reveals that OA requires extremely high levels of regulation and consistency, seldom found in the publishing industry. In fact publishers seem to often be quite dubious regarding their OA policy[29], and most of the times, *'do not know'* or simply *'do not have'* any real OA policy. Bronze OA is thus a form of Gratis OA, sharing characteristics with Gold OA (being publisher-hosted) and Hybrid for they are published in subscription-based journals (Piwowar et al., 2018). Also it is not clear whether the artefacts are supposed to be accessible temporarily or permanently (Piwowar et al., 2018).

Dark OA (B. C. Björk, 2017) refers to articles shared online despite copyright infringement. These are massively shared either over ASN or pirate libraries.

ASNs[30] are essentially regular social networks[31] with the particularity of aiming at scholars and being specialized in the kinds of interactions particular to them. They do so by displaying citation counts, measures of popularity, etc, and by notifying connected nodes of newly published material while (and most importantly) giving access to the shared material to those with a registered account. Registration on these sites is often, if not always, free. This material does no

[28] What they also called 'dark gold' or 'hidden gold' (Piwowar et al., 2018);

[29] "Sage and Wiley-Blackwell used these well-defined terms in a potentially misleading manner by using the term "hybrid 'gold' OA choice", which essentially refers to the standard hybrid model (...) Elsevier interpreted our "number of OA journals" to include self-archiving and hybrid journals, instead of full OA . Whereas our expression "OA journals" allowed this interpretation, Elsevier's subsequent claim that "100%" of their journals are OA demonstrates the potential and willingness of the publishers to make misleading interpretations when the evaluation criteria are not well-defined and standardized." (A. Björk et al., 2018);

[30] ResearchGate, Academia, Mendeley, Semantic Scholar, Google Scholar, etc

[31] Such as Tumbler, Twitter, Diaspora, etc;

require any particular license or warranty of quality and is not even required to be authored by the same person who published it and do not check for copyright compliance (Martín-Martín, Costas, van Leeuwen, & Delgado López-Cózar, 2018). Strictly speaking this means that anyone can share the material they want, even if the document was authored by someone other than the person who posted it and even if that article is under a copyright license that restricts it from being shared in such platforms.

Being granted access to such material may be achieved through a free process of registration in such sites (Google Scholar and ResearchGate even bypass this requirement), making these ASN another source from which readers can access scientific information (Martín-Martín et al., 2018). For the organization behind these sites – mostly business ventures (B. C. Björk, 2017), – maintenance is quite low, given that they do not offer warranty on the quality of their service. They also do not work on the basis of prestige but on number of users (the higher the number of participants in the network, the higher the relevance for a non-member to become a member of that network), favouring, as in many other online systems, concentration[32], control, scrutiny and privacy violations. This gives them a privileged position in scholarly publishing, and inside the academic community. Twitter, for instance, although not being a social network is being used increasingly by academics to notify peers of newly published studies and of events and initiatives – the OA community is quite heavily present in Twitter, for instance.

Literature disseminated in such a way lacks a proper license and may be copyright-infringing. Thus it has the property of being unstable, unregulated and unpredictable. Such literature can be available one day, and be shutdown on the other. So for instance a law like Article 13 (Europe, 2017), or a legal court case (for instance big publishers against ASNs[33]), could potentially block the access of many of these articles quite suddenly.

Shadow Libraries

> *'Letting the opposition control the terms of debate means the debate is lost. Letting other people define one's behaviour as piracy means that all discussion will be framed in terms of classical property theory, and it is the framing of classical property theory that is an essential problem to the debate.' (Warwick, 2014)*

> *'It's called stealing or piracy, as if sharing a wealth of knowledge were the moral equivalent of plundering a ship and murdering its crew.' (Swartz, 2008)*

Regarding Pirate/Shadow library nomencalture we prefer to use the term shadow library and we refer to the above citations to do so. Although the name 'pirate' is widely used to refer to

[32] See Chapter 3 on Research Method – (de)centralization;
[33] A coalition of big publishers has in fact already been created: see `http://www.responsiblesharing.org`

'online piracy', it seems that 'shadow' nomenclature is more adequate to refer to the actions of these radical libraries, that do not have anything to do with piracy.

ASNs are not essentialy invested in ethics, they are mostly business ventures, start-ups, who, like Mendeley for instance, end up getting bought by bigger companies, being part and parcel of a hegemonic form of globalization in what I shall identify with (de)centralization [34]. Shadow libraries on the other hand are part of some sort of a collective connoisseurship activity – centralizing in a single place a diversity of both important and rare literature artefacts – civil forms of desobedience and ethical stands directly linked to the preservation and access to human cultural production. Piracy in these terms is also identified in the literature as a form of redistribution [35], but it is also the consequence of 'a seriously flawed mainstream scholarly publishing model, which has failed to adequately adapt to the needs of the international research community' (B. C. Björk, 2017) and a symptom of a 'social, political, and economic dominance of an unbalanced class structure.' (Warwick, 2014).

Shadow libraries have been around online and offline for many years, the Internet though, provided the ultimate tools to make this practice universally widespread. Shadow libraries are nowadays quite prolific in terms of coverage and use although scarce in number and often living short, hard lives. Shadow libraries have historically especialized in scholarly publications and provide the kind of scholarly literature fitting the very recent Black OA nomenclature. These online libraries are for instance the Library.nu wich in the early 2000s was the biggest online repository for ebooks, hosting about half a million titles (about the size of the Library of Alexandria (Warwick, 2014); Aaaaarg.org; Monoskop; AvaxHome; Genesis Library (also known as LibGen), a major online library based in Russia, hosting primarily scholarly books and articles [36], and is characterized by the opacity of its operation. Looking into the literature, it is possible to infer that its rate growth is very pronounced: by 2013, Warwick (2014) stated that the Libgen entire repository would weight about 9 terabytes; by 2014 the value had risen to 42 terabytes of material, comprised of about 25 million documents. It then hosted about 36% of all Digital Object Identifier (DOI) articles (Cabanac, 2016), and circa 68% for the three major publishers. While by 31st January 2017, according to Gardner, Mclaughlin, and Asher (2017), the total number of documents would surpass the 61 million units. LibGen is also known for having hosted Sci-Hub's article database from 2013 until 2014 in its scimag. In 2015 Elsevier filed a civil suit against both Sci-Hub and LibGen (Himmelstein et al., 2018). Shadow libraries effectivenes reached a new peak with the creation of the Sci-Hub, another form of radical librarianism. Defining the Sci-Hub is not straightforward (Alexandra Elbakyan, 2017), especially throughout its history.

While early literature refers to it simply as a site, today, Sci-Hub hosts its own repository. Contrarily to the LibGen, the Sci-Hub's creator and main maintainer, Alexandra Elbakyan, has

[34] See Chapter 2, **(de)centralization**;
[35] See *Robin Hood OA* (Archambault et al., 2014) and (Oxenham, 2016);
[36] See scimag, a science 'magazine' section of LibGen;

adopted a quite open policy during the last three years – although the site has been around since Septembre 2011 – participating in several interviews, being active on Twitter, by releasing data dumps with usage statistics for the last few years and by working directly with researchers to help understand those statistics. The Sci-Hub is both an unique and unparallelled powerfull tool because of its most peculiar mechanism of retrieving paywalled articles on demand – its uptake levels render it a source of a deeper knowledge about scholarly publishing. This differentiates the Sci-Hub among the rest of the shadow libraries, because, unlike any other online library or repository, it can provide access to nearly all scholarly literature[37], in fact showing accessibility scores that nearly reach the total coverage. For instance Himmelstein et al. (2018)'s study found out that an impressive '91.0% of citations since 2015 were present in Sci-Hub's repository, which increased to 96.2% when excluding citations to articles in open access journals'. On a prior investigation, Bohannon (2016), found out that this online service was being used by people all arond the world, namely by students in several Northen universities.

[37](Himmelstein et al., 2018);

Chapter 2

Research Design

2.1 Research Question

Many of the studies reported in the previous chapters point out to the fact that open access is set out to become a major aspect of scholarly publication; a crucial condition for this is the increasing use of informational technologies in the process of reading and publishing articles – almost 20 years into the 21 st century this is barely an issue. Even though many people have access to the internet worldwide and accessing the internet is a daily routine for many in a growing number of countries, the internet can still be seen as being in a developmental and implementation phase[1]. In this context OA appears as response and attempt to develop internet regulation.

So OA appears in the West as a form conquering the open waters, a continuation of a broader ideology of openness and freedom of speech that has always circulated around the internet. This was achieved through the introduction of open licenses such as the CC, Open Repositories and Open Access Journals. While many publishers and funding agencies start incorporating OA into their business model, the emergence of shadow libraries and 'pirate' practices intensify as they both follow and lead the developments on copyright and IPR issues that echo in many other forms of industry and cultural production.

Many of the studies mentioned in the literature draw on notions relative to the fact that the scientific community is increasingly united around interchangeable objectives and forms of knowledge – thus the importance of an openly accessible and free channel of communication between scholars. This study will provide a critical position to these notions and assumptions by evoking some question relative to inequalities among countries – very sharp in contemporary international landscape (Milanovic, 2016).

Open access has become the field in which researchers produce a consistent discourse about the relationship of dependence between academia, commercial businesses and lobbying. The open access initiative, or movement, works this self-awareness of researchers by claiming back

[1] According to Figure 2.1 the percentage of the population with internet access is growing steadily not showing signs of slowing sown – nearly reaching 50%;

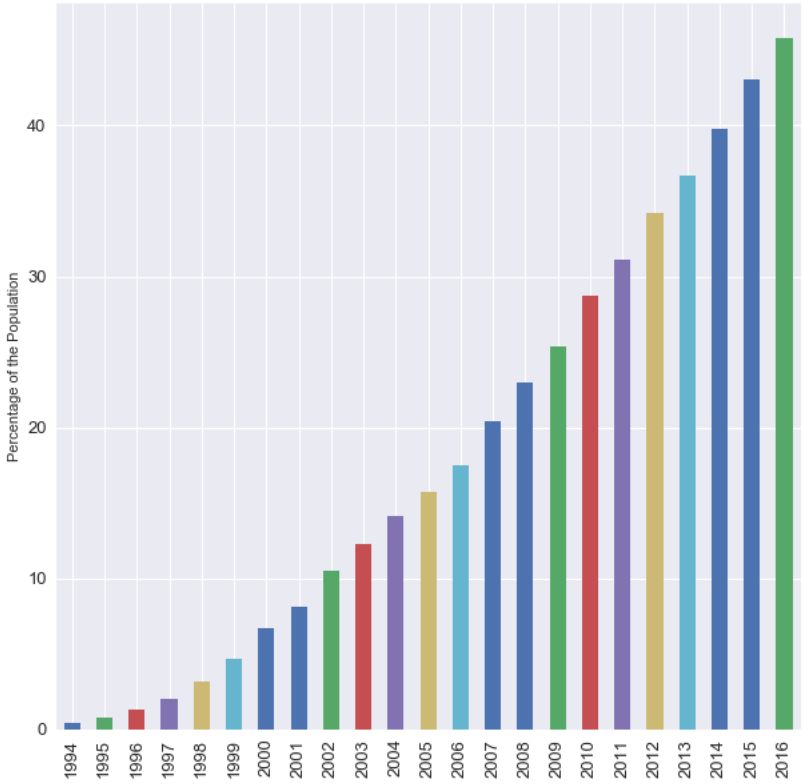


Figure 2.1: Worldwide Internet Access Progression (World Bank data)

the fruits of their work utilizing the ICTs as an innovational tool. It holds on to the principle that scholarly production and communication should be part of the public domain both due to the fact that it is publicly funded research, and that it is of public interest to maintain it as so. Curiously enough, most of research does not seem to by directly funded by governments[2] and some authors point out to the fact that only in rare situations does the very specific, often business/industry related academic articles are of interest to the general public (Allington, 2013).

Even though open access draws claims over the independence of academic work from corporative influence, only tangentially it does address this same problem; despite the fact that much of the discourses address and concentrates on the dependence of academics on big publishers, there is too an underlying tension reflecting the broader dependence of academic production on the interest of big money and global economy – a tension this study highlights and explore.

Did the open access initiative succeed?

The success of acoa is a critical question, for a successful OA would have different meanings for different stakeholders in academic publishing, we shall discuss this in higher detail. Although still locating scholars in an ethical-political position towards science (see the discussion between advocates against critics (Eisen, 2013; Poynder, 2017a, 2017b)) , open access has reached the point where it is no longer solely a movement among scholars. In fact, as for instance the

[2] See `https://scienceogram.org/blog/2013/05/science-technology-business-government-g20/`

implementation of mandatory acoa policy reveals, acoa is becoming the model of scholarly publication and communication, developed and adopted by major funding institutions, mainly in OECD countries, with the goal of making it the new standard for article publication.

Judging its success on its level of implementation and not so much on the quality of that implementation, the whole movement and idea of open access is proving to be at least quite proliferous. It is generally argued that the way this is being achieved is by disregarding some of the essential ideas that open access based itself on; one of these ideas would be to relieve researchers' from economical factors and restrictions. What effectively seems to have pushed forward the open access movement to its current point of high level integration, into different levels of institutionalization, is the sum of a series of timely events, that alone are not enough to explain the general adherence to open access, but when all brought together imprint a strong, collective pressure. The massive migration into the internet was a major step for this, and could to a certain extent explain the emergence of such a movement – because open access was simply just made possible by digital publishing and the possibilities provided by the internet – but on the other hand it is not enough to explain its major impact. In fact this question would require a whole investigation of its own, but these factors are sparsely referred by different authors in different contexts: the rising prices of subscriptions; open access's citation advantage, but also the success of Open Source in the field of software development and the principle of digital non-scarcity. All together they seem to have made open access unavoidable, both by universities and publishers.

The idea of open access rests on the bigger concept of scholarly publishing, a concept in itself quite problematic – because it rests on the unchanged, centenary system of article and monograph publishing. A system in which academia as a whole heavily rely on; explaining to a certain extent the advent of Open Science, Altmetrics, ASNs and other Social Media in academic communications.

As the morphology of scholarly publications evolves – the number of publications grows, OA is more widely implemented, new OA journals and online libraries emerge, – academia becomes more and more dependent on a reliable quality certification mechanism, capable of effectively separating 'good' research from 'bad' research. In practice this puts academic publishing facing a big fork on the road – either the industry gets more centralized by using journal prestige as the safekeeper of quality, or it gets more fragmented, decentralized and uses less traditional certification models, based on altmetrics and small communities of researchers to assess on the quality of articles. The latter is done by institutional and disciplinary repositories, and these do not substitute the role of journals, so, in a way, we can expect to see both of them co-habiting the ecosystem of scholarly publishing together. While big informational traders might acquire small repositories, indexes, libraries, etc, once these become meaningful competition, these are also able to emerge quite easily.

The centralization of publishing processes in big commercial publishers such as Elsevier, but also as Public Library of Science (Plos) may act on determent of diversity and variety of

forms of knowledge and may promote brain-drain from peripheral countries toward countries in the North, especially in the context of a wider economy of academic knowledge production. On the other hand, a more fragmented, more de-centralized and more innovative structure cannot (without causing harm to the economics of scholarly publishing) engage into conflict with the established *status quo* without hindering itself.

While there is a wide corpus of discourse and studies around OA in the context of a mechanism for scholarly publication, it seems that the getting access to literature perspective is less talked about in the literature. In part this might be explained by the fact that most of the OA movement is based in the US and in Europe[3]. In these countries the barriers to access are less pronounced since subscriptions are more common, as well as better internet access and internet literacy. The proliferation and reach of the Sci-Hub and other shadow libraries might be proof of some kind of malformation in the OA system, which despite its continuous growth show itself incapable of providing a *de fato* openness. The immediate reasons for this are for instance, embargo periods, which prevent access to meaningful literature, or simply the fact that researchers still rely on big publishers and Toll Access (TA) venues to better capitalize the fruits of their work.

RESEARCH QUESTION:

This investigation will balance the aspects of the open culture against the particularities of the OA movement, using it as a case study for the broader cultural movement, analysing the developments of OA against a background of critique, by taking into account concepts inserted into globalization, by developing self-reflective and objective knowledge. For that we engage in an empirical data analysis focused on answering the question of *whether or not OA is being implemented as part and parcel of hegemonic globalization or if on the other hand it should be seen as a tool for the proliferation of emancipatory and diverse forms of knowledge.*

The above research question is broad and encompasses the whole of our investigation. A smaller, micro objective of this investigation, nested inside the empirical analysis, is to provide insight into the motives that shape the behaviour of users in our sample[4].

The results and discussion of this investigation, eminently sociological, will hopefully be pertinent for those seeking a deeper knowledge of the cultural, economic and political morphology of the increasingly technological and interactive modern societies. In the next section we will be discussing the theoretical elements under which this investigation lies under.

2.2 Theoretical Framework

The presentation of the theoretical instruments is important to understand the shift of thought that the investigation experiences between the *literature review* and the *analysis chapter* and it

[3] Although, of course, there are exceptions to this such as CODERSIA and SciELO, to name a few;
[4] See Methodology section;

will shed light into the directions in which the investigation will be going. Even though it is not essential for the following of the analysis that will come in the next chapter, we recommend the reader to incorporate the theory presented in this section into all the analytical arguments and presentations pointed out throughout the rest of this dissertation. That being said and reinforcing the importance of this section for the cohesion of the comprehension of the argument, I will now enter into the description of the methods.

In order to understand whether OA is part of a dominant form of science or rather a tool for emancipating knowledge we will bring forth several theoretical sociological concepts to guide our empirical analysis. These concepts are the 1) globalization (taking into special consideration the concepts of agglomeration, subterranean trends, and the recent study on global inequalities); 2) the Network Society; 3) Information Economy; 4) Academic Dependency and; 5) Ecology of knowledge; Lastly I will introduce a model of (de)centralization that will provide insight into the recent and future developments of networking systems such as the ones affected by the open movements.

The central argument to my analysis rests the idea that OA is currently not only a cultural movement headed by librarians and academics, but that it is undergoing strong institutionalization, revealing signs of development and implementation near big institutions, multinational, international, not-for-profit, commercial, private and state sponsored, etc. OA is also shifting to a form of revenue, through the development of APCs, but also the reason to be of many non-governmental organizations (Non-Governmental Organization (NGO)s) and several increasing bundles of budgeting money. This I have shown in the chapter above.

