

1. COVID-19, TEXT AND DATA MINING AND COPYRIGHT: THE BRAZILIAN CASE *

Allan Rocha de Souza**

ABSTRACT

The COVID-19 pandemic has intensified the importance of text and data mining (TDM) techniques and tools, which are behind several key applications in the fight against SARS-CoV-2. As such, this paper discusses the importance of TDM tools in scientific and technological innovation, as well as how such technologies, which depend heavily on open access and circulation of information, are affected by current copyright protection on databases, especially in developing countries, taking Brazil as an example. To this end, the paper uses bibliographic and documental sources where TDM played a crucial role in research on the pandemic and in combatting disinformation. The work begins with an introduction on TDM, databases, and machine learning technologies, their applications, and their importance for innovation and for scientific and technological innovation. Next, it discusses the nature of database protection via copyright and its implications for the development of data-intensive research and technologies, as well as the role of limitations and exceptions in this process, as illustrated by recent initiatives taken by several countries. The authors conclude that, as it stands, copyright protection of

databases creates extraordinary obstacles to the access and use of data for research. In addition, we suggest that the promotion of limitations and exceptions in this area is central to the scientific development and innovation, for reducing the technological gap between countries and, specifically, for the success of the fight against this and other pandemics.

Keywords: *copyright, text and data mining, databases, limitations and exceptions, COVID-19.*

1. INTRODUCTION

The use of systematic techniques of data collection and analysis as an important element in the construction of scientific knowledge is not a recent phenomenon. In the 18th century, philosopher and theologian Joseph Priestley was already making extensive use of quantitative and qualitative constellations of data and data sets to substantiate his research.¹ Similarly, a century later, Ellen Garvey presented the book 'American Slavery As It Is: Testimony of a Thousand Witnesses' as an example of a product resulting from a systematic and meticulous process of collecting, organizing, correlating, and presenting data, following a specific purpose and narrative.²

Likewise, while data has always been seen as an important resource, it was already produced in too large a scale and pace to be used to its full potential, often

* This paper is part of a large research project on Copyright, Right of Access and Innovation, which has been partially funded by the Brazilian Copyright Institute (IBDautoral), the Arcadia Right to Research in International Copyright Project, and the National Institute of Science and Technology (INCT) *Proprietas*, together with CAPES, CNPq, and FAPERJ.

** Allan Rocha holds a PhD on Cultural Rights and Copyright (UERJ, 2010) and is currently a Copyright Professor at the Graduation Program on Public Policy, Strategies and Development and the Civil Law and Humanities Department of the Federal University of Rio de Janeiro (PPED/UFRJ), (DDHL/ITR/UFRJ). He also teaches Copyright of the IP Specialization Course at Pontifícia Universidade Católica (PUC-RJ). He is the Vice Coordinator of the National Institute of Science and Technology (INCT) *Proprietas*, Scientific Director of the Brazilian Copyright Institute (IBDautoral), a copyright consultant at Fundação Oswaldo Cruz (FIOCRUZ), and a lawyer. This article was co-authored by Luca Schirru and Miguel Alvarenga.

*** Luca Schirru holds a PhD on Artificial Intelligence and Copyright from the Graduation Program on Public Policy, Strategies and Development of

the Federal University of Rio de Janeiro (PPED/UFRJ) and is currently an Arcadia Post-doctorate Fellow in International Copyright of the Information Justice and IP Program and an LLM candidate on IP and Technology at Washington College of Law, American University. Additionally, he teaches Copyright of the IP Specialization Course at Pontifícia Universidade Católica (PUC-RJ). He is a lawyer and the incumbent Executive Director of the Brazilian Copyright Institute (IBDautoral).

**** Miguel Alvarenga is a PhD candidate at the Graduation Program on Public Policy, Strategies and Development of the Federal University of Rio de Janeiro (PPED/UFRJ). Besides being a lawyer and research fellow at the Brazilian Copyright Institute (IBDautoral), he is working on the thesis 'Data and Text Mining and Copyright'.

¹ Rosenberg D, 'Data before the Fact' in Lisa Gitelman, *Raw Data is an Oxymoron* (The MIT Press 2013) 15-40.

² Garvey EG, 'facts and FACTS': Abolitionists' Database Innovations' in Lisa Gitelman, *Raw Data is an Oxymoron* (The MIT Press 2013) 89-102.

requiring the employment of special analysis techniques to obtain results 'by approximation' (e.g., filtering, selection, and sampling). The main difference is that today, with the improvement of information and communication technologies, it has become possible to analyze vast amounts of data in real-time and obtain results at previously unseen levels of completeness and granularity.³

This scenario presents unique challenges for all countries, especially for developing ones, which tend to lack the necessary regulatory and institutional structures to thoroughly undertake the opportunities for research and innovation at present. In this context, the present study highlights the circumstances in Brazil in the hope of offering insight on the way developing countries might respond and avoid deepening the existing technological and knowledge gap. To this end, the work focuses on three main topics. Firstly, it looks into use and importance of text and data mining (TDM) tools for research and innovation. Secondly, it explores how current copyright-database protection impacts innovation and the struggle against the pandemic. Finally, it discusses the importance of copyright limitations and exceptions in this aspect.

To that end, the present paper begins by bringing forth selected cases from Brazil to illustrate the scientific and social importance of data-intensive technologies in the struggles against the COVID-19 pandemic and the spread of disinformation. It then provides a brief analysis of TDM technologies, their role in developing and operating AI systems, and how access to databases and other source materials are crucial in data analysis. We go on to observe the multi-layered protection afforded by the current copyright law structure over databases and the obstacles it presents to data-intensive innovation. Finally, before

concluding, the work stresses the role of copyright limitations and exceptions in TDM and their implementation in developed countries, as well as Brazil's current position regarding the matter, highlighting the need for developing countries to respond to current technological demands and bridge the divide in worldwide technological innovation.

2. SOCIAL VALUE AND INNOVATION IN DATA USE: EXAMPLES FROM BRAZIL

The possession and use of large amounts of data is currently seen by governments, businesses, and non-commercial entities as a valuable tool in innovation. These resources have been playing an increasingly vital role in several fields such as helping improve decision-making in the development of new technologies and providing support for faster, better, and novel scientific research. In the field of health, data resources have helped introduce improved medical diagnostics and genetic sequencing techniques to better understand various diseases.⁴

In the latter case, Brazil presents a recent example of how data-driven genetic research has been of valuable assistance during the COVID-19 pandemic. As the first cases of the disease were found in the country, a team of researchers from the Adolfo Lutz Institute and the University of São Paulo managed, in just 24 hours, to conduct a sequencing of the samples collected and discern the regions of origin of the virus, by analyzing the history of mutations seen in the organisms found in the samples and combining it with the observation of patient travel records, with the help of genetic data globally shared via the GISAID Platform.⁵ The said platform provides royalty-free access to its genetic database while

³ Martens B, 'The importance of data access regimes for artificial intelligence and machine learning' (2018) Digital Economy Working Paper 2018-09, JRC Technical Reports, 6-7 <<https://ec.europa.eu/jrc/sites/default/files/jrc114990.pdf>> accessed 20 June 2021; Dean J, *Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners* (1st edn, Wiley 2014) 11-14.

⁴ Chen H, Chiang R, Storey V, 'Business Intelligence and Analytics: from Big Data to Big Impact' (2012) 36 (4) MIS Quarterly: Management Information

Systems, 1168-1172 <<https://www.jstor.org/stable/41703503>> accessed 20 June 2021; Dean J (n 3) 3-5.

⁵ Goes de Jesus J, *et al.*, 'Importation and early local transmission of COVID-19 in Brazil' (2020) 62 Journal of the São Paulo Institute of Tropical Medicine <<https://www.scielo.br/j/rimts/a/ZXhgyfr6NznSJTFTfsbdsND/abstract/?lang=en>> accessed 20 June 2021.

creating simple rules for acknowledging the original creators of the data and restricting its usage by third parties.⁶

Such findings had several practical implications. Firstly, they suggested that the genomes analyzed presented variations in relation to the Chinese strain meaning that, at the time, the virus was already being internally transmitted in Europe, which pointed out possible locations for the implementation of travel restriction measures.⁷ More importantly, however, was the second implication: according to two of the main researchers of the study, successfully performing these sequencing methods constituted a crucial step towards understanding the main characteristics of the pathogen and how much it mutated – which was essential for developing effective treatments and vaccines.⁸

Another area where intensive data analysis proves itself increasingly useful is journalism, especially when it comes to fact-checking and combatting disinformation. In Brazil, multiple fact-checking and news agencies have made consistent efforts to verify content produced in social media and other outlets. In Brazil, the G1 Portal, owned by Globo – one of the largest media conglomerates in the country – has an entire section dedicated to this activity, called 'Fato ou Fake' ('Fact or Fake', in Portuguese). There, teams from different forms of media (e.g., radio, television, magazines, and the internet) collaborate in a joint effort to quickly obtain confirmation on content veracity.⁹

Another Brazilian example is AOs Fatos ('To the Facts', in Portuguese), which has a Radar based on algorithms constantly curated by linguists. The software collects publications and posts on several media such as WhatsApp, Facebook, YouTube, and others, looking for keywords that match content which is typically associated with false information on several topics, including those related to the COVID-19 pandemic.¹⁰ The agency then verifies the information and grades it from 1 (most unreliable) to 10 (most accurate or reliable).¹¹ The agency also has a bot on Twitter, called Fátima, dedicated to debunking false information on the platform that the agency has already checked.¹²