Another underlying idea to my argument is that OA is inscribed in a flagship process of a globalizing industry characterized both by a network ecosystem and by technological innovation. This section will provide other authors (and one original) theoretical tools to understand how power, as information and money, and any hypermobile commodity tend to concentrate, reproduce and accentuate inequalities while technological innovation might provide ways to democratize and reshuffle inequalities in alternatives to forms of dominance. In other words, how in the context of an economy of information, the profit shifts from economic growth surplus to technological innovation, which by its turn cannot be *ad aeternum* and thus the dilemma. I will expose some of these theories next.

2.2.1 Globalization

The origins of globalization can be traced back to the early history of trade. On a not so long stretch, it is also possible to coincide the the origins of contemporary globalization with the 'age of discovery', still more than 500 years ago. In fact, many modern concepts related to the global society have taken place during that time, for instance, the binary distinction between New and Old World[5] and of course *Coloniality*, and many other important concepts to contem-

[5]Now represented by the dichotomies North/South, or East/West;

porary academic subjects. Martell (2010) suggests that historically, globalization has its bases in the development of capitalism, industrialism and the 'institutions these systems brought along' (Martell, 2010); contemporary globalization differs from the earlier experiences of internationalization (of trade for instance) by the fact that international relations are currently not only in a much more higher number but also that they are relations of systemic interdependency. (Sassen, 2007), suggests that 'the global – whether an institution, a process, a discursive practice, or an imaginary – simultaneously transcends the exclusive framing of national states yet partly inhabits its national territories and institutions.'.

Academics use the term *First Global Economy* to refer to the period between the end of the XIXth century and the Great Depression, while referring to the current stage as the *Second Global Economy* [6]. If by 1914 the 'world capital, commodity and labour markets had become closely integrated' (Jones, 2005) – based on railroads, steamships, telegraph, ports, electricity, gas utilities and an internationalization of global economics, law[7] and politics[8] – around that same time implementation of restrictive migrational, mobility policies and protectionist measures started to be put in place by several countries (a process that intensified with the First World War) (Jones, 2005), promoting the plunging of international world trade and the end of the first wave of global economy.

The second wave of global economy emerged in the immediate decades that proceeded from the Second World War ('the mother of all technologies' (Castells, 2000a, p. 46)). This initial phase is characterized by a prominent role of the American economic[9], political[10] and cultural dominance[11] in the world order as an aftermath of the great conflict[12]. The second global economy, also the contemporary global economy and what is commonly stated in the literature as *globalization*, in its transitional phase (until mid 1970s), is also characterized by the ascendance of Japan into a central position in the international trade market. Most importantly, powered by the ICT revolution, banking and finance changed drastically, polarizing into points of control and coordination – global cities (Sassen, 1991, 2005, 2008).

By 1971 the economic system that persisted from the post-war international agreements that we here present as being predominantly American, suffered great restructuring – mostly in the form of deregulation[13]. In 1979, under Deng Xiaoping's rule China got back into the global economy and several years later the Soviet Union collapsed giving space to newer liberal policies in eastern Europe and Asia. In stock market, 'the late 1980s and early 1990s saw the

[6](Jones, 2005, p. 144) 'provides a figurative illustration of the historical role of firms in worldwide economic integration';

[7]Such as International Property Rights;

[8]Passports were note required to travel inside Europe several decades prior to the First World War;

[9]For instance the Bretton Woods Conference;

[10]Much of international organizations such as the United Nations (UN) were then constituted;

[11]Mainly through movies and music;

[12]'Service firms such as consultants, advertising agencies, hotels, and film distributors were significant conduits for the international diffusion of American management practices, values, and lifestyles.' (Jones, 2005);

[13]See for instance the 'Nixon Shock';

addition of markets such as Buenos Aires, Sao Paulo, Mexico City[14], Bangkok, Taipei, Moscow' (Sassen, 2005)[15], giving a new importance to the BRICS[16] emerging economies in the world trade stage. These events would send global economy back into high levels of trade and reshape the morphology of industry and at large, globalization. It was also during this period that much of multinational investment became mergers and acquisitions, which is particularly relevant in the context of scholarly publishing[17].

Theory of globalization

Although economy is a big part, there are other aspects of globalization that make it such a particularly far reaching and comprehensive concept. In fact globalization is considered by most social sciences as a complex process that affects most parts of people's social lives (Held, McGrew, Goldblatt, & Perraton, 1999). Some of the most relevant defining conceptual axis used by theorists of globalization – besides economics – are culture, politics (reflecting in the role of the welfare state and social democrat politics as well the repositioning of the nation-state), migration, environmental dilemmas, conflict (war, but also trade war[18]) and technological development. Globalization thus should not be viewed as a fixed structure (Holton, 2005) but rather a process; in fact, globalization is political and thus ideological. Holton (2005), Robertson (1992), Waters (2001), emphasize the question of awareness (consciousness) which is also extremely relevant to the understanding of the developments of OA – a highly ideological-political issue[19].

Globalization theory has emerged around the 1980s (Martell, 2010). Martell (2010) conceptualized the theory of globalization into 4 waves: First wave – globalists – are characterized by being hyper-globalists, by attributing a great deal of importance to global economics and mobile capital to the determent of national economies and nation-states sovereignty, while technologies act as a facilitator of these processes. Martell (2010) also suggests that opposite political-economic ideologues (Marxists and liberals) of this stream of thought tend to diverge into whether this new phenomenon should be regard as a good or bad thing. Regarding culture and the future, first wavers understand that a homogenization (hybridization) of cultures is taking place, that globalization is both new and inevitable and understand that the social sciences are being required to radically change the way they view society to 'cosmopolitan and global perspectives on social relations' (Martell, 2010).

First wavers are often seen as 'hyper' globalists, 'thin on empirical substantiation' and quite abstract in their claims (Martell, 2010). Second wavers can be understood as response to this. That is why they are called the sceptical wave of global theory. Authors scrutinized globalist ideas through the adoption of an empirical methodology arguing for instance that the power of

[14]For more on history of South America in the global economy see (Sassen, 1991)
[15]See also (Birdsall et al., 1993; Young, 1995);
[16]Brazil, Russia, India, China and South Africa
[17]See Chapter 2;
[18](Joseph E Stiglitz, 2017)
[19]See previous chapter;

nation-states is very much alive, even if socialist and radical policies were especially limited by the mobility of capital that would eventually flee restrictive measures. Most importantly, sceptics realized that globalization does not have the same type of responses everywhere, nor by all cultures – global inequalities are stark and rising; Foreign Direct Investment (FDI), for instance, concentrates in western economies. Economically, they argue that free trade has contributed as much as protectionism for the growth of several economies[20], namely China and the US (Joseph E. Stiglitz, 2003) , and that essentially the phenomenon is international rather than global.

The third wave[21] ,while seeking to provide a more sound and intermediate interpretation of globalization between the two previous ones, offers a picture of a more fragmented globalization, reinforcing much of the sceptical views, but on an attempt not to reject globalization as a phenomenon. Even though economic interdependency between countries is a reality, and that state sovereignty is changing – shared between new stakeholders of international law and industry – this does not mean that these are final nor universal in the way they apply and affect countries. The future of globalization is seen as open-ended[22] (Martell, 2010), the hierarchy of nations is shifting from centre/peripheral (binary) to 'a three-tier structure, including a middle group of more successful economies' (Martell, 2010) – the Brazil Russia India China and South Africa (BRICS).

This group of countries share unique characteristics between them, such as the fact that they are reemerging strong economies, and the fact that they are not located in the West, but an also key feature of these countries is their population size, which makes them particularly important in a globalized supply chain, both as producers, i.e, labour market, extraction sites, and as consumers.

Second and Third waves do share a lot of aspects. Martell (2010) argues that transformationalists give a lot of points to sceptics. The two currents seems to essentially differ in the way they define what is happening, or what they are analysing: either internationalization, or globalization. The sceptical position dies off in itself in the sense it is sceptical and destructive about the subject it verses on, and exists for and thus ceasing to exist as a fully-fledged theory. So, can the discourse of globalization claim it not to be a thing? This takes us into the fourth and final wave.

The third wave proposes a more qualitative methodology to the analysis of globalization that the previous ones, and in this sense, the fourth wave fills some of the voids enabled by that stance. The fourth wave, Martell (2010) argues, can even be seen as an ally of the second and third ones. This stream of thought, is based on the notion that the idea of globalization is important for its realization in the real world – i.e. that the self-reinforcing discourses are important for the movement towards building it. This perspective is a post-structuralist, postmodern and social

[20] Again pointing out to the importance of state intervention in the global stage;

[21] Transformationalists (Held et al., 1999) in the sense they defend an idea of globalization that is transforming society

[22] See for instance (Joseph E Stiglitz, 2017) on trade war;

constructionist one (Martell, 2010), and gives emphasis to the consciousness and self-reflexivity and power of an ideology. In this sense, this perspective is also critical for it defies the discourse of the idea of globalization, through a discursive claim.

Recent Concepts and Discussions

The above presentation shows, conceptually, the evolution of the globalization theory. Much of the discussion surrounding globalization will tend to follow much of the arguments presented from a generic point of view in the above section of the text. Recently some concepts have been under the sight of scholars more than others and some of those fit better the purpose of our investigation, that is to determine whether OA is a tool of emancipation or dominium in the context of the global supply chain of scientific knowledge production. Many of these concepts are not new – after all, globalization is not made of new events alone – but these gain new meanings as new measuring tools emerge and phenomena get more distinct. For instance the study of global inequalities has been leveraged by recent studies of Piketty and Goldhammer (2014) and Milanovic (2016). Milanovic (2016) suggests the emergence of 'global plutocrats' (Milanovic, 2016), the global top 1 percent that gained the most from globalization. New tools have also been developed to understand institutional change (AOKI, 2007), such as game theory and complex theory (see for instance (Byrne, 1998)).

Network and Information Society

A crucial notion of globalization is the formation of an information economy and an informational society. Castells (2000a) speaks of a 'new society' composed of a new structure to describe in broad terms the changing society emerging from the turn of the millennium. The main constitutive dimensions of this 'new society' can be enumerated as follows: 1) technological paradigm; 2) globalization; 3) the internet; 4) the demise of the sovereign nation-state; 5) progress in scientific knowledge, and the use of science to correct its own one-sided development (Castells, 2000b). So as we can see, concepts intermingle.

Other authors speak of an informational society, certainly a more sceptical way of approaching the 'new society' emerging from the global economy. The term tends to be more restricted to Western societies and western economies (Berry, 2008), and like the theoretical concept of Manuel Castells, it refers to a post-fordist era of industry and assumes a shift from industrial to informational economies (a shift from mass production manufacturing and consumption to information economy cored in innovation (Himanen, 2001)) and what Castells (2000a, p. 69) entitles a 'new socio-technical paradigm'. But 'the information society remains a capitalist society, which Castells argues has grown from three major causes; the information technology revolution, the restructuring of capitalism in the 1980s and the long-term effect of social and political movements in the 1960s and 1970s.' (Berry, 2008). Information society and the shift of global economics (UNCTAD, 2017), seem to be increasingly important to understand con-

temporary society's discourses and cultural production, although its reasons are complex and intimately related to the globalization concepts brought to discussion in the above paragraphs.

2.2.2 Coordination and agglomeration

The network society also involves notions of coordination and agglomeration – these will be important in the way we address OA in this dissertation. "The widely accepted notion that agglomeration has become obsolete when global telecommunication advances should allow for maximum dispersal, is only partly correct. It is, I argue, precisely because of the territorial dispersal facilitated by telecommunication advances that agglomeration of centralizing activities has expanded immensely. " (Sassen, 1991). Agglomeration is suggested in the concept of 'global plutocrats' and in the studies of (Piketty & Goldhammer, 2014). While designed to explain the agglutinating mechanisms characteristic of banking and finance around mid 80's, and the emergent logics of power and division of labour inside the global value chain, the concepts of coordination and agglomeration might be used to explain other types of heavily data dependent industries such as is academic research, publishing and increasingly any data related industry – arguably, even politics within restricted scenarios that would explain some neocolonial studies.

2.2.3 Academic Dependency

> *'Despite the growth in scientific production in many peripheral countries, Latin America, Asia and Africa currently contribute less than 20% of the articles published in SSCI. As a result, striving for academic autonomy has been a complex and uphill task for peripheral sociologies, while it is simply taken for granted in American or French Sociology '. (Beigel, 2018)*

As a stream of thought and a field of research, academic dependency finds its foundations on the social studies of science, comparative studies of higher education and critical epistemology (Beigel, 2018). It takes into account the unequal structure of human cultural/economic production and draws an analogy of the latter in the field of knowledge production, management, and ownership. This unequal structure composed by institutional, material and symbolic processes (Beigel, 2018), results from national/regional responses the internationalization (Beigel, 2018) of social, political and economic aspects of human interaction in contemporary society, namely under the context of a Neo-liberal, capitalist form of globalization.

In order to define academic dependency – a condition in which the social sciences of certain countries are conditioned by the development and growth of the social sciences of other countries to which the former is subjected –Alatas (2003) starts by introducing the notion of intellectual imperialism, an analogous phenomenon to political and economic imperialism – 'policy and practice of the political and economic domination of colonial by more advanced nations since

the 16th century through military conquest and subjugation' – i.e. colonialism (Alatas, 2003). Today, intellectual imperialism is viewed as no longer being applied in a direct manner – such as exerting direct control over schools and universities by colonial powers in their colonies – but in such an indirect manner it is possible to find academic dependency in countries that were not colonies during the western-defined historical Era of colonialism (Alatas, 2003). This means that a South inside the North (Sinha-Kerkhoff, Alatas, & Jha, 2010) is emergent, and feeding into alternatives models of knowledge production, globalization and internationalisation. When using the metaphor North/South, or the comparative structure centre/periphery, its is more salient how a system of academic dependency can be laid down on a *dégradé* of influence where works from the centre exert more influence than those in the periphery.

Academic dependency seems to be experienced both at intellectual (or theoretical) levels and at practical levels (Sinha-Kerkhoff et al., 2010). On the practical levels what stands out the most is the abyssal distance between South and South and the practical difficulties experienced in bringing the South together, in comparison to the shallower (even though often great) distances between North and North or even North and South researchers. This is caused for instance by restrictive visa policies among South countries whose international relations are often more limited than for instance European countries.

On the intellectual level, while intellectual imperialism refers to a more general approach on a people's domination of though by another in their world of thinking (see (Alatas S.H. 2000: 24) in (Sinha-Kerkhoff et al., 2010), academic dependency draw attention to the more specific, post-colonial period world of academic knowledge production, above all, in the social sciences. Academic dependency at this level is to be understood in the context of a critique to Eurocentrism and colonialism – according to (Sinha-Kerkhoff et al., 2010), it is in this context that we can understand the core concept of Alatas' subject-object dichotomy, the reflection of an unequal division of object/subject among the international stage of knowledge production[23]. Still at the intellectual level this form of dependency operates at 6 dimensions: (a) dependence on ideas (metatheory, theory, empirical social science and applied social science – the most important dimension of academic dependency (Alatas, 2003)); (b) dependence on the media of ideas (such as books, scientific journals, proceedings of conferences, working papers and electronic publications); (c) dependence on the technology of education; (d) dependence on aid for research and teaching; (e) dependence on investment in education; and (f) dependence of scholars in developing societies on demand in the knowledge powers for their skills (brain-drain).' (Alatas, 2003; Sinha-Kerkhoff et al., 2010). For Alatas (2003) there is also 'a psychological dimension to this dependency whereby the dependent scholar is more a passive recipient of research agenda, methods and ideas from the social science powers'.Yet critical to the structure of academic dependency is the three-fold division of labour: '(a) the division between theoretical and empirical intellectual labour; (b) the division between other country and own country stud-

[23] When non-Europeans do appear in the texts they are invariably the objects of study of European thinkers and not knowing subjects (Sinha-Kerkhoff et al., 2010);

ies; and (c) the division between comparative and single case studies.' (Sinha-Kerkhoff et al., 2010).

2.2.4 Network (de)centralization

In this subsection I will give an initial definition of the concept of (de)centralization to explain network developments in the digital environment (digital economy and ecossystem). The concept tries to be innovative in the sense it attempts to provide a better understanding of network developments on the long run. The subsequent analysis will provide additional ground to sustain and explain the claims that this theoretical concept raises.

(De)centralization is a dichotomy. The core assumption of the concept is that in a network there are two main tendencies – centralization and decentralization. Centralization is always a menace to a network, in the sense it is the contrary to it – an opposite force or tendency. In other words, a network tends to centralization the same way that a strong coordination point tends to create and rely on satelite bodies or points. Centralization might appear cheaper, simpler and more controllable – in brief, power attracts more power. Decentralization might attract smaller concentrations of capital – such as start-ups – create new opportunities that are not capable of mobilizing bigger concentrations of capital and power. Innovation and technological developments are core enablers of decentralization.

What seems to be the case is that if information is power, then information is tending to centralization and not to decentralization. This is core to the rest of the argument of this thesis, because, new technologies, such as Open Source seem to have been proving the opposite of what social theory says about agglomeration and cumulative inequalities. Internet and information technologies have been on the core of some of the global trends in contemporary globalization. But it might be that these technologies are not decentralizing information, but rather they are expanding information and the reach of existing technology, thus also making new sites for extracting economic profit possible. In the same way that economic growth seems to provide greater riches, in the end, capital tends to concentrate and create new forms of appropriation, dominion and inequalities. Seen under this scope, information technology would seem to be accompanying this trend. Expansion of technological means and tools gives the experience of a Network Society, of a network technology and science of information, but in the long-term is expanding and creating yet newer forms of wider and extreme forms of concentration.

Lets consider a network. It seems right to say that a network tends to become more and more centralized if we assume that power tends to concentrate. Information tends to concentrate. Google is powerful because it can centralize everything under one click. In the case of OA a researcher has all articles available to himself. Centralization is a natural process of trade and extraction of profit from a given system, or ecossystem – this tendency is pushed back by genetic evolution and technological development but also the destruction of considerable parts of this networks – natural disasters, calamities, war, death. In this case, the concept is thought of to fit

progressive shifts, such as those caused by technological developments.

Technological revolutions reshuffle the cards of power. Because the powerless have new opportunities to extract power of the new *status quo*. The mainstream channels are clogged and very condensed within themselves and their own *modus operandi* – thus extracting power and capital from new forms of technological capabilities – often in small amounts at high risks – is not profitable for bigger compnaies (this does not exclude that these also make use of high-end technology). Technological revolutions expand the 'mining' sites for power, thus creating new forms of networking – new forms of decentralization – but also new fields of exploration and opportunities for concentrating more power.