Despite all the clear advances, the mere possession of data is insufficient to make such projects viable. This is because raw data when considered in isolation, has no intrinsic meaning or utility in itself since the information that is sought is only revealed when data is adequately contextualized and interpreted.¹³ In other words, the great value currently attributed to data depends on its being processed by sophisticated collection and analysis tools capable of sifting through all that content and making sense of it.¹⁴

3. TEXT AND DATA MINING, AI, AND DATABASES

A crucial part of the above-mentioned process involves the use of **text and data mining** (TDM), which is defined as the set of techniques dedicated to finding patterns of interest from large amounts of data in a complex process of information collection. The analysis begins by the selection, cleaning, and integration of the relevant data

⁶ GISAID, 'Terms of Use' <<https://www.gisaid.org/registration/terms-of-use/>> accessed 20 June 2021.

⁷ Toledo K, 'Sequenciamento identifica genomas diferentes nos dois casos brasileiros de coronavírus' *USP Newspaper* (3 December 2020) <<https://jornal.usp.br/ciencias/ciencias-da-saude/sequenciamento-identifica-genomas-diferentes-nos-dois-casos-brasileiros-de-coronavirus/>> accessed 6 September 2021.

⁸ Brazilian Ministry of Health, 'Sequenciamento do coronavírus possibilita o desenvolvimento de vacinas' *Blog da Saúde* (2020) <<https://web.archive.org/web/20210309015721/http://www.blog.saude.gov.br/index.php/perguntas-e-respostas/54104-confira-a-entrevista-sobre-o-sequenciamento-do-coronavirus>> accessed 9 June 2021.

⁹ Globo, G1 'G1 lança Fato ou Fake: novo serviço de checagem de conteúdos suspeitos' <<https://g1.globo.com/fato-ou-fake/noticia/2018/07/30/g1-lanca-fato-ou-fake-novo-servico-de-checagem-de-conteudos-suspeitos.ghtml>> accessed 9 June 2021.

¹⁰ AOs Fatos, 'Radar' <<https://www.aosfatos.org/radar/#/>> accessed 9 June 2021.

¹¹ AOs Fatos, 'Nosso Método' <<https://web.archive.org/web/20210510104827/https://www.aosfatos.org/nosso-metodo/>> accessed 9 June 2021.

¹² AOs Fatos, 'Fátima' <<https://twitter.com/fatimabot>> accessed 5 June 2021.

¹³ Rowley J, 'The wisdom hierarchy: representations of the DIKW hierarchy' (2007) 33 (2) *Journal of Information Science*, 170-171 <<https://journals.sagepub.com/doi/abs/10.1177/0165551506070706>> accessed 20 June 2021.

¹⁴ Dean J (n 3) 4-5.

into a single location and converted into an intelligible format. Then TDM is applied to find correlations and patterns from which a variety of information can be extracted and evaluated by the analyst or by the machine, and then presented to the user.¹⁵

With the increase in computing power and the ever-growing production of data seen today, it follows that the amount of data one has to work with usually demands processing power that far exceeds what human beings can achieve, making it unfeasible to operate a large database or mine its contents without proper assistance. The use of artificial intelligence (AI) systems is to assist in these tasks – instead of letting the AI systems perform entirely on their own – is therefore on the increase, through a so-called machine learning process. Machine learning refers to the use of procedures and techniques that enable a machine to process information and, as the name suggests, learn from it, extracting information that will serve as a basis for task-solving and pursuing assigned objectives in a flexible way.¹⁶

The use of machine learning has been advocated to address several issues, including the COVID-19 pandemic. Recent publications propose using AI systems in the formulation of predictive models of case growth¹⁷ as well as for developing new medication¹⁸ and diagnosing methods for COVID-19, be it through information analysis¹⁹ or imaging.²⁰ In Brazil, Aos Fatos demonstrates that, when vast quantities of data are produced and replicated every second, sorting through what is true or

false as well as signaling incorrect content that reappears regularly, can be quite a challenge for a human team of experts to do consistently. The use of AI algorithms, e.g., the ones operating the Radar has become increasingly important in fact-checking.

On the other hand, while AI and algorithms are generally capable of rapidly processing unimaginable amounts of data, they do not have the same sophistication of thought that humans possess. Machine-learning tools usually require millions of inputs in order to apprehend simple information that an ordinary person would be able to gather in no more than a glance.²¹ The selection and classification of large amounts of data are, therefore, crucial for an algorithm to learn and achieve the intended results of its operation.²² Drawing on the work of Russel and Norvig,²³ it is possible to assert that data is a central element today in the operation and training of AI systems, assuming its position as a key player in data analysis which was once exclusively attributed to the algorithm.

It should be noted, therefore, that the AI itself is but one part of the architecture that makes text and data mining possible. One must also consider several other components such as the servers responsible for searching for relevant data, the knowledge base that informs the parameters to be used in the processing, search and evaluation of patterns, the user interface, and, most importantly, the sources of the data that is collected which is usually stored in databases.²⁴

¹⁵ Han J, Pei J, Kamber M, *Data mining: concepts and techniques* (Elsevier 2011) 7; Kroenke DM, *et al.*, *Database Concepts* (8th ed., Pearson 2016) 493; Kelleher J, Tierney B, *Data Science* (MIT Press 2018) 241-242.

¹⁶ Samuel AL, 'Some Studies in Machine Learning Using the Game of Checkers' (2000) 4 (1/2) *IBM Journal of Research and Development*, 207-226; Kaplan A, Haenlein M, 'Siri, Siri, in my hand: Who's the fairest of the land? On the interpretations, illustrations, and implications of artificial intelligence' (2018) 62 *Business Horizons*, 15-25.

¹⁷ Marni M, Chokani N, Abhari RS, 'COVID-19 Epidemic in Switzerland: Growth Prediction and Containment Strategy Using Artificial Intelligence and Big Data' (2020) *medRxiv* <<https://www.medrxiv.org/content/10.1101/2020.03.30.20047472v2>> accessed 20 June 2021.

¹⁸ Ho D, 'Addressing COVID-19 Drug Development with Artificial Intelligence' (2020) 2 (5) *Advanced Intelligent Systems* <<https://onlinelibrary.wiley.com/doi/full/10.1002/aisy.202000070>> accessed 20 June 2021.

¹⁹ de Moraes Batista AF, *et al.*, 'COVID-19 diagnosis prediction in emergency care patients: a machine learning approach' (2020) *medRxiv* <<https://www.medrxiv.org/content/10.1101/2020.04.04.20052092v1>> accessed 20 June 2021.

²⁰ Wang S, *et al.*, 'A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)' (2020) *medRxiv* <<https://www.medrxiv.org/content/10.1101/2020.02.14.20023028v5>> accessed 20 June 2021.

²¹ Martens B (n 3) 3.

²² Drexel J, *et al.*, 'Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective' (2019) *Max Planck Institute for Innovation and Competition Research Paper No. 19-13* <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3465577> accessed 20 June 2021.

²³ Russel S, Norvig P, *Artificial Intelligence* (Regina Célia Simile tr, 3rd ed, Elsevier 2013) 25.

²⁴ Han J, Pei J, Kamber M (n 15) 7-9.

In a digital format, databases can be defined as structures created out of information collected from various sources. In practical terms, their purpose is to allow the collected data to be preserved and accessed in a more organized fashion, enabling various pieces of information to be effectively cross-referenced. Such organization includes the creation and storage of metadata and indexes, as well as descriptions of the applications used. These groupings or collections of data and metadata are part of a large system that includes computer software dedicated to creating, processing, and administering these databases – which comprise the database management system (DBMS): the applications that work as an interface between users and the DBMS, and the users themselves.²⁵

Databases play a key role in the architecture of data collection and analysis, serving two primary purposes: first, as points of origin for extraction, query, and operational activities, and secondly, as **data warehouses** i.e. points of destination where all the data collected from multiple sources are reunited, so that its contents can be adequately modeled and prepared for the upcoming analysis, and the results of data mining can be stored for future cross-referencing.²⁶ The assembly, acquisition, and maintenance of large databases are, therefore, vital components within the framework that is built to make data mining possible.

4. MULTIPLE LAYERS OF PROTECTION: REGULATORY OBSTACLES TO DATA USE

While several organizations largely see the collection and possession of large quantities of data as a crucial asset for improving their activities, it is also true that the process

of building and maintaining databases is often “messy”, demanding considerable investment of time, technical resources, and, more importantly, financial and human capital.²⁷ It should, therefore, come as no surprise that certain rights holders regard this as an incentive to seek the limiting or constraint of access to their databases, so as to prevent third parties from using the contents of such databases without authorization.²⁸

This usually stems from the understanding that data, as a set of abstract representations, is non-rival (their consumption does not prevent subsequent use by third parties) and, in theory, 'non-excludable' (controlling and restricting their acquisition would be very difficult or impossible).²⁹ This peculiar nature of data would in turn mean that simply copying a database from elsewhere would cost much less than creating a new one, leading to possible underinvestment. Such a concern would lead to the adoption of institutional measures, at national and global levels, aimed at mitigating the issue.³⁰ One such measure was the insertion of databases and their content into an institutional-technical system of intellectual-property rights protection that operates, according to Estelle Derclaye, in three main 'layers': copyright, technological protection measures, and anti-circumvention measures.³¹

At the first layer, since the introduction of the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) in 1994, collections of data have been globally granted, alongside computer programs, the status of protected works under copyright law, and any protected works contained inside these repositories retain their protection as well. Even before that, Brazil already had a similar provision for databases at the time under the Brazilian Copyright Act of 1998 (Article 7, XIII

²⁵ Kroenke DM, *et al.* (n 15) 3-32.