Thus the internet, and in this case, OA , create yet new ways of centralization. Whether the internet is centralized or not seems to be discussable but what is central to the argument is that steady centralization of power, capital and information are constant and thinking of this problematic in the long run gives a clear insight to the social issues that will arise from globalization.

2.3 Methodology

In order to better define the role of OA in the current run to information we ask the rhetorical question of whether OA fits a model of dominant globalization or a model for an emancipatory one. In methodological terms we took several steps that can be divided into 3 different natures – interviews, documental analysis and empirical data analysis. In this section we shall lay out the steps that made up this investigation.

2.3.1 Interviews

Although data analysis is very important in contemporary studies, data alone can sometimes give misleading suggestions about the subject in discussion – without a good foundational knowledge of the subject, wrong conclusions might be taken out of context. To avoid this sort of biases, and in exchange gain practical knowledge of a subject that can be considered a niche subject we engaged into a number of open interviews. The interviewees were mainly scholars from western universities: Dr. Paul Ginsparg[24] was interviewed via email communications, Dr. Martin Eve[25] was interviews through video-conference; Dr. David M. Berry[26] and Dr. Judith Townend and Dr. Chris Marsden Law professors at our host university – US – were interviewed in person. Contact was also engaged with the US library staff specialized in OA management.

Regarding my interview with Dr. Paul Ginsparg, the creator of arXiv – the arXiv started more than ten years before the BOAI commonly stated as the genesis of the OA movement – one of the things that stroke me the most was his detachment from other OA realities, in fact he does not talk much about OA in any circumstance, because, in fact, his repository was born long

[24]Cornell University, creator and maintainer of the arXiv.
[25]Birkbeck University, creator of the Open Library of Humanities (OLH);
[26]University of Sussex (US), author of several books on critical theory and digital media;

before the BBBs. His model of OA, though, is what many advocates in early stages of OA aimed at, in the sense that it is economically sustainable and makes use of the technologies available to deliver what people want, on a bottom-up mechanic. The money used to fund the project has the objective of paying for infrastructure, not for paying the artificially created immaterial commodities based on copyrighting – common place in disciplines where OA is being enforced. The arXiv is made up of so-called pre-prints, not journal articles. Articles go first into the arXiv and then go into a journal. Why are there no other venues like this one, is often explained through its dimension (it is so big), and other informal, symbolical mechanisms – the arXiv it has been around since before publishers ever noticed it, so now it seems as if they can't really do anything about it. Dr. Ginsparg himself seemed to me to be quite unaware or at least uninterested by the tension between western universities and big publishers, and even referred to me the essential role of journals in accreditation of academic papers.

My interviews with prof. Judith Townend and prof. Chris Marsden, both from the department of Law, and both sharing the same office, were very much complementary. Our conversation led mostly to a discussion on aspects of Internet Law and Copyright (Edwards & Waelde, 2009) such as as the arguably old 2001 Copyright Directive (Parliament & Council, 2001) and the Digital Millennium Copyright Act (105th Congress, 1998) in the USA, as well as the then newly proposed EU's Article 13[27] (Europe, 2017). On this context we discussed some major legal battles, mainly on the media industry, such as the "Napster" case, "Pirate Bay", Viacom (against Youtube) and also realized how the 'news media' industry, in so many ways was going through similar challenges as the scholarly publishing industry. Regarding the OA movement, we discussed Aaron Swarz, and the Elsevier Boycott. We also discussed some OA initiatives in law environment (such the africanlii.org, austlii.org, etc.). Their knowledge on Internet law was deeper than in OA particularly, which they advocated for, even though prof. Judith was part of an editing board. Regarding OA it seemed they were participants more than cold observers.

Dr. David Berry had a somewhat similar background, in fact I found his book in the law section of the library. As with Dr. Townend and Dr. Marsden, Dr. Berry had a very developed notion of copyright. Besides this he also had a very deep knowledge on Software Copyright and IPR. I asked him for a critical analysis of the whole OA scene, which I was still lacking at the time, and he offered me his holistic point of view on copyright and OA. For him the political-economy[28] of OA (and knowledge economy at large) is the key for a critical analysis of the subject. Under this scope it is possible to transpose some concepts such as control, power, agglomeration and expulsion (or exclusion) into the subject of OA. For him it is very clear that OA is being pulled under the condition that a market still exists for publishers. From the point of view of profit making, scholarly publishers did not suffer any radical change[29]. He stressed the capacity of this model to perpetuate Southern economies into a consumers model,

[27] See `https://edri.org/about/`

[28] As he said 'you need to follow the money';

[29] In fact the literature review corroborated that notion;

concentrating production of creative and informational goods in the North. We also spoke about shadow libraries, a concept (and a cultural scene) that is very important to conceptualise the Sci-hub[30].

My interview with Martin Eve was very interesting. Because he is really specialized in OA we dove deep into specific points such as peer-reviewing, Journals/Repositories differences, STEM and HASS particularities (mostly explained by informal mechanisms), and also about pre-prints such as the case of the (low) adoption of pre-prints by researchers in the area of the humanities[31] and how OA in the context of books or monographs can be important in these areas of academia. We also discussed some details on the OLH[32], and the new economic model that it puts at test.

My interview with Hellen Webb from the Sussex's Library, was very much an appendix to my interview with prof. Eve, in the sense that she was the coordinator of OA in the library of Sussex, and she new Dr. Eve's job quite closely and was quite fond of it. This interview was good to get a feeling of what is going on inside libraries, i.e. between researchers, publishers and the university. Webb was responsible for the OA management as I was able to understand[33], she didn't share information with me the same way as the scholars did. The job of the OA department is to archive published articles online, contacting and giving counsel with the researcher and negotiating with the publisher. Funded by RCUK, this venture seems to deal with quite a lot of money, mostly paying publishers to make articles OA.

Besides a general feeling and sensibility to all these matters the main ideas I withdrew from these interviews were thus that **1)** the explanations based on symbolical, informal mechanisms, tend to naturalise problems and hardly lead to solutions. This challenged the use of Bourdieu's notion of 'distinction' and 'symbolic capital' as part of the theoretical framework. **2)** At least in the UK context, there seems to be reasons to believe that mandatory OA, is helping APCs going up because there is a huge detachment of the actors in the trading and negotiation process – the university mandates an OA regime, the researcher will do it, but on the venue that is most likely to benefit his career. The library will ask the RCUK for the money to make that article OA, and the publisher will decide how much he wants (negotiating with the library's Research Support, that is spending money that legitimises and empowers the latter). On the other hand spending on subscription based journals never ceased to increase (as Hellen Webb told me and some reports show, with some suggesting that we might be reaching a turning point.). So publishers are indeed double-dipping, or at least Hybrid OA, as is most commonly stated, is impossible to keep track of. Besides this seems a reinforcing economic cycle that benefits the big publishers capable of a) buying prestigious journals (where researchers want to publish) and b) controlling all sources of income at the same time (i.e. good coordination facilities) and explains why it is so difficult for

[30] `https://www.reddit.com/r/Scholar/comments/3bs1rm/meta_the_libgenscihub_thread_howtos_updates_and/`

[31] Single authorship culture;

[32] See `https://www.openlibhums.org/`

[33] She was from 'research support';

small journals and publishers to start-up. The last conclusion follows: when will public money stop going into the hands of publishers as proposed by the 'advocates' of OA? The answer seems be 'never'. Governments are indeed paying more. OA has turned into a new source of expense. This makes sense only if we think of it as a process in which corporations (the industry) have lots of bargaining power near the political power and democracy is not making its way through. We have seen this happen in the auto industry, we have seen this in finance. And this is happening in the scholarly publishing industry too.

2.3.2 Empirical Data

Digital Object Identifier

The DOI is an identifier used mainly for scholarly articles (similar to International Standard Book Number (ISBN)s). It is an identification stamp managed by the International DOI Foundation whose objective is to provide a persistent, unique, standardized identification code to all scholarly articles. The benefits of its usage are mainly managerial for instance providing the object's associated metadata, enabling better coverage for data management. DOI is a centralized attempt to universal standardization of scholarly articles although it is most commonly found in STEM areas than HASS. It is also possible to find DOI's in other sorts of scientific publications such as datasets, book reviews and even public sector information. The DOI is composed of a prefix and a suffix. While the first identifies the registrant of the identifier, such as the publisher, the second identifies the object associated with the DOI. Here we provide as an illustration, the most common article in our dataset: 10.1080/01419870.2016.1199889

The Sci-Hub sample characterization

Partially due to its apparent capacity to "democratize knowledge", but also because of its apparent capacity to generate new forms of agglomeration, we analyse this morphology of OA withdrawn from unregulated OA to the light of a tension between different players.

We chose the Sci-Hub's statistics to act as a proxy measure to what a broader OA market might look like. Besides the economic and pragmatic aspect of its availability, one of the key aspects for choosing this dataset for this purpose has to do with the disaffiliation of the Sci-Hub service to any institution, discipline or country. This is not entirely true though, because the Sci-Hub seems to rely on DOIs to find articles, thus Sci-Hub, to some degree is related to the institution issuing DOIs, and since these are more used in the natural sciences than the social sciences and humanities, probably more used by publishers in the global North than in countries of the South, and thus more used in English written articles, it is quite safe to say that this dataset over-represents STEM disciplines, English language articles and western works. These three elements – DOIs being more used in STEM, in countries of the North and in English based articles, are all cumulative factors, interrelated and applying pressure in the same direction, thus

accentuating one another.

These elements are part of what critical theory defines as imperialism, they tend to aggregate and total knowledge, repelling what is outside its scope. But is this a form of bias or is it a reflection of this central, totalitarian power? Indeed one might say that the map in Figure 2.4 represents the map of a gold OA. Another factor to this is the fact that services such as SciELO, mostly represent humanities and non-English based articles, and since these are OA, and commonly unprovided with a DOI, it is a tendency that they do not appear in Sci-Hub's channels. Another way to put this is despite (Himmelstein et al., 2018) claims that Sci-Hub provides access to nearly all scholarly literature this is only partially true – only when taking into account the mainstream literature. It would be more accurate to say that the Sci-Hub provides access to nearly all of paywalled STEM literature (and most STEM research is made in the North); especially the more requested one, the most recent publications. In other words, it would be fair to say that the Si-Hub provides free access to all the mainstream hits of the publishing industry – much like 'Napster' and the 'Pirate Bay' used to. This seems to be a feature of this kind of service, for back in its days, these also did not give excellent access to remote cultural artefacts, these were more prone to channel the hype, new and profitable cultural goods.

This perspective sets a distance between the Sci-Hub and other shadow libraries. Technically, both the Sci-Hub and the 'Pirate Bay' are capable of eventually hosting this kind of rare cultural good, but in practice, its main function, or the main 'magic' behind both these sharing systems is to provide access to something that lots of people want. The conclusion to this is that the there is more to scholarly publishing than there meets the eye, namely in the form of minorities and plurality, which are constantly under the radar of broader research. Not all articles need to be relevant worldwide, not all articles must be written in English – not all articles need to be digital. So this dataset would never be capable to reconcile all of the multiplicity and diversity of scholarly publishing in itself, which in a way would be the aim for this research, but that also, is taken *a priori* as an impossibility.

This being said, the conclusion follows that the Sci-Hub has created a space for itself in the global market of academic publishing, with it bringing the disruptive power of data related technologies, and its revolutionizing powers. In the context of this investigation, the Sci-Hub as a data source is also the more indicated, for it offers the more immediate and wide spectrum of analysis available online. The next subsection is going to tackle the more technical issues behind data analysis.

Data management

The data Analysis was made on Python, making especial use of the Pandas module to deal with datasets; numeric plotting was made using Matplotlib and Seaborn; all maps were created thanks to the Cartopy technology. City names were retrieved using the services of OpenStreetMap Foundation. All other country-specific data was retrieved from the World Bank institutional site. All datasets are publicly available at their site of retrieval indicated throughout the document.

All the analysis is publicly available at `joao-bernardes/scihub/`, may be used, re-used and shared. These subsections will lay the data structure used for this study[34].

In January the 18th, 2018, the Sci-Hub announced through its twitter account it was publishing its 'download log' (SciHub, 2018a) for the year of 2017. We downloaded the gzip file containing the database through `https://zenodo.org/record/1158301#.W39Ub7gnZhE'` (Tzovaras, 2018) for convenience. The file contained 8 columns (Timestamp; DOI; IP identifier; User identifier; Country according to GeoIP; City according to GeoIP; Latitude; Longitude) per 150875862 rows. According to online information (SciHub, 2017) we are assuming that each row represents the occurrence of a resolved download[35]. Since all DOIs with a prefix '10.1016/j[36]' (SciHub, 2018b) were omitted in the 18th of January 'download log', the Sci-Hub published a new dataset containing the missing values (another 43623865 rows), which we added to the first dataset obtaining a total number of 194499727 rows[37]. From the total number of rows we created a random sample, providing a 99% of confidence interval[38] in analysis over the raw material.

OA status of articles

The second step involved getting to know the prevalence of OA in our sample, for this we downloaded the full Unpaywall dataset (January 25th update)[39]. The dataset came with several columns: we only used two, the DOI and the OA status, in order to be able to assess how many of the articles contained in our database were already openly licensed. For this we first divided Unpaywall dataset into 30 parts (so a normal laptopt could deal with the large size of the Unpaywall dataset) and rebuilt the dataset with only two columns - the article's DOI and its OA status. We then crossed the DOI's contained in both datasets and plotted the results (Figure 2.2).

98.72% of Sci-hub's Statistics DOIs were present in the Unpaywall Dataset. The Unpaywall considers an articles is OA when it is 'free to read online, either on the publisher website or in an OA repository' (Piwowar et al., 2018), the authors consider this a conservative approach, since it excludes all Gratis OA and accepts only licensed OA – either Gold, Green or Hybrid. According to this definition the percentage of OA articles downloaded through the Sci-Hub is 16.67%, quite bellow the 28% estimated OA ration for all DOI articles for the same year.

The gap between values reflects the fact that the Sci-Hub users rely on the service in order to circumvent paywalls in scholarly literature. According to Greshake (2017) they also tend to

[34]Please visit `https://github.com/joao-bernardes/scihub/blob/master/src/2017.Statistics/BUCKET_2017.statistics` for schematic information on analysis structure

[35]We attempted to establish contact with the Sci-Hub team to clarify this sort of issue, but always without success. The information available online is quite clear although not 100% unambiguous

[36]The prefix for Elsevier;

[37]In February the Sci-Hub had announced on twitter that *circa* one million papers were missing. These were not recovered;

[38]Using a Sample Size Calculator at: `https://www.surveysystem.com/sscalc.html`

[39]https://github.com/joao-bernardes/scihub/blob/master/Release%20notes%20for%20Unpaywall%20dataset.pdf

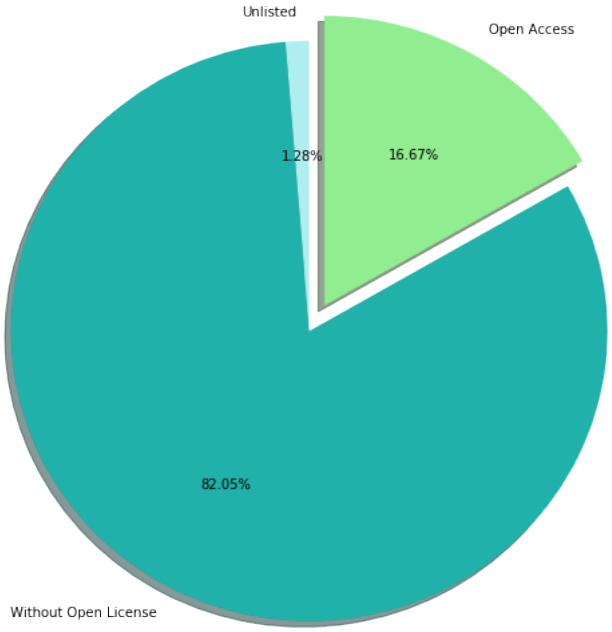


Figure 2.2: Prevalence of Open Access Articles in Sci-Hub's 2017 Dataset. (By comparison with Unpaywall data)

access mostly recent literature. Himmelstein et al. (2018) calculate that the Sci-Hub repository holds 49.1% of articles in OA journals listed in the Crossref (i.e. gold OA), if we compare the two (i.e. the existence of an article in repository, vs the frequency with each it is requested by users) we confirm that tendency of users to rely on the Sci-Hub to retrieve TA literature. Gold OA articles are rarely requested by Sci-Hub users, assuming that they use the service for circumventing paywalls, but also for matters of convenience. The Sci-Hub is not a search engine (Alexandra Elbakyan, 2017). Users are required to insert a very specific information about the article they are seeking – since usually Gold OA, available at the publishers' venue, are easily findable by search engines such as Google, it is expected that users of literature will not use Sci-Hub in these situations.

Since Figure 2.2 represents the frequency of occurrence of a request of a Libre OA relative to TA/Gratis OA, this also reveals that users are slightly more interested in accessing TA articles than OA ones when using the Sci-Hub, or that they require a centralized venue in order to find and get access to this kind of literature. An explanation to this also resides in the fact that many of the Sci-Hub's articles are recent articles, and thus still under embargo periods, although on the other hand, more recent articles tend to be published under OA. The Sci-Hub log has more representativeness of recently published articles (Greshake, 2017), whose OA status' ratio is even higher than average (28% being the 'conservative' estimate, and around 50% the less conservative) so we can see that Sci-Hub is being requested quite intensively to access paywalled articles, since it rests on 16.67% of OA articles. On the other hand if we analyse the disposition

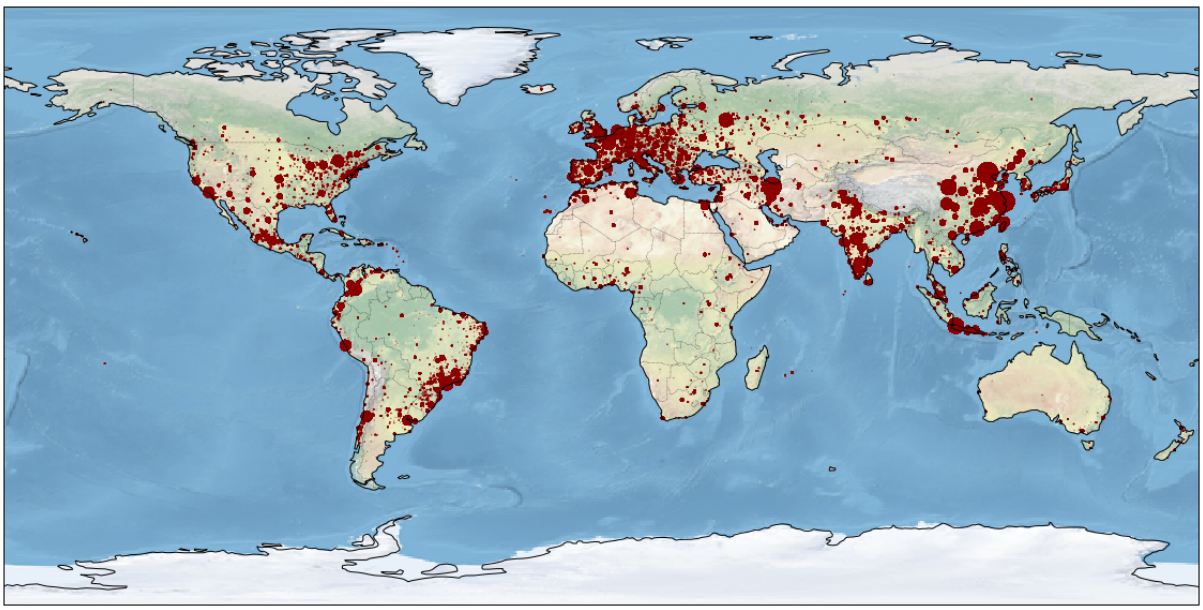


Figure 2.3: Sci-Hub Only Open Access Consumption

of articles in our sample that are OA against those that are not, we can see that there is no intense disparity between the Figures 2.3 and 2.4 .

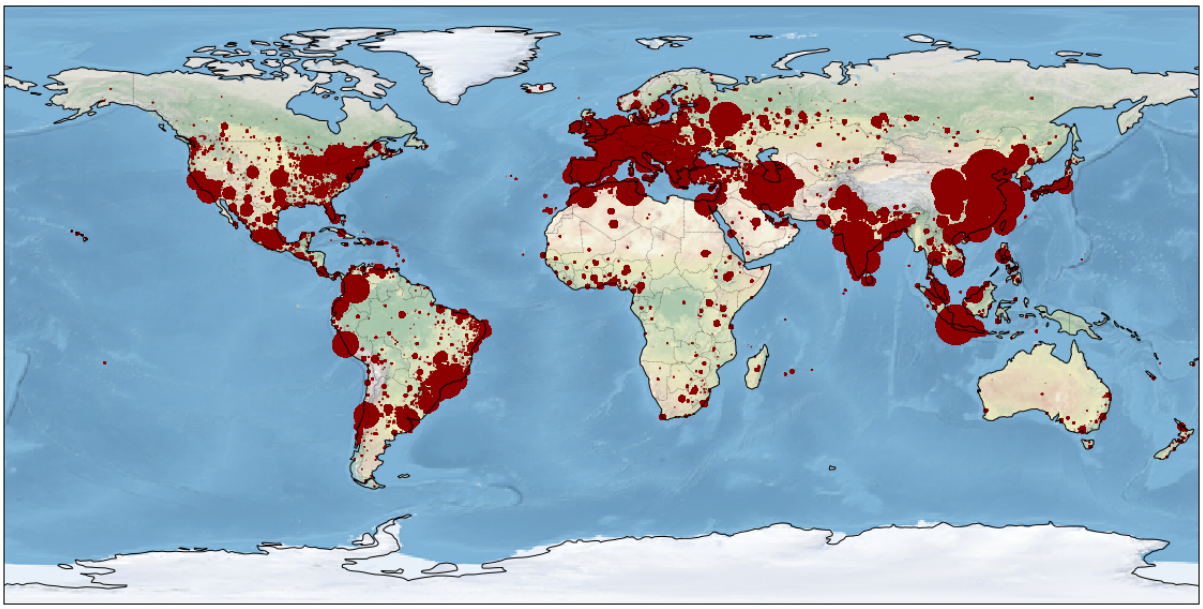


Figure 2.4: Sci-Hub Only Toll Access Consumption

Using Outsource Data

The nature of this investigation is different from most studies on either the Sci-Hub or Open Access, mostly because we address article consumption numbers for world regions instead of drawing our focus on the detailed study of the western-centric production of articles, or on the

data put available by the Sci-Hub in order to assess on the uptake of pirate article consumption. Nevertheless, aspects related with the origin of articles were not simply ignored, and while most detailed analysis in these matters is to be presented in a subsequent study – for more on this see future work section in Chapter 4 – there is also a broad approach in the morphology and prevalence of publishers in the uptake of OA in the subsequent analysis section. To make this possible we extracted the prefix of the DOIs contained in the Sci-Hub Statistics and merged it with a list of publishers names[40].

Lastly we compared the number of downloads with other socio-economic indicators for several countries. The data was extracted from the World Bank official repositories[41] as well as from The United Nations Educational Scientific and Cultural Organization (UNESCO)[42]. The indicators used were Gross Domestic Product (GDP), Internet Access and Research & Development (R&D) Investment, which were treated and correlated with the main dataframe in order to explain the determinant factors that influenced the download counts.

In order to study downoads at the city level we relied on the Maxmind service[43] to provide population numbers.

Caveats

Some draw-backs have already been presented in this section of the Chapter. The dependence of the whole analysis in a core database provided by a single person makes problematic some consistency aspects. Open technologies tend to depend on the participation of a community in order to check for defaults and make sure everything works smoothly. The Sci-Hub does not have a community and its software is not open-source, which makes it more vulnerable to problems – for instance in the quality of data. For instance, the database was released without a clear name for each column, leaving the process of naming them to third-parties, unrelated with the process of extracting the data.

That takes us to the second problem – the extraction of data is not controlled by a multiplicity of eyes – a corner stone in the foundation of the quality of open-source technology. Several times during the year there were announcements of issues with the data – one of which announced that all articles with a prefix belonging to Elsevier had not been taken under consideration. This had a huge impact on the overall shape of the data, and while this seems odd, it is impossible to rectify what exactly happened that excluded those records and how exactly they were recuperated.

Lastly, a major problem is that it is not possible to understand whether the origin of a download request is really from the point the record points to or if otherwise the user was using rerouting technology such as Virtual Private Network (VPN) that mask users real location and offers a mirror instead – although Bohannon (2016) claims that 'according to Elbakyan, fewer

[40] available at `https://gist.github.com/TomDemeranville/8699224`
[41] `https://data.worldbank.org/`
[42] `http://data.uis.unesco.org/`
[43] `https://www.maxmind.com/en/home`

than 3% of Sci-Hub users are using those' it is not possible to verify this for sure – or other kinds of download behaviour that do not match the 'expected' *download to read, use or re-use*.

Another caveat relates with the date this dataset is relative to, for, although making it more interesting to be working with such recent data (the year of 2017, while this study is written in 2018) it also makes it very hard to find relatable data. For this, many of the statistical correlations were drawn with country-relative indicators from years other than 2017 (We used 2014 and 2015 data).

Nevertheless, still, this dataset provides an astonishing opportunity to learn more about scholarly publishing, the state of knowledge production and open access morphology – all this without the requirement of a huge infrastructure of data gathering.

The next Chapter will include the presentation and analysis of the empirical data discussed in this chapter.

Chapter 3

Analysis

In this chapter we will analyse the reasons that make OA an increasingly popular tool for scholarly communications; what, why and how different morphologies of OA derive from its use. We will do so by giving particular emphasis to the relationship that researchers, countries and markets have with technological innovations, namely OA, by first defining a board of players and a set of dichotomies we understand influence the logics, rules, tensions and dynamics occurring between them (see Chapter 1). The analysis is run and tested alongside empirical data retrieved from the Sci-Hub service, which we analysed as a proxy to what consumptions of OA look like globally.[1]

In this investigation we wish extrapolate some of the conclusions that we withdraw from the Sci-Hub Statistics Log of 2017 to the wider universe of OA in scholarly publishing. The ultimate goal of this investigation is to understand whether, OA is contributing to a scholarly communications industry that promotes a dominant form of Globalization or the opposite – an emancipatory one. For this we introduced the concept of (de)centralization, a concept that proposes that in a network social system, two opposing forces are at work – centralization of power and capital, and decentralization, particularly caused by the disruptive and power of technological advancements in industry [2].

3.0.1 Results

Validity of Analysis

In order to be able to assess whether or not our dataset is suited to provided insight into the ecosystem of scholarly publication and knowledge production we propose the hypothesis that what affects the amount of articles downloaded by each country is the amount of investment on R&D. This hypothesis would prove that what affects the use of the Sci-Hub online service

[1] What the Sci-Hub is, its relevance for this study, its limitations and caveats, is discussed in more detail in the methodology section and should be kept in mind during the reading of this section although not being essential in order to understand it;

[2] See Schumpeterean destruction;

is not the lack of investment of universities to provide researchers with journal subscriptions or general access to literature, but simply that higher piracy levels in scholarly publication per country, reflect a wider participation in R&D and a greater 'effort' of countries to invest in knowledge production. In other words, by proving our hypothesis accurate (that the amount of illegal downloads relate with the amount invested in R&D) we open up the possibility to extrapolate conclusions withdrawn from the Sci-Hub Statistics Log of 2017 to the wider universe of Scholarly Publishing.

We set out to understand what factors would affect the amount of downloads made by users in for each country. We assumed that the factors that would affect mostly the behaviour of users would be the effort each would put into academic research. For this we correlated the number of downloads per country with the amount spent on R&D in absolute terms and in a percentage of the country's GDP as well as each country's total population.

Since the population number was affecting strongly the amount of downloads[3], we calculated its *per capita* values.

The amount spent on R&D, population, and number of downloads were highly inter-related among the three, probably due to the presence of outliers such as China and Brazil, but mostly China. This suggests that countries limits are concepts poorly suited for this analysis, or at least that China requires an internal sub-division in order to not distort other values. Looking at the results, one possible interpretation would be that countries with a lot of population tend to spend more in total than countries with less population[4]. Even so, countries that spend more in R&D, such as China, Germany, Japan, Brazil and France, tend to do more downloads (or should we call them *readings*)[5]. Finally if we take a look at the test results of correlating the number of downloads and population, the results reinforce the blatant relation between the two[6].

If we take a look at the spending *per capita* or in percentage of GDP the distribution of values changes quite a bit and the correlations between variables loose a lot of intensity. For instance, the amount countries invest in R&D *per capita* correlates at 0.34[7] with the amount of downloads *per capita*. Another meaningful value arises from the correlation between the number of downloads and the effort a country invests in research – R&D in percentage of GDP. While scoring quite low when using the number of downloads in absolute terms, the result becomes particularly compelling when the number of downloads is normalized to number of inhabitants of a country – suppressing China as an outlier – and giving rise to a steady distribution[8]. Lastly, all correlations presenting a negative direction (i.e. that the amount spent on R&D would lower the amount of illegal downloads) always displayed extremely low R^2 values floating on 0,05, 0,02 and 0,01.

[3] $R^2 = 0,92$;
[4] Corr = 0,92, $R^2 = 0,86$, p-value < 0,05;
[5] $R^2 = 0.81$, correlation = 0,89 and a p-value <0,01;
[6] Corr = 0,97, $R^2 = 0,93$, p-value < 0,05;
[7] With a p-value < 0,01 and $R^2 = 0.42$:
[8] Corr = 0,24, $R^2 = 0,42$, p-value < 0,01;

We conclude from this analysis that while we were not able to distance ourselves absolutely from the factor 'number of inhabitants' of countries –and that China as an outlier was influencing our results in this matter – the usage of the Sci-Hub as a mechanism to circumvent paywalls is not tendentially used in countries where R&D investment is lower. On the contrary, those countries that invest the most in research and science are those who tend to be using the Sci-Hub more often.

We open thus the possibility to accept the subsequent analysis as reflecting wide trends in scholarly publishing and not only in 'shadowed' activities.

Inactivity - reasons for low OA consumption

In continuation with the previous results obtained by comparing amount of downloads and R&D investment by country, we measured how GDP interrelated with number of downloads. The results show that when correlating GDP with the number of downloads, we get a strong correlation[9]. Thus, this reinforces the notion that richer countries with better access conditions also have higher prevalence of usage of the Sci-Hub service.

Access to the internet seems to be among the capital conditions to access scholarly literature through the Sci-Hub – since without internet, there is no access. This might not be an absolute true though if we take into consideration what Bohannon (2016) says regarding Iran high access rates. He points out the possibility of access rates in those regions being even higher, given that there may be many servers mirroring the Sci-Hub. These mirrors could be channelled locally, not necessarily using the internet. The correlation between internet availability and downloads was proven to be very low. In fact, looking into Fig 3.1 and 3.2 it is possible to see that a low percentage of population accessing the internet will not necessarily translate into an absence of requests. This might be explained by the fact that most users of the Sci-Hub services – scholars – are already part of a small fraction of population under the influence of several factors that make it possible for them to access the internet. For instance, while Guinea-Bissau, Somalia and Eritrea have the lowest rate of Internet Access according to the World Bank data[10], they still are represented in our sample – although the most requests are concentrated in a single city in each country.

Besides, accessing scholarly literature is something that does not require great infrastructure or internet speed. Usually full-text 'PDFs' weight no more than 2 Mb. Despite OA advocates' initial claims that scholarly literature could be of the interest of the several publics other than scholars, this seems to be doubtful. Mainly, scholarly literature seems to be accessed by scholars alone. The benefits to the general public would be indirect, since the benefits of a more informed academic structure would be likely to overflow to wider spheres of society. Especially in southern countries.

[9] Corr = 0,91, R^2 = 0,84, p-value < 0,01;

[10] see `https://github.com/joao-bernardes/scihub/blob/master/src/outsrc/Internet_Access.ipynb`

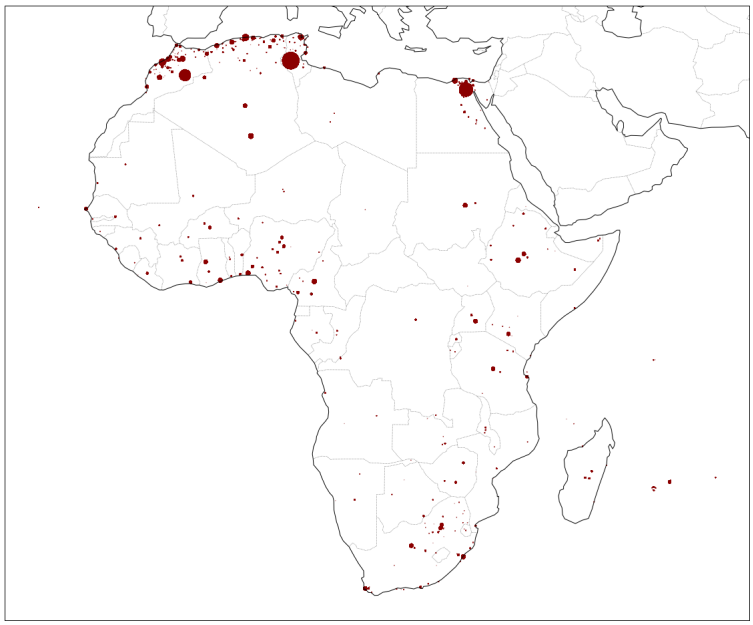


Figure 3.1: Africa: Map of OA Consumption

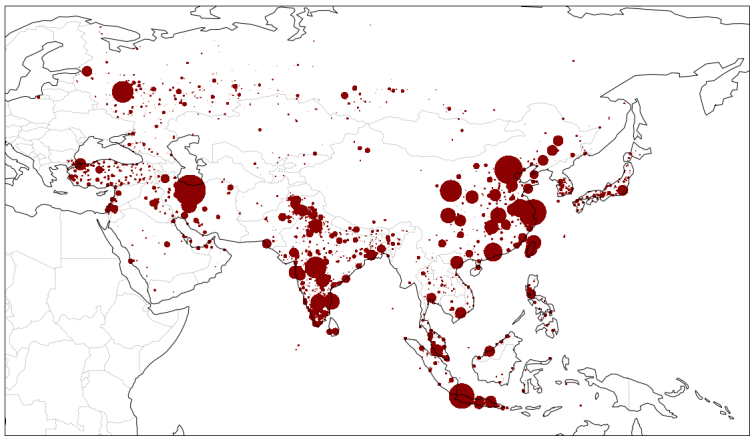


Figure 3.2: Asia: Map of OA Consumption

For southern countries it seems that OA could be benefiting the speed of access, but not access itself. For instance, in the African case, Florence Piron (2018), Piron et al. (2017) speaks of 'several cognitive injustices' that students have to endure that do not relate with internet or TA paywalls. In fact, it is known, that since early days of scholarly publishing in the post-colonial era, publishers have traditionally 'offered special packages' that include both ease of access to some journals both from the reader and the author point of view.

So OA might no be the answer to the questions that most sub-Saharan Africa is looking for. And this might reflect the tension of the movement against commercial publishers. The TA model might not be hindering sub-Saharan countries as much as other forms of injustice and inequality that OA might be helping to replicate. Answering this dilemma is the macro objective of our investigation.

Looking above at Fig. 3.1 and below at Fig. 3.3 it becomes clear that there is a territorial expression of OA use – a territorialization that emphasises an Eurocentric morphology. There seem to be two main poles – Northern Africa and Western Sub-Saharan Africa. Below, Fig. 3.3 stresses the gap between northern African countries and the rest of the continent. The African countries with the larger amounts of downloads are Egypt, Algeria, Tunisia and Morocco. While the first four are all North African countries located in the coast of the Mediterranean sea, facing Europe, Nigeria, notoriously less active than the prior, is at the centre of a second pole, in Western Africa.

Although analysis at city level may seem to be more accurate than at country level, given the high levels of concentration observed in the geographical disposition of our data, in fact there are some practical difficulties to provide meaningful results at such a detailed level. For instance, some rows in our database are missing coordinates values, or a city name. This might translate in that some important focus points are absent from some tables, maps or plots.

According to table 3.1(a), the African city concentrating the highest amounts of requests of scholarly literature is Cairo (Egypt) – although, according to Fig. 3.1, the greatest number of downloads is actually coming from the region of Chott el Djerid, Tunisia – while Morocco has three different cities (Tadighoust, Casa Blanca and Fes) in the top ten African Cities – showing a more dispersed activity in comparison with Egypt.