²⁶ Han J, Pei J, Kamber M (n 15); Kroenke *et al.* (n 15) 492-494; Kelleher J, Tierney B (n 15) 8-9.

²⁷ Dean J (n 3) 12.

²⁸ Stucke ME, Grunes AP, 'Debunking the Myths Over Big Data and Antitrust' (2015) (2) CPI Antitrust Chronicle, 3.

²⁹ Ostrom E, Hess C, 'Private and Common Property Rights' (Workshop in Political Theory and Policy Analysis, Indiana University 2007) 9; Coriat B,

'From Natural-Resource Commons to knowledge Commons: Common Traits and Differences' (International Seminar on Property and Commons: new issues of shared access and innovation, Sorbonne 2013) 13.

³⁰ Wu T, 'Law and Economics of Information' in Parisi F, *The Oxford Handbook of Law and Economics* (Vol. 2, Oxford Handbooks 2017) 4; Derclaye E, *The Legal Protection of Databases: A Comparative Analysis* (Edward Elgar 2008) 35.

³¹ Derclaye E (n 30) 196.

and § 2) along with the inclusion of software in the scope of copyrightable products (Article 7, XII).

Here, one specific consideration must be made: in all cases mentioned above, copyright law only applies to databases when their content is selected or arranged in a manner that constitutes intellectual work. In other words, they must be sufficiently new or 'distinguishable', bringing something so original that they could not be mistaken for other works of the same genre.³² This is not the case with most digital databases, which usually serve to store content automatically, following standardized methods of collection and organization of content. There are three reasons for this: first, because the value of that data comes from the inferences drawn from it, not from its arrangement (as seen earlier); second, because the selection criteria for data collection is often purely quantitative ('the more, the better'), and lastly, because it is natural that such databases prefer to adopt specific storage and organization standards for reasons of accessibility and compatibility.³³

Regardless, the protection afforded by copyright law to databases has two implications: first, while many databases will not satisfy the legal criteria for protection, some will – and that means accessing and using their content will require previous and express authorization from the rights holders, even if the said content is comprised entirely of unprotected material (such as public information or just raw data). Secondly, as stated above, protected works inside a database retain their respective rights independently which means that using them still requires permission from the owner of each piece. If Lawrence Lessig had indicated the inherent

difficulty of locating and negotiating with every rights holder in a simple collection of films featuring one artist,³⁴ this would elevate to impossibly high levels once we consider that data analysis usually covers an enormous quantity of data – and, quite possibly, of protected works and owners as well. Last but not the least, substantial amounts of the data, including metadata, are either public, of public interest or outside the purview of intellectual property (IP). That being said, the reality of information technology has led to efforts in creating systems dedicated to controlling and restricting, by technical means, unauthorized access, and usage of protected works. In other words, we have, at the second layer of protection, the implementation of technological protection measures (TPM), dedicated to preventing copyright infringement rather than relying on the law to penalize it.³⁵ In databases, the DBMS, being responsible for managing the database and its access, could technically serve this role.

While the above can be effective in avoiding infringement, one must be critical of the abusive use of such tools. From the imposition of abusive clauses or the restriction on access and use of non-protected data or works in the public domain to the installation of spyware, it has already been pointed out that such mechanisms often use the architecture of databases to control access and use of protected content beyond the scope that would normally be afforded by copyright law, restricting legitimate uses of lawfully acquired material.³⁶

Regardless, technical barriers are not invulnerable as anyone with sufficient knowledge and resources can surpass them. And this is where the third layer of

³² Borges Barbosa D, 'A noção de Originalidade e os Títulos de Obra, em particular, de Software' (2005) <<http://denisbarbosa.addr.com/originalidade.pdf>> accessed 9 June 2021.

³³ Derclaye E (n 30) 45; Truyens M, van Eecke P, 'Legal aspects of text mining' (2014) 30 Computer Law & Security Review, 163-164; Banterle F, *Data ownership in the data economy: a European dilemma. EU Internet Law in the digital era* (edited volume based on the REDA 2017 conference, Springer 2018, forthcoming) 5.

³⁴ Lessig L, *Free Culture: How Big media uses technology and the law to lock down culture and control creativity* (The Penguin Press 2004) 103.

³⁵ Bechtold S, 'Digital Rights Management in the United States and Europe' (2004) 52 The American Journal of Comparative Law.

³⁶ Lessig L (n 34), 157; Vieira Branco S, *O Domínio Público no Direito Autoral Brasileiro – Uma Obra em Domínio Público* (Lumen Juris 2011) 269; Rocha de Souza A, Schirru L, 'Os direitos autorais no marco civil da internet' (2016) 12 (1) Liinc em Revista, 41 <<http://revista.ibict.br/liinc/article/view/3712>> accessed 20 June 2021; Pereira de Souza CA, 'Abuso do direito autoral' (DSc thesis, University of the State of Rio de Janeiro 2009); Rocha de Souza A, *A Função Social dos Direitos Autorais: Uma Interpretação Civil-Constitucional dos limites da proteção jurídica. Brasil: 1988-2005* (Coleção José do Patrocínio; v.4, Ed. Faculdade de Direito de Campos 2006).

protection, according to Derclaye, operates in the form of anti-circumvention measures. These measures are established by copyright law to bar the suppression, modification, or destruction of technological protection devices that seek to prevent infringement, as well as the production, sale, and distribution of devices that may be used for such purposes.³⁷ An example can be found in Article 107 of the Brazilian Copyright Act, which outlaws the alteration, suppression, modification, or destruction of technical devices put in place to avoid unlawful copying of a work.

Despite being put in place by copyright law, anti-circumvention measures belong to a third layer because the protection they offer is not aimed at databases or software themselves but at the technological devices that prevent unauthorized access or copying of protected works. Therefore, they would constitute a kind of 'paracopyright', which operates outside the core object of copyright, but still within the system.³⁸ Moreover, in countries such as Brazil, there is no provision specifying that copyright limitations and exceptions override this clause, thus creating the possibility for these measures to override user rights effectively granted by law.

5. LIMITATIONS AND EXCEPTIONS FOR RESEARCH AND INNOVATION: BRIDGING (OR DEEPENING) THE DIVIDE

When considered together, these layers of protection become especially problematic once we consider that TDM typically require the copying, extraction, and

modification of existing content in third-party databases, implying that such acts would require prior authorization from the owners. Consequently, there is a potential danger of violation of reproduction rights, especially if a substantial part of the collection is copied which is quite common since many TDM processes aim to obtain as much relevant information as possible. On the other hand, if we are dealing with an original database, both the reproduction of relevant material and the discarding of content irrelevant to the analysis may also constitute copyright infringement, as they may replicate or alter the selection or arrangement of the database from which the material was extracted, implying infringement of both the right to reproduce and adapt the work.³⁹

According to José de Oliveira Ascensão, all these are intensified by the reduction of copyright limitations, which exist precisely to balance the exclusivity resulting from the copyright system due to public interest⁴⁰ and to reconcile it with other equally important fundamental rights⁴¹. This is evinced by the fact that the extension of copyright to databases, at least in Brazil, was not followed by a limitation stipulating the conditions under which the access and use of databases and their contents would be allowed, even if only for scientific research purposes. This becomes particularly egregious if one considers that TDM generally does not interfere with the normal economic exploitation of a copyrighted work as it mainly involves using archives as a source for data analysis. There is no use of the expression of these works – which is the actual object of copyright protection, as the idea/expression distinction dictates.⁴²

³⁷ Derclaye E (n 30).

³⁸ Lessig L (n 34); Brown K, 'Digital Rights Management: Trafficking in Technology That Can Be Used to Circumvent the Intellectual Property Clause' (2003) 40 (3) *Houston Law Review*, 804 <[review.org/article/4808-digital-rights-management-trafficking-in-technology-that-can-be-used-to-circumvent-the-intellectual-property-clause](https://www.review.org/article/4808-digital-rights-management-trafficking-in-technology-that-can-be-used-to-circumvent-the-intellectual-property-clause)> accessed 20 June 2021.

³⁹ Geiger C, Frosio G, Bulayenko O, 'The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market – Legal Aspects' (2018) Centre for International Intellectual Property Studies Research Paper No. 2018-02, 6-7 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3160586> accessed 20 June 2021.

⁴⁰ de Oliveira Ascensão J, *Direito da Internet e da sociedade da informação: estudos* (Forense 2002) 33, 136.

⁴¹ Rocha de Souza A, de Serpa Pinto Fairbanks A, 'The Marrakesh Treaty Ratification in Brazil: Immediate Effects' (2017) 4 (5-6) *Panorama of*

Brazilian Law <<https://www.e-publicacoes.uerj.br/index.php/pbl/article/view/34438/24355>> accessed 20 June 2021; Rocha de Souza A, 'Resgatando a função cultural nos direitos autorais' in Moura Vicente D (Org.) *Estudos de Direito Intelectual em homenagem ao Prof. Doutor José de Oliveira Ascensão. 50 anos de Vida Universitária* (1ed, v.1, Almedina 2015); Geiger C, Frosio G, Bulayenko O, 'The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market – Legal Aspects' (2018) Centre for International Intellectual Property Studies Research Paper No. 2018-02, 6-7 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3160586> accessed 20 June 2021.