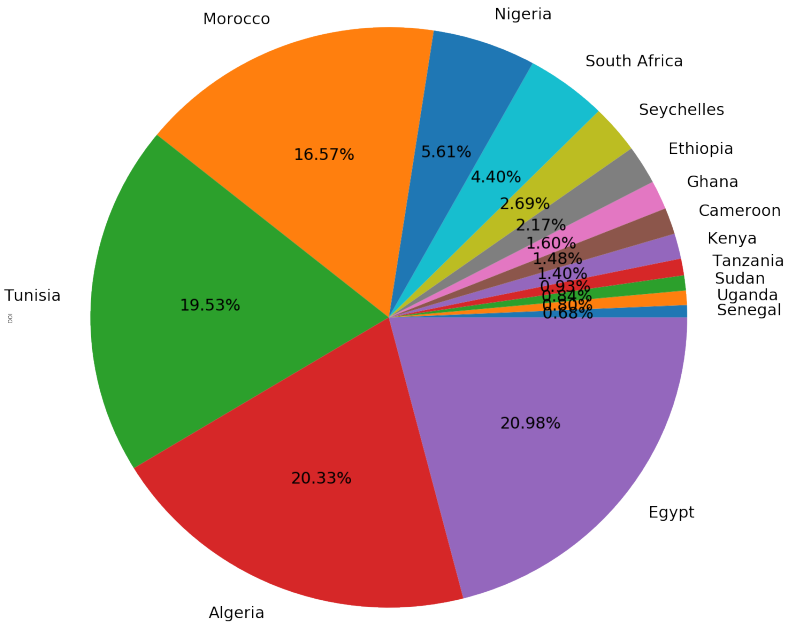


Figure 3.3: Africa: Internal Ratio

On a more curious note, the Seychelles island, as can be seen in Table 3.1 (b) has shown a tremendous amount of downloads in comparison with its population size. In fact when assessing the number of downloads *per capita*, Fig. 3.4, we realize that the Seychelles island has had more than 7 times the amount of downloads *per capita* than the countries next to it – Singapore and Luxembourg. The latter two belonging to a different group of countries, with high internet access rates, higher GDP and R&D investment. By plotting a time log of those downloads we realize that much of it is done between the period of about 4 months (between April and August) – we realize that most of them are also requested by the same IPID – finally, none of these have coordinates (thus not represented in the maps – would be in fifth position in Table 3.1[11]). This phenomenon might be explained by the use of VPN services, the percentage of population with internet access in the Island is about 55%, not far from, for instance, Morocco, although as we have seen, internet access does not really relate with article consumption.

So, what exactly are the reasons that may be blocking regions from accessing scholarly articles? As we can infer from the first block of this section, population highly influences the amount of downloads. We can expect no activity in the Saharan desert, Antarctica, etc. But what other factors might influence OA article consumption. Should we think of scholarly publishing as a form of economic activity, and taken into consideration that, as we have seen, globalization is very much characterized by conflict, one factor that might influence article consumption, might be the existence of local conflicts in certain areas – war.

The morphology of consumption under war circumstances, it seems, reduces the general

[11] see `https://github.com/joao-bernardes/scihub/blob/master/src/outsrc/World_pop.ipynb`

country	city	Downloads
Egypt	ةنيدم رصن	7735
Morocco	تسوغيدات	5245
	Casablanca	1654
Algeria	Algiers	1639
Tunisia	Tunis	1043
Egypt	ىديس ربإج	972
Morocco	Fes	953
Algeria	نيع لقمأ	896
Ethiopia	Ezhana Welene	765
Cameroon	Mbam-et-Kim	748

country	Downloads
Egypt	15249
Algeria	14776
Tunisia	14197
Morocco	12046
Nigeria	4076
South Africa	3199
Seychelles	1956
Ethiopia	1574
Ghana	1163
Cameroon	1073

Table 3.1: Africa: a) Top Active City/Country | b) Top Active Countries

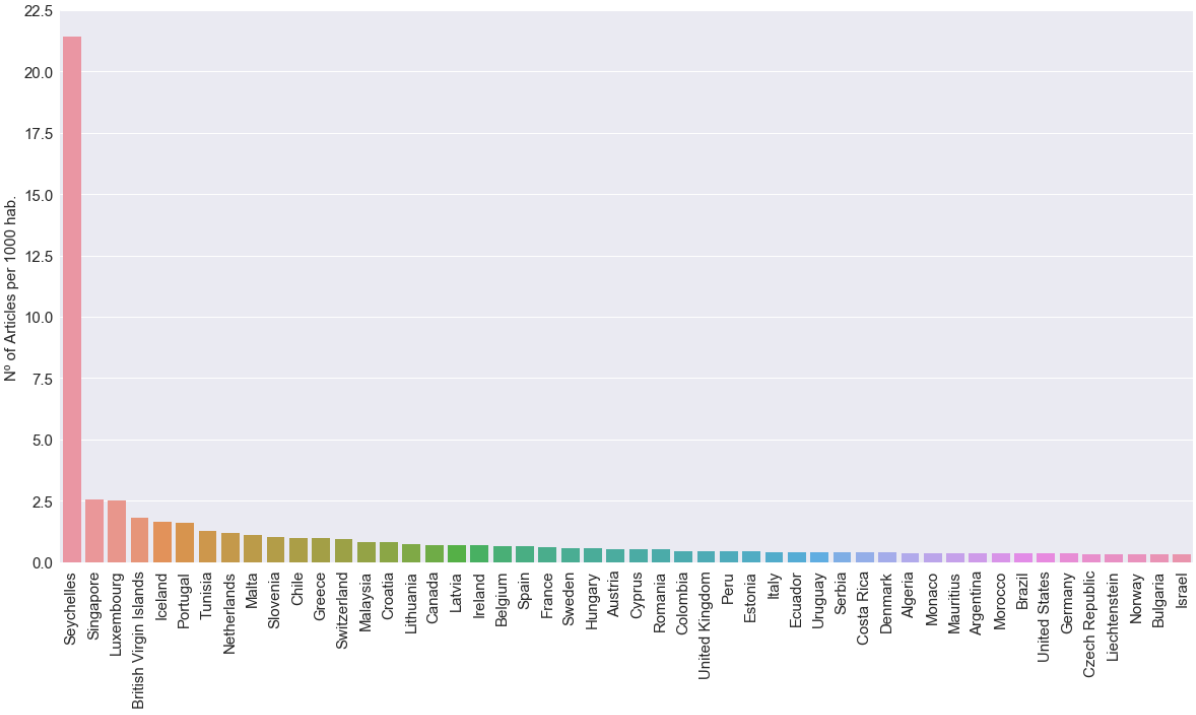


Figure 3.4: World: Top OA consumption per capita

activity of a country and tends to polarize consumptions into a single city, creating more or less heavy contrasts between warzones and neighbouring countries. For instance, in the African case, the only country scoring zero downloads was Central African Republic, a country in civil war since 2012 and still heavily affected by conflicts and foreign occupation. But countries such as Algeria and Egypt are also involved in conflict, mostly at the southern borders, and while this reduces activity in those borders it may reinforce download activity in the war-ridden areas.

The Middle East scenarios seem to reflect some of these tendencies. See for instance Syrian territory, in the midst of a fairly active region. Neighbouring countries such as Israel, Jordan, Lebanon, Iraq and Turkey all show high consumptions of scholarly articles *per capita*[12], revealing a somewhat high R&D activity near borders of warzones.

On a more holistic analysis of the contrasts resulting from war, conflict and the intrinsic inequalities brought by the globalized political-economy we could also relate some of this inactivity with the higher activity levels witnessed in other countries potentially involved in those conflicts, not only directly but indirectly. So for instance, in the Syrian and Yemenite[13][14] case, while some of the above cited countries may be selling arms and ammunition or even contributing with troops on sight of the conflict, arms and ammunition is understood to come from Eastern European countries originally. The core idea to retain here is that this hierarchy of power may be traced to critical points of exploitation where actual participation in economic activity is, in zero sum game perspective, zero. While these appear to be excluded from the Global Value Chain (GVC), they can be in fact be understood as mining and extraction regions, natural resources to be explored. In this perspective it becomes clear that 'expulsions' are not a mere description of a non-participation, but a participation in non-equal terms with other actors, or players.

Proceeding our analysis into the reasons that might block regions' activities, and in continuation with the arguments proposed by Collyer (2018), we turn our attention to Oceania, more particularly Australia, in order to understand some reasons that might shape the low article consumptions witnessed in this area. Australia does seems to show Southern characteristics in terms of participation in scholarly literature – low article consumption in absolute and *per capita* terms, heavily concentrated in specific littoral regions that suggest a borderline presence in an external ecosystem of economic activity. This drives us to take into consideration the intrinsic territorial characteristics of global economic activity. Oceania reveals extremely low levels of consumption of articles, especially when taking into consideration the growing participation of the western Austronesian regions with countries such as Indonesia or Malaysia (See Figs. 3.2 and 3.5). In the Australian case, downloads are spread throughout a rather large number of different cities which is quite particular to this country, looking into Table 3.2 we see that in absolute terms, there are 12 cities among the top 15 cities requesting articles.

[12] See Fig. 3.11

[13] See `https://www.theguardian.com/world/2016/jul/27/weapons-flowing-eastern-europe-middle-east-revealed-arms-trade-syria`;

[14] `https://www.bbc.com/news/world-middle-east-22906965`;

Another interesting finding is that the region with the most downloads *per capita* is New Caledonia, a French territory comprised of several islands. New Caledonia's downloads do not seem to be related with VPN services – unlike in the Seychelles scenario, there is nothing odd about its records and the existence of the University of New Caledonia in Noumea suggests that these requests are coming mostly from students from that university.

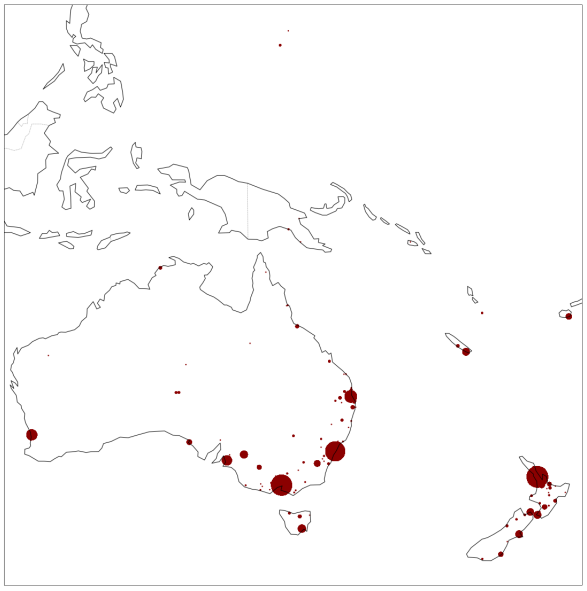


Figure 3.5: Oceania: Map of OA consumption

country	city	Downloads
Australia	Melbourne	620
New Zealand	Auckland	602
Australia	Sydney	558
	Alexandria	400
	Hornsby	317
	Brisbane City	189
	Adelaide	176
	Perth	163
	Brighton East	85
	Hobart	78
	Monash	77
	Barangaroo	74
	Clayton	73
New Caledonia	Noumea	66
New Zealand	Christchurch	63

Table 3.2: Oceania: Top Active Cities

Inactivity - Review of Results

Next we are going to inquire about some activities in the Latin America (LA), but first let's review some of our results so far.

Article consumptions in the regions analysed up until this point are characterized by being generally low, concentrated in big cities and territorially integrated in the periphery of external ecosystems of economic activity. This explains, for instance the existence of higher levels or article consumption in Northern African countries, connected to Europe by sea. As a rule of thumb, Sub-saharan Africa appears to be excluded from this activity.

Conflict can be interpreted in our data but does not translate into the pure absence of activity, instead, shaped like a tornado, the focal point of the conflict shows extremely low values of article consumption, while the neighbouring region reveal fairly large levels of consumption, mostly concentrated in big cities. This leads us to think of expulsions as a form of hierarchical exploitation in which the last link in the chain does not participate in equal terms with other players.

What would be expected, if OA was being used as a tool for diversification of scholarly publications, would be that countries from peripheral areas such as those in Oceania would be benefiting from OA. Therefore we would be expecting a slightly higher consumption of literature, especially since Australia and New Zealand are English speaking countries, and most of the articles in our database are expected to be written in English. In fact, it seems that most African countries are facing way deeper epistemological issues (Piron et al., 2017), and mostly, only counties with english or french roots (mostly inherited from colonialism) will be able to benefit from it.

Latin America

The LA case is quite different from the previous and quite particular in many aspects. While in most African territory, it seems, access obstacles are caused by deeper social shortages and lack of physical infrastructures, the adoption of OA in much of LA can be understood as something radical. Radical in the sense that the model has been adopted at governmental level, giving rise to effective alternative channels for scholarly publications (see for instance Scientific Electronic Library Online (SciELO) and Latindex). Unlike in Africa, most Latin American countries are not comprised by traditional and meagre publishers' 'developing countries special packages' that allow special prices for access and publications in countries unable to pay the usual fees. On the other hand, due to heavy investments in higher education during the past few years (Castells, 2014) in some Latin American countries, the technological development among economic activities and population in these regions is high and thriving. Finally, without the the roadblocks caused by the tensions of the conflict between major commercial publishers and national governments and funders in the regions of the Global North, countries such as Chile and Brazil end up having a pronounced expression in the map of OA not only in LA terms, but at a global level

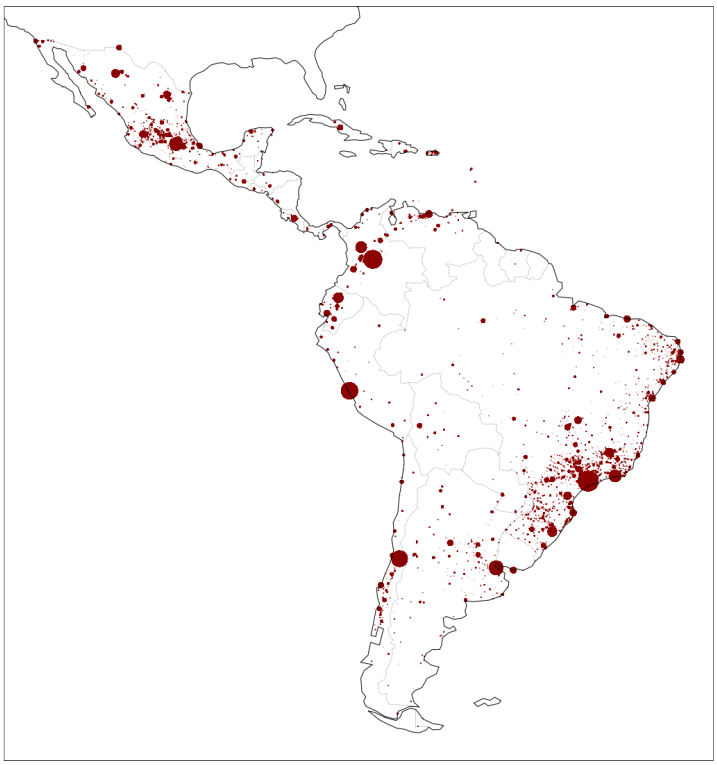


Figure 3.6: Latin America: Map of OA consumption

too (See Fig. 3.6).

An issue that might be hard to track in our database is the absence of non-DOI based articles flowing inside LA countries. The existence of such articles is particularly compelling in LA, due to the fact that, as stated above, OA has been adopted in this region radically and created a 'pathway for the distribution of academic knowledge inside the global South' (Collyer, 2018). The consequence of this is that by creating a parallel channel of communication, and a decentralization from hegemonic channels of communication, it becomes harder to track these activities, most of them, under full-fledged OA. The problem here was stated above in the Methodology section. The DOI as an identifier serves the purpose of helping to keep track of articles, in order for studies such as these to be able to performed without requiring huge investments of time and man-power. To keep track of all SciELO article flows would require an investigation of its own. The real reasons for many LA publications lacking a DOI is mostly a mystery, nevertheless it has been aknowledged as a rule of thumb that the adoption of the DOI would be something beneficial.

This highlights some of the limitations of our database, for it reinforces centralization, it emphasizes hegemony, and activities of the global North. In fact, the more processed the data gets, the more the results will be likely to reinforce this centralization. This process of under-representation of variety is both symptom and consequence of what we wish to study here, and is likely to be affecting several other studies in these matters.

Among South American countries, Brazil is responsible for *circa* 47% of all article con-

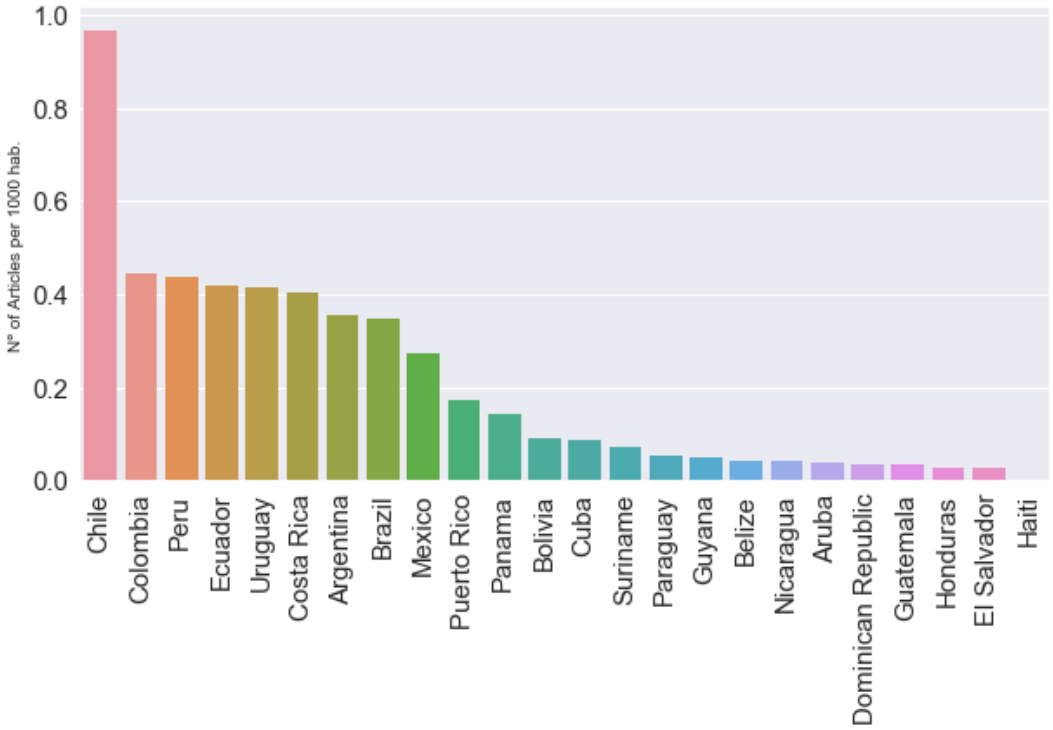


Figure 3.7: Latin America: Top OA consumption per capita

sumption (38% in the LA context), on the other hand, in population terms, Brazil represents more than half of the continent's population (30% in the LA context), which makes it a natural outlier. When calculating *per capita* values, the results are also very in line to what would be expected (See Fig. 3.7). Chile scores first, reflecting its dynamic policy, the so-called 'Miracle of Chile' quite ahead of neighbouring countries. What is striking to see, though, are the sharp stage formations in the graph. First we see a huge gap between Chile (0.9 articles per 1000 inhabitants) and the 8 following countries – Colombia, Peru, Ecuador, Uruguay, Costa Rica, Argentina, Brazil and Mexico where downloads per 1000 inhabitants stay in between the 0,4. This 'second platform' is followed by a third one whose values stay below generally bellow the 0,1 (similar to those witnessed in many countries in Africa).