⁴² European Copyright Society, 'General Opinion on the EU Copyright Reform Package' (2017) 5; Flynn S, et al., 'Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action' (2020) PIJIP/TLS Research Paper Series, 48, 2

It is also important to note that, while some projects can rely on relatively open databases widely adopted by fellow researchers (such as in the COVID-19 genetic sequencing case) or do possess human and financial resources to carry out fast and intensive data-gathering work on their own (such as 'Fato ou Fake' and 'Aos Fatos'), not all research projects fall under these two possibilities. In fact, in the academic field, it is common to find journals or other means of scientific communication that sometimes impose prohibitive fees for access to the texts they disseminate, often leading to the necessity to be affiliated to some institution capable of licensing multiple publishers.

In such a scenario, researchers whose investigation work may require the use of copyrighted material are faced with three options: (1) to avoid using protected material altogether, thus limiting the available material for analysis and possibly compromising the effectiveness, quality, or even the viability of the research; (2) to seek authorization from the rights holders – which is becoming increasingly unfeasible, as previously noted; or (3) to use protected material without authorization and subject to eventual litigation, bringing potentially prohibitive financial and temporal costs.

The result is an institutional framework that generates insecurity for users and imposes bureaucratic barriers to legitimate access to databases and the use of their content, undermining data collection and analysis, now so rooted in the process of scientific research and of

innovation in general.⁴³ Considering such burden, the final decision may well be not to engage in the research at all.

With that in mind, proposals for creating copyright limitations and exceptions for TDM, or at least regulatory measures dedicated to enabling their use in some capacity, have been tabled and implemented in several industrialized countries over the last two decades. In the United States, for instance, reproduction and use of copyrighted material for text and data mining and data analysis have already been considered fair use, as seen in *Kelly v. Arriba Soft Corp*, *Authors Guild, Inc. v. Hathi Trust* and in *Authors Guild, Inc. v. Google, Inc. (Google Books)*.⁴⁴

Meanwhile, in Europe, there have been discussions regarding the creation of a regulation applied to the European Union as a whole. In the steps leading to the creation of a new Copyright Directive, academic institutions such as the European Copyright Society and the Max Planck Institute have tabled proposals of their own regarding the use of copyrighted materials for TDM, including situations in which lawful access to the data is not possible or for commercial purposes.⁴⁵ Daniel Gervais went so far as to take limitations and exceptions already in place for TDM in the United Kingdom, France, and Germany as examples for special topics to be considered when crafting a regulation of that sort, such as the rights involved, the nature of use, the ability to forfeit the law via contract and position regarding anti-circumvention measures, among others.⁴⁶

<https://digitalcommons.wcl.american.edu/research/48/> accessed 20 June 2021. On the idea/expression distinction, see: Banterle F (n 33) 4; Hugenholtz B, 'Data property: Unwelcome Guest in the house of IP' in Julia Reda, *Better Regulation for Copyright: Academics meet Policy Makers* (TheGreens/EFA 2017) 69.

⁴³ Brown K (n 38) 821-822; Costa Lewicki B, 'Limitações aos direitos de autor: releitura na perspectiva do direito civil Contemporâneo' (DSc thesis, State University of Rio de Janeiro, 2007); Pereira de Souza CA, 'Abuso do direito autoral' (DSc thesis, University of the State of Rio de Janeiro 2009) 221-253.

⁴⁴ *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 822 (9th Cir. 2003); *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 90 (2d Cir. 2014); *Authors Guild, Inc. v. Google, Inc. (Google Books)*, 804 F.3d 202 (2d Cir. 2015); Carroll MW, 'Copyright and the Progress of Science: Why Text and Data Mining is Lawful' (2020) Washington College of Law Research Paper No. 2020-15,

American University Washington College of Law, 916-922 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3531231 accessed 20 June 2021 .

⁴⁵ European Copyright Society (n 44); Hilty R, Richter R, 'Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules Part B Exceptions and Limitations (Article 3 – Text and Data Mining) (14 January 2017)' (2017) Max Planck Institute for Innovation & Competition Research Paper No. 17-02 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2900110 accessed 20 June 2021.

⁴⁶ Gervais D, 'Exploring the Interfaces Between Big Data and Intellectual Property Law' (2019) 10 (1) *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*.