Since the case of Brazil seemed so relevant in the context of South and Latin America we studied in more dept the Brazilian case and what we saw was that most downloads were coming from the mega city of "São Paulo" (Fig. 3.8). Fig. 3.9 reveals a progressive slope in the context of cities, with South American countries at the top, such as Colombia, Chile, Peru and Argentina, reinforcing the idea that the analysis at the city level is often the more appropriate in comparison with the 'country' measure. The idea of global cities, and concentration is patent in the analysis of this chart.

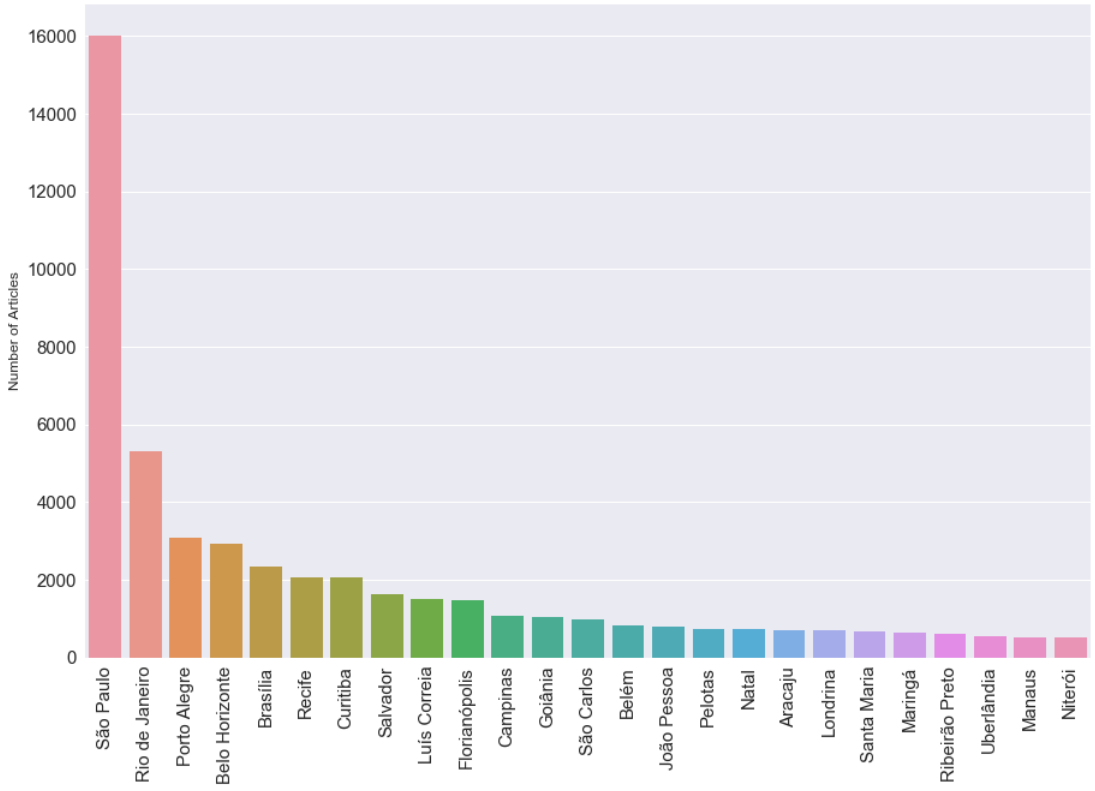


Figure 3.8: Brazil: Top OA City consumption

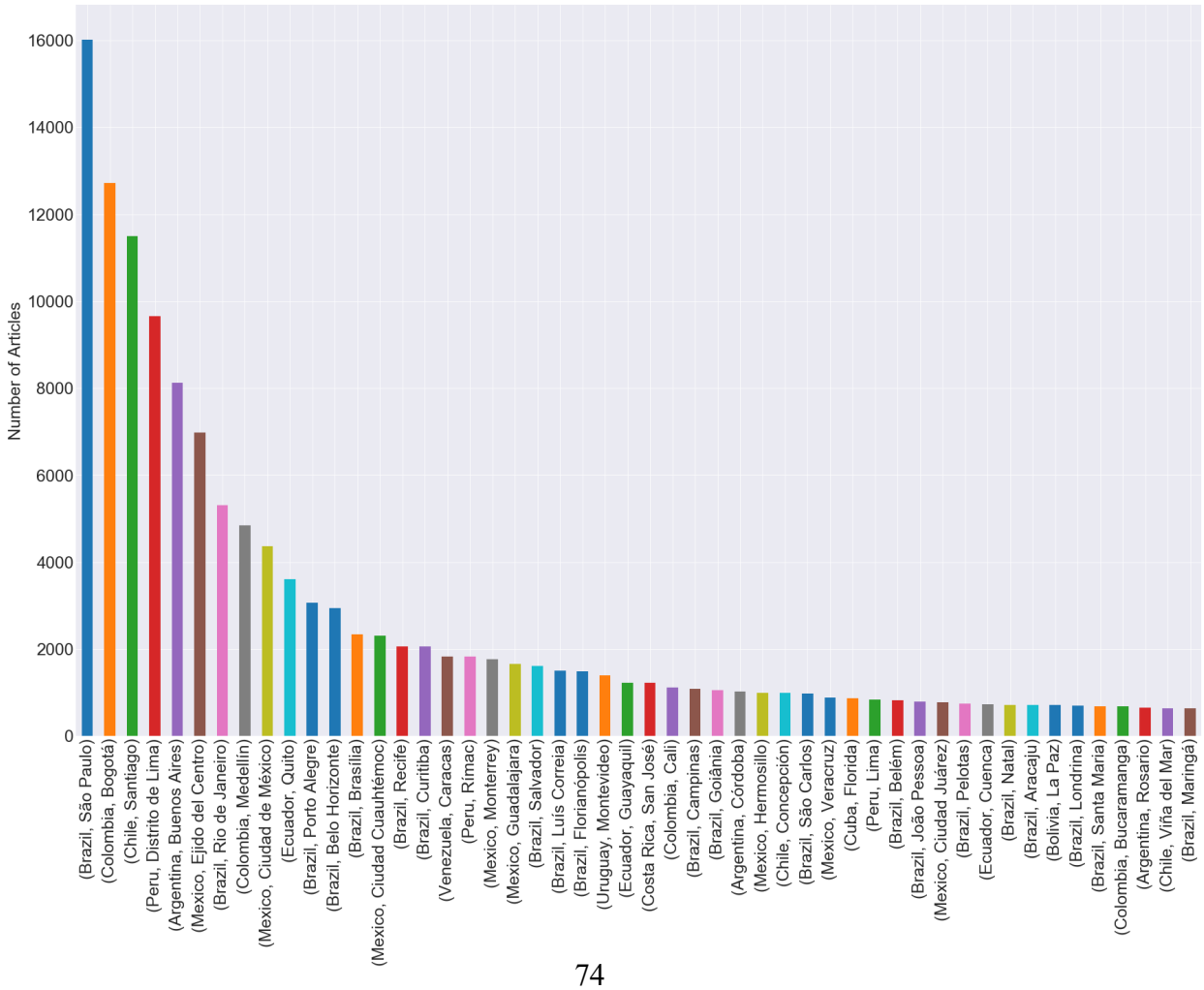
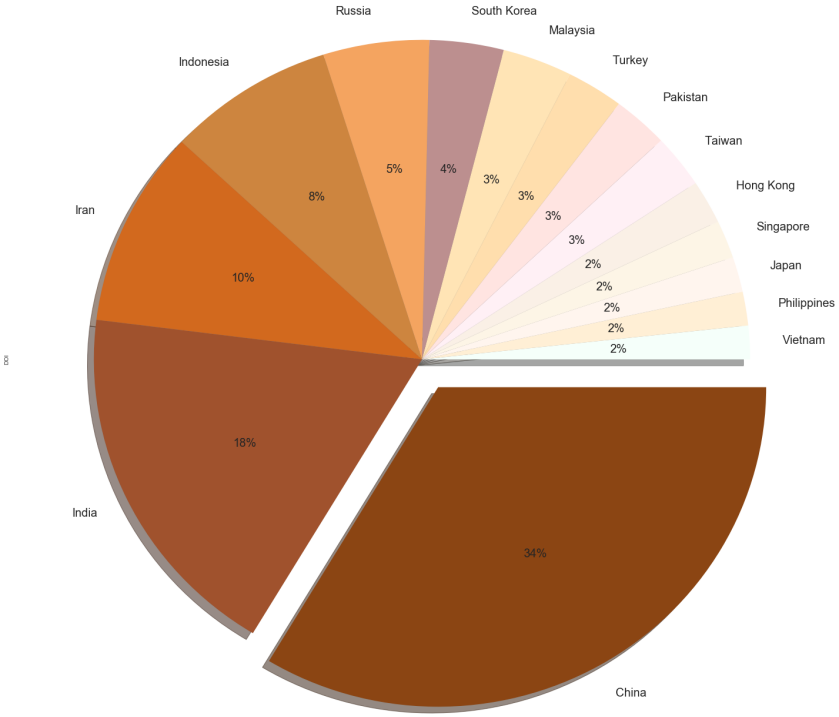


Figure 3.9: Latin America: Top Country/City OA Consumption



H

Figure 3.10: Asia: Top Country Consumption

East and South East Asia Big Economies

This analysis does not intend to cover comprehensibly the situation of each country, but give an interpretation to some of the most notorious findings in our data that might help us understand the role of OA in the context of a global chain of knowledge production an to understand whether OA might induce the production process into a route of concentration. With this goal in mind we turn our attention to some of the biggest Asian economies – many which have a pronounced representation worldwide.

To analyse South East Asia some caveats most be taken into consideration. Since we used the world bank database in order to assess on the population of each country, Taiwan is not included in all *per capita* analysis, since it is included in China mainland – the same applies to Hong Kong. This explains why Japan, Singapore and South Korea are right at the top of the Fig. 3.11, with the exception of Taiwan and Hong Kong.

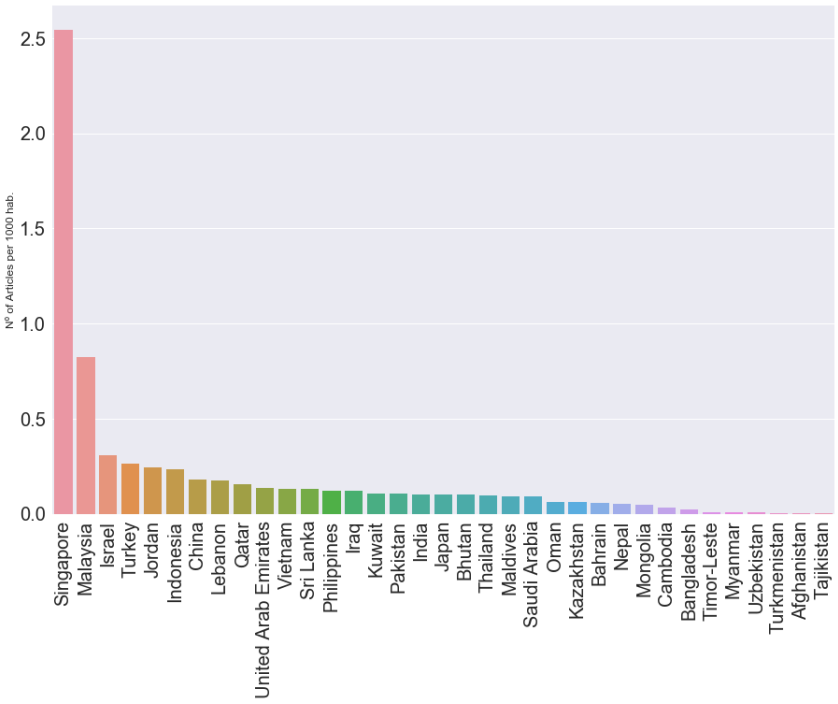


Figure 3.11: Asia: Top OA Consumption per capita

On a rough attempt, if we divide the number of downloads by population number it is possible to estimate the ratio of downloads *per capita* in Taiwan (0.83) and Hong Kong (2.17). Hong Kong thus jumps up to second place in the Asian context (see Fig. 3.11) and third in world level (Table 3.3), while Taiwan rests third behind Singapore in the Asia context and 15th in world terms.

China and India comprehend more than half of Asia activities 3.10, followed by Iran, Indonesia, Russia, South Korea and Malaysia. In the Iranian case, 81% of requests are coming from Tehran (66%) an Tiran (15%). This is important to note, since Iran is a very big country and economy, but according to our data, what we really are witnessing is a concentration of economic activity in a single city.

Scholarly publishing in these regions is characterized by being mostly STEM related, contrary to most of the OA movement in Latin America that is mostly characterized by flowing HASS related literature. Most research is being produced by countries in the North and also most literature is english based, most literature is also written in alphabetic characters – this shapes the production of knowledge. Article consumption in most regions of Asia, namely the most populated, is intrinsically different from European or North American consumptions. The prospects of use of OA are also different from those of the Latin America.

Japan, the so-called Four Asian Tigers, as well as the Tiger Cub Economies, and China are countries whose economies are increasingly important both at regional level and global level, affecting other countries economies both at international and national level. These regions also cradle intense anthropological and social diversity. The map of OA is not one of diversity though, it is a representation of a field of economic and political dominance. In this context

country	Downloads_1000hb
Singapore	2.543821
Luxembourg	2.521934
Hong Kong (estimate)	2.170000
British Virgin Islands	1.825064
Iceland	1.637211
Portugal	1.620892
Tunisia	1.273970
Netherlands	1.202490
Malta	1.104571
Slovenia	1.036867
Chile	0.965550
Greece	0.965167
Switzerland	0.927503
Malaysia	0.823508
Taiwan (estimate)	0.830000
Croatia	0.815168
Lithuania	0.737629
Canada	0.710194
Latvia	0.704189
Ireland	0.675435
Belgium	0.657861
Spain	0.655022
France	0.612958
Sweden	0.567238
Hungary	0.557342
Austria	0.546315

Table 3.3: World: Downloads *per capita*

it makes sense to give particular emphasis to the role of India, an extremely diverse country, also carrying extreme forms of poverty and inequality, hosting a cheap source of labour.

Could OA help democratize the voices of diversity in these regions is hard to tell, there seems to be no apparent reason for it to be the case, at least in the current scenario of a dominant proliferation of western-centric literature. The interdependency on financial and economic factors seems to reinforce Northern epistemologies and dominance. In the case of China, for instance, lack of civil rights such as freedom of speech or a free internet also seem to be blocking the path for a diversification of knowledge – in this sense, OA is being used as part of a wider imperial plan and only makes sense as a natural shift of economic and industrial production in the sense that Schumpeter described the continuous reconstruction and redesign of models of production.

In other words, *'ceteris paribus'*, for the high R&D investing Asian countries such as China or Japan (See Table 3.4), OA seems to be a model to which these countries will need to adapt. While from the access point of view they might benefit a drop of costs, from the point of view of production and publication this might reflect into high expenses with mandatory Gold OA APCs. While European countries already have a budget on publication, southern countries will be inhibited from producing original knowledge, and passive participation is likely to be promoted.

country	2015 GERD (billions of US$PPP)
China	408.257710
Japan	170.003023
Germany	114.778152
France	62.313712
Brazil	41.104129
Italy	30.466534
Canada	26.250891
Spain	19.734812
Netherlands	16.991344
Turkey	16.604485

Table 3.4: Top R&D Investors

The landscape of OA in these regions might be beneficial to these countries since it provides the tools to enhance economic activity through, for instance, access to pharmaceutical and health related literature, or energy and climate related technology and help boost the speed of research. For instance in India, if the country invests in an alternative communication system, based on Green OA, ie, pre-prints, manuscripts, in order to create South to South connections, OA might turn into a beneficial tool since it could provide mechanisms for the enhancement of economic activity. On the other hand, mandatory OA could translate into a inhibition of participation in South to North, or South to South communications.

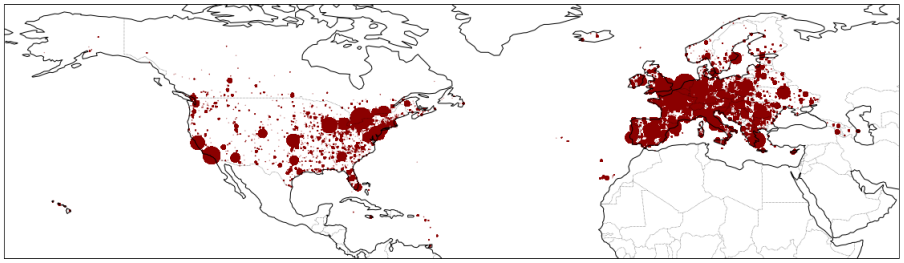


Figure 3.12: Global North: OA Activity

Russia's activities seem to reveal the strong territorial influence of the West, because of Moscow's strong activities, but mostly due to the absence of activity in the Eastern regions of Russia.

The Global North

To analyse the statistics in the regions of the North (Fig. 3.12) we mixed North American countries that were not part of the LA with European countries in order to avoid the sharp disparities between the United States and the rest of the continent. Besides, since we did not engage into State level analysis for the case of the United States, studying North America alone would practically mean to study the United States and Canada alone, and that would not give us much useful information. Gardner et al. (2017) make a State level analysis of the United States case on a similar data referent to a different time window.

When compared in absolute terms, the number of downloads coming from the United States is 2 to 3 times higher than the following countries. This reveals a certain inappropriation of the comparison since the United states have a much larger population size. Further down the list we observe that countries no longer appear in population size order, since most of these are located in Eastern Europe, such as Poland and Ukraine. Instead these fall the so called semi-peripheral, *buffer* countries such as Spain, Portugal and Greece.

	country	Downloads
65	Luxembourg	1403
126	British Virgin Islands	54
86	Iceland	536
18	Portugal	16859
14	Netherlands	20280

Table 3.5: Global North: Downloads sorted by values *per capita*

We proceed with an analysis of *per capita* values. According to Fig. 3.13 the Luxembourg scores the highest, followed by the British Virgin Islands, Iceland and Portugal, but the relevance of each varies quite a lot (see Table 3.5). The British Virgin Islands count a total of

only 54 downloads, and Luxembourg (1403 downloads) concentrates most of its downloads in Schoos (1038). Unlike most similar situations there are no universities in Schoos' whereabouts that would explain such high records (such as Luís Correia in the Brazilian case), also, ration between downloads and IP addresses is extremely low (0.12 in Luxembourg, against 0.55 sample ratio[15]. There is also the remote possibility of data mining research experiments being pulled out in remote regions of Luxembourg, so we look for signs pointing in that direction. In fact, when looking into the timestamp of each request performed by the top downloading IP in Luxembourg we realize that these are being done with few minutes or even seconds of interval among each other – within 3 months – also there is a consistency among DOI prefixes – the IPs are downloading physics material. This might be suggesting that although automatic, these downloads are being performed by the same user or group of users. Nevertheless this is not the case for any other IP groups which means that we are still getting suspicious activity coming from Schoos, Luxembourg. It is thus highly likely that the user or group of users making use of the Sci-Hub during the 3 months time-period were using a proxy or VPN service, consequently being located in a implausible city (Schoos) while maintaining the same IP address, and that the other IP addresses were also related with similar types of identity masking.

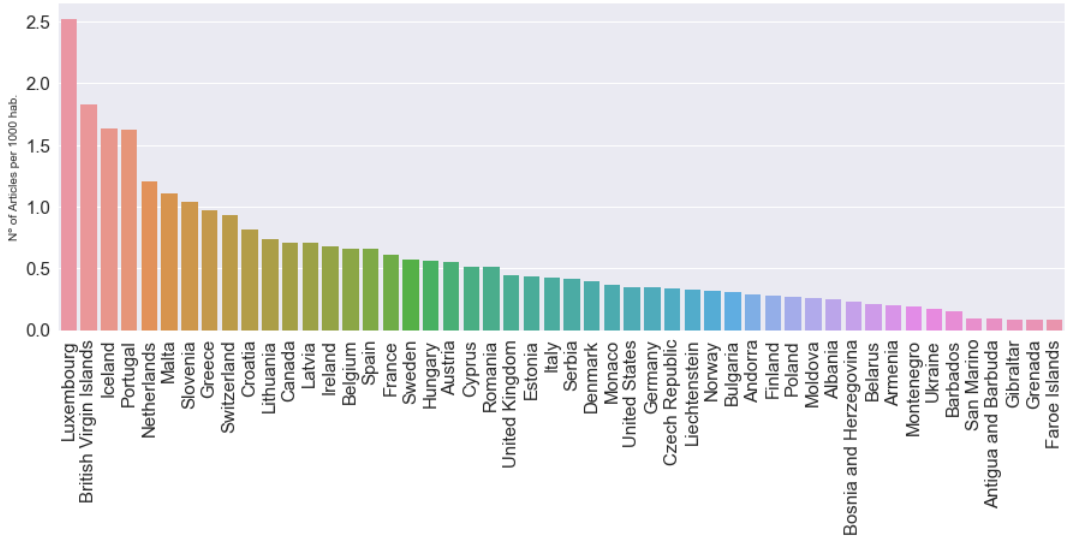


Figure 3.13: Global North: Top Downloads per capita

It does not seem to be that the users were using a script to perform downloads on several articles, as would be expected in larger scale, professional big data experiments since there is a lack of consistency in the timestamp relative to each download. It would be interesting to study the database more thoroughly with the objective of detecting large scale, script based, bundles of download, in order to inquire some big data and data mining related research.

Continuing with our analysis it is worth noting that the distribution of countries along Fig. 3.13 reveals a blatant absence of the bigger economies in the North. This is explained by the

[15]0.04 in Schoos;

country	city	Downloads
France	Paris	16248
United Kingdom	London	15684
Netherlands	Amsterdam	13971
Canada	Toronto	13563
United States	Los Angeles	8962
Spain	Madrid	8665
Greece	Athina	7379
Portugal	Lisboa	6946
United States	Chicago	6735
Germany	Haina (Kloster)	5973
United States	San Jose	5427
	Wilmington	5403
	Buffalo	5038
Spain	Barcelona	4854
United States	Potwin	4643

Table 3.6: Global North: Download per City

fact that the data reflects OA consumptions, ie. either because bigger economies do not need to rely in the Sci-Hub to access scholarly literature, or because semi-peripheral countries are more dependent on accessing literature in order to give certification and credibility to their own work. The second half of this explanation is especially important, for it suggests the existence of an *academic dependency* inside the North – a process in which researchers are dependent on other countries' research to be able to produce 'good science'.

Lastly, by taking a look at consumptions by city (Table 3.6 we are able to observe that differently from what we saw in Country *per capita* distribution, the main economies of the North are represented (with Toronto, Canada, appearing for the first time in top 5), among semi-peripheral countries. This suggests that cities and countries in the Northern group have different behaviours. There seem to be two possible explanations for this landscape. One possible explanation aligned with the argument that researchers in richer countries do not need to access scholarly literature through parallel channels of dissemination. Accordingly it seems that the data would be revealing that big cities in rich countries tend to create the opportunity for a black market anyway. Either because of business or smaller lower funding universities that appear next to big research poles such as is the canonical case of the MIT, or simply because researchers find it easier to use the Sci-Hub in specific situations.

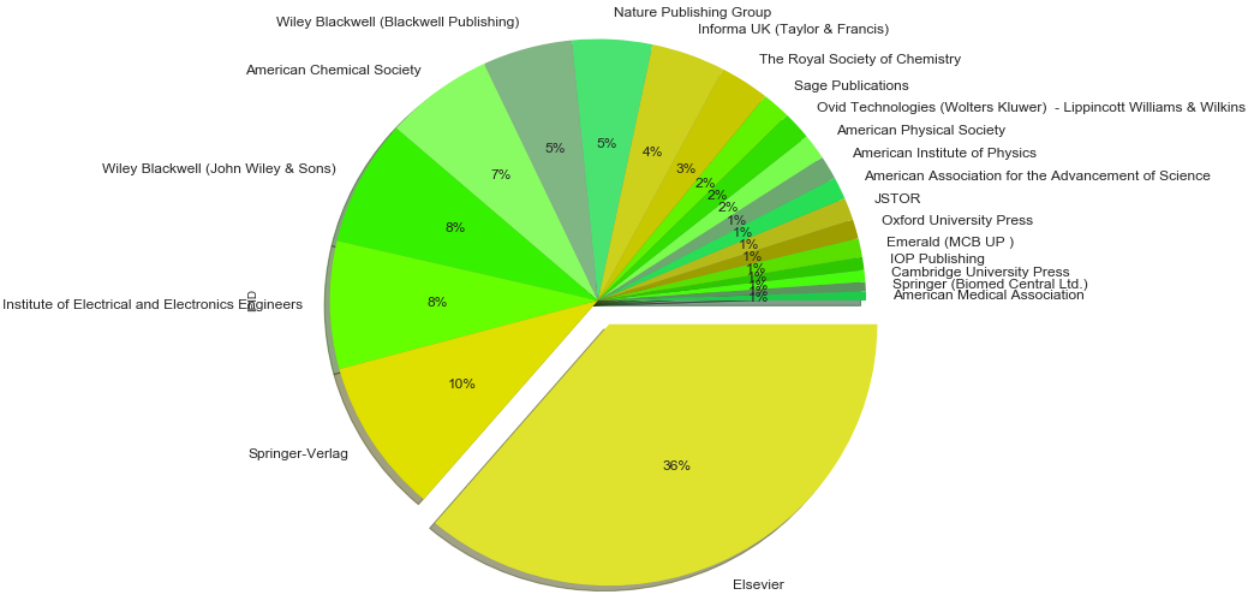


Figure 3.14: Academic Publishers Market Share

To speak of OA in tho global North is to speak of an ongoing battle between the stakeholders on knowledge production, namely social movements of researchers and librarians, public funding bodies, such as universities and governmental funding agencies, and lastly, the scholarly publishers and the particular interests of established business. Attempting to understand whether OA is part of a hegemonic form of globalization and scientific production we turn our attention to the representation of scholarly publishers in our Sci-Hub database (See Fig. 3.14).

Fig. 3.14 plots the percentage of articles shared between the top publishers. An analysis focused on the disposition of publishers in our database reveals a sharp tendency to oligopolistic and Pareto distribution. Situated at the top we also observe the presence of both for-profit corporations and not-for profit professional associations and institutes. About half of the market is owned by three major publishers: Springer, Wiley and Elsevier. In third position, IEEE, The Institute of Electrical and Electronics Engineers, is a not-for-profit professional organization based in New Jersey:

> *'IEEE depends on conferences and publications for three-quarters of its revenues. Both face serious competition. For the future health of IEEE, we must diversify our revenue sources.'* [16]

While during the year of 2009 it reached US $340 million, according to the Wikipedia, by 2018, its revenue would be of US $413 million revealing a huge growth in general business volume.

[16] `http://theinstitute.ieee.org/members/presidentscolumn/ieee-is-also-a-business502`

'Revenues, expenses, and profits are basic metrics for any business. But for IEEE, our success also is measured by how well we adhere to our core values. The most important of the values is service to humanity—which has been imbued in IEEE since the inception 125 years ago of the American Institute of Electrical Engineers, one of IEEE's predecessor societies. That's why in this 125th anniversary year, I initiated the IEEE Presidents' Change the World Competition to recognize students who develop unique solutions to humanitarian or community problems.' [17]

'IEEE and its members inspire a global community to innovate for a better tomorrow through its more than 423,000 members in over 160 countries, and its highly cited publications, conferences, technology standards, and professional and educational activities. IEEE is the trusted "voice" for engineering, computing, and technology information around the globe.' [18]

ACS, American Chemistry Society, 5th, is also a non-profit society that specializes in chemistry, 'with total consolidated revenues of $569 million'[19] by the year of 2017. 6th, is a publishing enterprise owning publishers such as 'Taylor and Francis' and 'Ashgate Publishing', for instance, characteristic for being created in 1998:

'Informa operates in a growing, global market for professional, commercial and academic knowledge and information. Our aim is to generate sustainable value for shareholders and maintain positive, long-term relationships with customers, suppliers and our communities. We do this by harnessing the ideas and contributions of our colleagues, our culture, brands, customer and partner relationships, financial capital and infrastructure to deliver products, services and outcomes that customers value.'[20]

Wolters Kluer, in 7th, is based in the Netherlands, annual revenue of US $4.3 billion, and is also extending business into Asian territory[21]. It's activities are to:

[17] http://theinstitute.ieee.org/members/presidents-column/ieee-is-also-abusiness502
[18] https://www.ieee.org/about/index.html
[19] https://www.acs.org/content/annualreport/en/executive-director-s-message.html
[20] https://informa.com/about-us/group-strategy/
[21] https://www.publishersweekly.com/pw/by-topic/industry-news/financial-reporting/article/58357-globalpublishing-leaders-2013-wolters-kluwer.html

'(...)provide information, software, and services that deliver vital insights, intelligent tools, and the guidance of subject-matter experts.'[22]

It is important to note from this quick analysis on how the top publishers is populated by businesses claiming to provide global informational services, also some of them are non-profit organizations experiencing acute growth levels. This tends to confirm the theory of agglomeration and coordination in global businesses. While the core focus of the for-profit corporations is to generate value for shareholders, not-for-profit entities' is not so clear although they do show the same core predatory tendencies of growing in order to deal with intense competition.

Another central and puzzling idea is the fact that all of these entities, especially the corporative ones, are based in several western countries, such as the US, the UK, Netherlands and Germany. These are also the countries whose governments and libraries are battling these big corporations, giving a hint that these enterprises, although located (often hard to locate precisely in a single place) in a certain country, are detrimental to the State/government ecosystem in which they are inserted.

Another tendency that may be noticed is the fact that these entities appear to be recovering from the 2005/2008 financial crisis, which coincides with much of the early developments of the institutionalization of OA. OA posed a threat in a problematic conjecture but seems to fade and be absorbed by established market tendencies as is suggested in Aspesi and Luong (2014). This relation between governments, funding agencies and the publishing industries *status quo* is puzzling because more than rejecting oligopoly, these actions are furthering and reshuffling a new way of publishing that might enhance the tendency for agglomeration.

It its pertinent to stress how the majority of these private corporations are expanding their market to 'Asian-pacific' regions, where political regimes are Neo-liberal, business friendly, also where major R&D investment is being made and where publishing market is still in a rather unexplored state.

Lastly, in order to grasp the size of academic publishing in a global market, we ask *how big these enterprises are*. In terms of comparison, while Relx (Elsevier's parent company) is $44.77 Billion, Nature Springer is $6.679 Billion, Apple is about $900 Billion, Spotify is about $30 Billion. In terms of revenue during the year of 2017, Relx had a total revenue of $7.35 billion, $2.48 billions of which belong to Elsevier, ie. roughly 30% of total revenue of the parent organization. Wolter Kueler (during 2016) had a $4.3 billion revenue, while IEEE's revenue was of almost half a billion same as ACS (both not-for profit organizations), $569 million, which makes them small in comparison with the others, but not so small, and in fact, giants in terms of non-profit organizations.

[22] `https://www.cchifirm.com.au/other/about-wolters-kluwer`

Conclusion

Open movements in general are tools that are based upon the ICT – they make use of tools such as the Internet to build a corpus of knowledge and technology that is solely or mainly based on *free*[23] material. The core characteristic of this free material is that is stays immune to privatization and commercialization, free artefacts are simply not proprietary. Being immune to commercialization does not mean it cannot be commercialized at all, but simply not directly, through privatization and copyright impositions. *copyleft* symbolizes this spoof on copyright, patenting and proprietary IPR. The power of the concept, boosted by the ICT, combined with legal tools (open licenses), has the capacity to at least shake the waters of contemporary informational production and its in its resilience – the immunity to privatization, etc. – that lies its ultimate fascinating feature, for open movements can easily sprout from individual action and grow into complex ecosystems of cultural production.

Information is becoming increasingly important in contemporary industrial production (Castells, 2000a, 2000b), which brings the issue of open movements to the front of various economic and social battles. The case of the OA movement is not an exception to this, although there seems to be a sharp difference between OA and other types of open movements. The fact that it is being heavily institutionalized and that this institutionalization is being adopted on a top-down manner, we argue, is key to understand in what way this is so.

The 3 main characteristics that make it peculiar to other initiatives is that OA publishing has been mostly adopted late in the rather short history of open movements (during the early 2000s with the BBBs), although the practice has been around since early 1990s with the flagship example of the arXiv. The movement was adopted by librarians and researchers inside their work environment, and not by hobbyists, as is usual in other initiatives[24]. Lastly, OA in the academic context is becoming extremely expensive, for it is being incorporated into traditional corporate models of industrial production of academic publication, through the application of publication fees (the APCs). To sum up, what differentiates OA from the rest of the open movements is its intention of opening up the access to all academic literature, instead of sticking to granting open access to all literature that is published under an open license. Accordingly it can be said that OA is a flagship attempt to impose an open model as an industry standard.

[23] Free as in freedom;

[24] To clarify, this does not mean that FOSS is made by amateurs, in fact it is developed and used by professionals and enterprises all around the world constantly, but the software is developed by professionals outside their work context, giving rise to great working groups such as for instance Debian (Coleman, 2013)

As we have seen, OA is being incorporated into all sides of publishing practices in such a way that much of its changing capabilities end up being suppressed, which raises the question of whether or not OA can really improve channels of communication and grant equality of voice and access among researchers throughout the world.

Our results have shown that there is a significant amount of the world's population that will not harvest the benefits of OA due to lack o basic social services and conditions, for they endure posterior forms of epistemological disadvantages (Piron et al., 2017). In other words, the disparities of economical development between North and South can be of such a scale that tools such as OA can benefit one type of countries, while simply not fitting the realities of the socio-economic structures of the other. For this, education would be required, but most countries investing in education towards open science are mainly located in the North.

Since it is such a highly symbolical process of production[25], entering academic publications and knowledge production can become increasingly hard for southern researchers in an OA model of publication, and traditional publication models might be more appealing to researchers in countries located in these highly disadvantaged regions.

Finally, mandatory OA seems to be thought of at high political levels as a tool to be used in the context of the European Digital Single Market – which could explain to some degree the disparities between North and South. This can be seen for instance in the contradictory measures[26] of embracing mandatory OA at European level while sanctioning Internet freedom giving more power to publishers and copyright.

The case of OA in North markets is very particular and characterized by the battle between the stakeholders of academic publication. The landscape of publications is very important in this case, since, more than half of scholarly publications are owned by 3 companies either located in Europe, or in the United States. For instance, Elsevier is based in the Netherlands, while RELX Group, its parent group, is based in the UK. Holtzbrinck Publishing Group, parent group of Springer Nature, Macmillan and BioMed, to name a few, is based in Germany, while Whiley's headquarters are in the US. These are also the countries in which the OA movement has had stronger support. While these are strong companies, this does not seem to affect positively the countries in which they are based on.

While for once OA was proposed as the cure to this internally corrosive problem, currently, it seems, OA might be in many aspects reproducing the problems that existed before it.

Relatively to levels of inequality inside the European region we witness signs of academic dependency (Alatas, 2003; Sinha-Kerkhoff et al., 2010) between the semi-peripheral countries of Europe, such as Portugal, Spain and Greece, although further investigation would be required to ensure this.

Finally the so-called emerging markets such as the BRICS, but also several other countries

[25] See (Eve, 2014);

[26] For more on this discussion see `https://www.eff.org/files/2018/06/13/article13letter.pdf` and (Poynder, 2018);

in the LA seem to be in a position to be benefiting the most out of OA's institutionalization, since they do not face the battle with publishers that is taking place in the North, and might be able to raise enough power and capital to make their own channels of certified knowledge and communication. It also provides a global, standard platform for these markets to compete. It is worth noting how these countries are intrinsically different, since the LA are betting heavier in the social sciences, open digital libraries, Spanish and Portuguese literature, while Asian countries mainly bet on OA journals, STEM subjects and English literature. In other words, while the LA is using OA in order to create South to South channels of communication, Asian countries are investing in South to North communications. This reveals two very distinct strategies and appropriations of OA.

Relatively to our research question and macro objective, it seems that our hypothesis has been corroborated – there are many instances of OA that vary along the landscape of a wider global ecosystem of interdependency and industrial production. While in some situations OA might promote diversity and South to South and South to North communication channels, in other it might favour North to North and North to South communications.

Future Work

Our investigation only tangentially approaches the matter of scholarly communications from the article's side. In other words, our main focus throughout this investigation was on the reader's side of OA and not so much on the author's side. Future work will involve the gathering of information related to the DOI of articles in the Sci-Hub 2017 statistics dataset. Providing insight into what sort of articles people are reading, the discipline, journal – as well as the publisher, but that was already covered in this investigation – and authors. This will provide answers on questions related to discipline, subject, and other topics related to *'who is reading who?*.

References

105th Congress. (1998). Digital Millenium Copyright Act.

Alatas, S. F. (2003). Academic Dependency and the Global Division of Labour in the Social Sciences. *Current Sociology*, *51*(6), 599–613. doi:10.1177/00113921030516003

Alexandra Elbakyan. (2017). Some facts on Sci-Hub that Wikipedia gets wrong.

Allington, D. (2013). On open access, and why it's not the answer.

AOKI, M. (2007). Endogenizing institutions and institutional changes. *Journal of Institutional Economics*, *3*(01), 1. doi:10.1017/S1744137406000531

Archambault, É., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L., & Roberge, G. (2014). Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013. *Copyright, Fair Use, Scholarly Communication, etc.*

Aspesi, C., & Luong, H. (2014). Reed Elsevier: Goodbye to Berlin - The Fading Threat of Open Access (Upgrade to Market-Perform), 1–23.

Barlow, J. P. (1996). Electronic Frontier Foundation. *A Declaration of the Independence of Cyberspace*, *56*(3), 1–2. doi:10.5860/CHOICE.48-1189

Beall, J. (2012). Predatory publishers are corrupting open access. *Nature*, *489*(7415), 179. doi:10.1038/489179a

Beall, J. (2017). What I learned from predatory publishers. *Biochemia Medica*, *27*(2), 273–278. doi:10.11613/BM.2017.029

Beigel, F. (2018). Academic Dependency.

Berry, D. M. (2008). *Copy, Rip, Burn: The Politics of Copyleft and Open Source*. doi:10.2307/j.ctt183q67g

Birdsall, N. M., Campos, J. E. L., Kim, C.-S., Corden, W. M., MacDonald, L. [, Pack, H., … Stiglitz, J. E. (1993). *Main report* (tech. rep. No. 12351). The World Bank.

Björk, A., Paavola, J.-M., Ropponen, T., Laakso, M., & Lahti, L. (2018). *Opening Academic Publishing: Development and application of systematic evaluation criteria*.

Björk, B. C. (2012). The hybrid model for open access publication of scholarly articles: A failed experiment? *Journal of the American Society for Information Science and Technology*, *63*(8), 1496–1504. doi:10.1002/asi.22709

Björk, B. C. (2017). Gold, green, and black open access. *Learned Publishing*, *30*(2), 173–175. doi:10.1002/leap.1096

Björk, B.-C., & Solomon, D. (2014). *Developing An Effective Market For Open Access Article Processing Charges*. Zenodo. doi:10.5281/zenodo.51788

Bohannon, J. (2013). Who's afraid of peer review? *Science (New York, N.Y.) 342*(6154), 60–5. doi:10.1126/science.342.6154.60

Bohannon, J. (2016). Who's downloading pirated papers? Everyone. *Science*, *352*(6285), 508–512. doi:10.1126/science.352.6285.508

Brienza, C. (2011). Communication or credentialing? On the value of academic publishing, 2011.

Byrne, D. (1998). COMPLEXITY THEORY AND THE SOCIAL SCIENCES: An introduction, 215.

Cabanac, G. (2016). Bibliogifts in LibGen? A study of a text-sharing platform driven by biblioleaks and crowdsourcing. *Journal of the Association for Information Science and Technology*, *67*(4), 874–884. doi:10.1002/asi.23445

Castells, M. (2000a). *The Rise of the Network Society. The Information Age: Economy, Society and Culture. Volume I*. Wiley.

Castells, M. (2000b). Toward a Sociology of the Network Society. *Contemporary Sociology*, *29*(5), 693. doi:10.2307/2655234

Castells, M. (2014). Rethinking Development in the global information age, Centre for Higher Education Transformation (CHET).

Chen, X., & Olijhoek, T. (2016). Measuring the Degrees of Openness of Scholarly Journals with the Open Access Spectrum (OAS) Evaluation Tool. *Serials Review*, *42*(2), 108–115. doi:10.1080/00987913.2016.1182672

Clark, J., & Smith, R. (2015). Firm action needed on predatory journals. *BMJ*, *350*(jan16 1), h210–h210. doi:10.1136/bmj.h210

Coleman, E. G. (2013). *Coding Freedom -THE ETHICS AND AESTHETICS OF HACKING*. PRINCETON UNIVERSITY PRESS.

Collyer, F. M. (2018). Global patterns in the publishing of academic knowledge: Global North, global South. *Current Sociology*, *66*(1), 56–73. doi:10.1177/0011392116680020

Commission, E. (2007). On scientific information in the digital age: Access, dissemination and preservation.

Commission, E. (2009). ICT INFRASTRUCTURES FOR e-SCIENCE.

Commission, E. (2010a). A strategy for smart, sustainable and inclusive growth.

Commission, E. (2010b). *Riding the Wave*.

Commission, E. (2012a). A Reinforced European Research Area Partnership for Excellence and Growth.

Commission, E. (2012b). On access to and preservation of scientific information.

Commission, E. (2012c). Recommendation on access to and preservation of scientific information in Europa, 1–125. doi:http://dx.doi.org/10.4403/jlis.it-8649

Commission, E. (2016). H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020.

Commission, E., & Commission, E. (2012). *Survey on open access in FP7*. Eurepean Commission. doi:10.2777/81083

Commission, E., From, C., Commission, T. H. E., The, T. O., Council, T. H. E., Economic, T. H. E. E., … The, O. F. (2012). No Title.

Commisson, E., & Commisson, E. (2011). *National open access and preservation policies in Europe: Analysis of a questionnaire to the European Research Area Committee*. Eurepoean Commission. doi:10.2777/74027

Cooper, I. D. (2015). Bibliometrics basics. *Journal of the Medical Library Association : JMLA*, *103*(4), 217–218. doi:10.3163/1536-5050.103.4.013

Costa, S. M. S., & Leite, F. C. L. (2008). Brazilian Open Access Initiatives: Key Strategies and Actions. *ELPUB 2008: 12th International Conference on Electronic Publishing. Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0, 25-27 June 2008, University of Toronto, Toronto, Canada*, (June 2008), 288–298.

Doctorow, C., Gaiman, N., & Palmer, A. (2015). *Information Doesn't Want to Be Free: Laws for the Internet Age* (Reprint edition). San Francisco: McSweeney's.

Edwards, L., & Waelde, C. (2009). *Law and the Internet - A Foundation for Electronic Commerce*. Hart Pub.

Eisen, M. (2013). Beall's Litter.

Ellen Finnie Duranceau. (2008). The "Wealth of Networks" and Institutional Repositories: MIT, DSpace, and the Future of the Scholarly Commons. *Library Trends*, *57*(2), 244–261. doi:10.1353/lib.0.0030

Europe, O. (2017). European Copyright reform: Impact on Free and Open Source Software and developer communities.

Eve, M. P. (2014). *Open Access and the Humanities*. doi:10.1017/CBO9781316161012

Ezema, I. J., & Okafor, V. N. (2015). Open Access Institutional Repositories in Nigeria Academic Libraries. *Library Collections, Acquisitions, & \ldots*.

Finch, J. (2012). *Accessibility, sustainability, excellence: How to expand access to research publications'. Finch Group report: Government response on open access - Publications - GOV.UK*.

Finn, M., & Fisher, R. (2018). Open-Access and long-form (book) publishing in the 2027 REF exercise, 2.

Florence Piron. (2018). Open Access in the Francophone Global South: Between Collective Empowerment and Neocolonialism.

Gardner, G. J., Mclaughlin, S. R., & Asher, A. D. (2017). Shadow Libraries and You : 568–587.

Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association*, *295*(1), 90–93. doi:10.1001/jama.295.1.90

Ginsparg, P. (2011a). It was twenty years ago today ... doi:10.1163/174552512X640323

Ginsparg, P. (2011b). It was twenty years ago today ... doi:10.1163/174552512X640323

Greshake, B. (2017). Looking into Pandora's Box: The Content of Sci-Hub and its Usage. *F1000Research*, *6*, 541. doi:10.12688/f1000research.11366.1

Guédon, J. C. (2004). The "green" and "gold" roads to open access: The case for mixing and matching. *Serials Review*, *30*(4), 315–328. doi:10.1080/00987913.2004.10764931

Harnad, S. (2014). Open access: What, where, when, how and why. In *Ethics, science, technology and engineering: An international resource*.

Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., … Hilf, E. R. (2008). The access/impact problem and the green and gold roads to open access: An update. *Serials Review*, *34*(1), 36–40. doi:10.1080/00987913.2008.10765150

Held, D., McGrew, A. G., Goldblatt, D., & Perraton, J. (1999). *Global Transformations: Politics, Economics, and Culture* (1 edition). Stanford, Calif: Stanford University Press.

Hetman-Krajewska, J. (2016). Open Access to Scientific Works: Axiology, History and Chosen Aspects of Copyright Solutions. *Odprti dostop do znanstvenih del: aksiologija, zgodovina in izbrani vidiki avtorskih rešitev. 11*(1), 81–95.

Heylighen, F. (2007). Why is Open Access Development so Successful? Stigmergic organization and the economics of information. *arXiv preprint cs/0612071*, 12.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431. doi:10.1038/520429a

Himanen, P. (2001). *The Hacker Ethic and the Spirit of the Information Age*.

Himmelstein, D. S., Romero, A. R., Levernier, J. G., Munro, T. A., McLaughlin, S. R., Greshake Tzovaras, B., & Greene, C. S. (2018). Research: Sci-Hub provides access to nearly all scholarly literature. *eLife*, *7*, e32822. doi:10.7554/eLife.32822

Hodson, S., Jones, S., Collins, S., Genova, F., Harrower, N., Laaksonen, L., … Wittenburg, P. (2018). Turning Fair Data Into Reality: Interim Report From The European Commission Expert Group On Fair Data. doi:10.5281/zenodo.1285272

Holton, R. J. (2005). *Making Globalization*. Macmillan International Higher Education.

Jones, G. (2005). Jones.pdf. https://homepage.univie.ac.at/peter.eigner/WS/Jones.pdf.

Kircz, J. (2005). Institutional Repositories, a new platform in Higher Education and Research The right question to your solution Institutional Repositories, a new platform in Higher Education and Research.

Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLoS ONE*, *10*(6), e0127502. doi:10.1371/journal.pone.0127502

Lessig, L. (2006). *Code* (Version 2.0). OCLC: ocm77638613. New York: Basic Books.

Lindtner, S. (2015). Hacking with Chinese Characteristics: The Promises of the Maker Movement against China's Manufacturing Culture. *Science Technology and Human Values*, *40*(5), 854–879. doi:10.1177/0162243915590861

Lindtner, S., Anderson, K., & Dourish, P. (2012). Cultural appropriation: Information technologies as sites of transnational imagination. (p. 77). doi:10.1145/2145204.2145220

Lipscomb, C. E. (2001). Mergers in the publishing industry. *Bulletin of the Medical Library Association*, *89*(3), 307–8.

Lynch, C. A., & Lippincott, J. K. (2005). Institutional repository deployment in the United States as of early 2005. *D-Lib Magazine*, *11*(9). doi:10.1045/september2005-lynch

Martell, L. (2010). *Sociology of Globalization*. London: Polity Press.

Martín-Martín, A., Costas, R., van Leeuwen, T., & Delgado López-Cózar, E. (2018). *Evidence of Open Access of scientific publications in Google Scholar: A large-scale analysis*. doi:10.17605/OSF.IO/K54UV

Milanovic, B. (2016). Global Inequality: A New Approach for the Age of Globalization. *Panoeconomicus*, *63*(4), 493501. doi:10.1017/CBO9781107415324.004

Morrison, H., Salhab, J., Calvé-Genest, A., & Horava, T. (2015). Open Access Article Processing Charges: DOAJ Survey May 2014. *Publications*, *3*(1), 1–16. doi:10.3390/publications3010001

Mukherjee, B. (2012). India ' s Efforts in Open Access Publishing, 1–27.

Munroe, M. H. (2007). The Academic Publishing Industry: A Story of Merger and Acquisition. https://www.ulib.niu.edu/publishers/.

Nassi-Calò, L. (2013). How much does it cost to publish in Open Access?

O'Carroll, C., Kamerlin, C. L., Brennan, N., Hyllseth, B., Kohl, U., O{u2019}Neill, G., … Directorate-General for Research and Innovation. (2017). *Providing researchers with the skills and competencies they need to practise Open Science*. OCLC: 1044479238.

Oxenham, S. (2016). Meet the Robin Hood of science. https://bigthink.com/neurobonkers/a-pirate-bay-for-science.

Parliament, E., & Council, E. U. (2001). On the harmonisation of certain aspects of copyright and related rights in the information society. *Official Journal of the European Union*, *L167*(February), 12–25.

Piketty, T., & Goldhammer, A. (2014). *Capital in the twenty-first century*. Cambridge Massachusetts: The Belknap Press of Harvard University Press.

Pinfield, S., Salter, J., & Bath, P. A. (2016). The "total cost of publication" in a hybrid open-access environment: Institutional approaches to funding journal article-processing charges in combination with subscriptions. *Journal of the Association for Information Science and Technology*, *67*(7), 1751–1766. doi:10.1002/asi.23446

Piron, F., Diouf, A. B., Dibounje Madiba, M. S., Mboa Nkoudou, T. H., Aubierge Ouangré, Z., Tessy, D. R., … Lire, Z. (2017). Le libre accès vu d'Afrique francophone subsaharienne. *Revue française des sciences de l'information et de la communication*, (11). doi:10.4000/rfsic.3292

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., … Haustein, S. (2018). The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ Preprints*. doi:10.7287/peerj.preprints.3119v1

Poynder, R. (2017a). Open and Shut?: Has the open access movement delayed the revolution?

Poynder, R. (2017b). Open and Shut?: Q&A with PLOS co-founder Michael Eisen.

Poynder, R. (2018). How do we interpret the fact that the EU has just voted for a link tax a week after it outlined plans to mandate that all research papers are made available as CC BY? https://creativecommons.org/licenses/by/3.0/ …https://twitter.com/creativecommons/status/1039835296 Tweet.

Raymond, E. (1999). The cathedral and the bazaar. *Knowledge, Technology & Policy*, *12*(3), 23–49. doi:10.1007/s12130-999-1026-0

RCUK. (2005). RCUK Position Statement on Access to Research Outputs, 1–15.

Research Councils UK. (2012). RCUK Policy on Open Access. (1), 1–13.

Robertson, R. (1992). *Globalization - Social Theory and Global Culture*. SAGE Publications Ltd.

Santos, B. d. S. (1985). Estado e Sociedade na Semiperiferia do Sistema Mundial: o Caso Português.

Sassen, S. (1991). *The Global City*. Polity Press.

Sassen, S. (2005). The Global City: Introducing a Concept. *The Global City*, (2), 18.

Sassen, S. (2007). *Sociology of globalization* (1st ed). Contemporary societies series. OCLC: ocm65064787. New York: W.W. Norton.

Sassen, S. (2008). The Global City. *A Companion to the Anthropology of Politics*, 168–178. doi:10.1002/9780470693681.ch11

Sassen, S. (2014). *Expulsions : Brutality and complexity in the global economy*.

SciHub. (2017). Note: In Sci-Hub access logs released previous year, all requests are resolved requests, i.e. user successfully downloaded PDF with that DOI. Tweet.

SciHub. (2018a). Sci-Hub download log for 2017 year: http://sci-hub.tw/downloads/2017.statistics.tab … two most read 2017 papers are on climate change http://sci-hub.tw/10.1175/JCLI-D-16-0526.1 … and obesity http://sci-hub.tw/10.1056/NEJMra1514009 …pic.twitter.com/gxtyfrR4n5. Tweet.

SciHub. (2018b). Update to 2017 download stats: It turned out that all DOI starting from '10.1016/j' pattern were marked as erroneous and were filtered out by exporter script. that happened by mistake. here the all missed downloads: http://sci-hub.tw/downloads/2017.statistics.1016.j.tab … Tweet.

Sinha-Kerkhoff, K., Alatas, S. F., & Jha, D. N. (2010). A Common Pool of Knowledge or Academic Dependency? An Introduction to the Problem, 19.

Stallman, R. (2002). *Free software, free society : Selected essays of Richard M. Stallman*. doi:10.1353/lib.2006.0005

Stephan, P., Veugelers, R., & Wang, J. (2017). Reviewers are blinkered by bibliometrics. *Nature News*, *544*(7651), 411. doi:10.1038/544411a

Stevan Harnad. (2006). ROAR to DOAR - Open Access Archivangelism.

Stiglitz, J. E. [Joseph E]. (2017). Roosevelt Institute Working Paper, 13.

Stiglitz, J. E. [Joseph E.]. (2003). *Joseph E. Stiglitz (2002) Globalization and Its Discontents*. doi:10.1046/j.0391-5026.2003.00107.x

Suber, P. (2008). Thinking about prestige, quality, and open access. *SPARC Open Access Newsletter*, (125).

Suber, P. (2012). *Open access*. MIT Press essential knowledge series. OCLC: ocn754518563. Cambridge, Mass: MIT Press.

Swartz, A. (2008). Guerilla Open Access Manifesto. (July), 2.

Tickell, P. A. (2016). Open access to research publications: Independent advice, 39.

Tzovaras, B. G. (2018). Sci-Hub download log of 2017. doi:10.5281/zenodo.1158301

UNCTAD. (2017). *Information Economy Report 2017*.

Wallerstein, I. (2004). *World-Systems Analysis: An Introduction*. a John Hope Franklin Center Book. Duke University Press.

Warwick, H. (2014). *Radical Tactics of the Offline Library*.

Waters, M. (2001). *Globalization*. Psychology Press.

Young, A. (1995). The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience. *The Quarterly Journal of Economics*, *110*(3), 641–680. doi:10.2307/2946695