In the end, the Directive 2019/790 for the Digital Single Market⁴⁷ brought about two new TDM limitations and exceptions to be implemented by their Members: Article 3 is specific for the use of lawfully accessed materials for scientific research and preservation, while Article 4 is a blanket regulation allowing text and data mining for any purpose, provided that such right was not already reserved by the rightsholders. Meanwhile, the Japanese Copyright Act, which, in 2009, had previously included a TDM provision in the form of Article 47-7,⁴⁸ has incorporated, in 2018, several articles concerning the possibility of reproducing protected works for any non-expressive purpose, including computational data analysis – more specifically, in Articles 30-4, 47-4 and 47-5.⁴⁹

6. BRIDGING (OR DEEPENING) THE DIVIDE

In Brazil, as very likely in many other developing countries, the situation is different from what we observe in developed countries. Although Brazil's National Plan for the Internet of Things in 2019 establishes the free flow of data as one of its main foundations (Article 1),⁵⁰ no regulation aiming at promoting such a principle has been implemented so far. That same year, while the government opens a public consultation for Copyright reform with special mention to the need to adapt to AI and data-intensive technologies, the call remarkably focuses on raising user awareness about online copyright

infringement⁵¹, but brings nothing regarding the subject. It was only in 2021 that the Brazilian Strategy of AI recognized the need for discussing the implementation of a TDM exception in an official capacity.⁵² While there are still no effective regulations in this regard, Bill No. 21/2020, which seeks to regulate AI operation in Brazil and is currently under discussion at the Brazilian Senate, features one particular provision in this direction: its Article 5, VIII states that the usage of data, databases or protected texts in order to train AI systems shall not constitute copyright infringement, as long as it does not interfere with normal exploitation of the work.⁵³

Discussions of this nature are becoming increasingly imperative not only in Brazil, but in low and middle-income countries in general, especially considering the regulatory divide in copyright balance pointed out by Sean Flynn and Michael Palmedo: while all countries seem to have more open and user-friendly copyright regulations over time, developing nations seem to be stuck in a 30-year gap in comparison to more developed countries.⁵⁴ As regards TDM for research purposes or innovation, the pattern seems to be reinforced, as most countries with TDM limitations and exceptions (or with a broad research clause that is open to such techniques) are either located in the Global North or largely industrialized. Most developing nations do not have such openness: in South America, for instance, Ecuador stands out as the only country in the region with a clear TDM

⁴⁷ EU Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC [2019] OJ L 130.

⁴⁸ Act No. 48 of 6 May 1970 (Copyright Act) Amendment of Law No. 73 of 2009 (Japan) <<http://www.japaneselawtranslation.go.jp/law/detail/?printID=&ft=1&re=02&dn=1&x=0&y=0&co=01&ia=03&ky=copyright+act&page=24&vm=02&lv=02&id=3379>> accessed 9 June 2021.

⁴⁹ Act No. 48 of 6 May 1970 (Copyright Act) Amendment of Act No. 30 of 2018 (Japan) <<http://www.japaneselawtranslation.go.jp/law/detail/?printID=&ft=1&re=02&dn=1&x=0&y=0&co=01&ia=03&ky=copyright+act&page=24&vm=02&lv=02&id=3379>> accessed 14 February 2021.

⁵⁰ Decreto No. 9.854, de 25 de junho de 2019 Institui o Plano Nacional de Internet das Coisas e dispõe sobre a Câmara de Gestão e Acompanhamento do Desenvolvimento de Sistemas de Comunicação Máquina a Máquina e Internet das Coisas (Brazil) <http://www.planalto.gov.br/ccivil_03/_Ato20192022/2019/Decreto/D9854.htm> accessed 9 June 2021.

⁵¹ Brazilian Ministry of Citizenship (Special Secretariat of Culture), 'Consulta pública sobre Lei de Direitos Autorais é prorrogada até 15 de setembro de 2019' (2019) <<http://cultura.gov.br/consulta-publica-sobre-lei-de-direitos-autorais-e-prorrogada-ate-15-de-setembro/>> accessed 9 June 2021.

⁵² Portaria GM, No. 4.617, de 6 de abril de 2021. Institui a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos (*Diário Oficial da União* 2021) (Brazil).

⁵³ Brazilian Federal Senate, 'Projeto de Lei No. 21 de 2020. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências' (2022) <<https://legis.senado.leg.br/sdleg-getter/documento?dm=9063365&ts=1651257291876&disposition=inlin e>> accessed 11 May 2022.

⁵⁴ Flynn S, Palmedo M, 'The User Rights Database: Measuring the Impact of Copyright Balance' (2017) 14-17; Flynn S, 'Statement of Sean Flynn, Principal Investigator, Global Expert Network on Copyright User Rights' (Second Session of the WIPO Conversation on IP and AI 2020) 1-2.

exception in the form of Article 212, No. 9, VIII of the 'Código Orgánico de la Economía Social de los Conocimientos, Creatividad e Innovación', which allows libraries and archives to perform text mining in some circumstances. A few other countries in the region do have limitations that are at least conducive to text and data mining, such as Article 71(O) of the Chilean Copyright Law and Article 16(a) of Colombia's Law No. 1915/2018 – both of which allow transitory reproductions of works. Other countries in the region, such as Brazil, Argentina, and Uruguay, keep very restrictive limitation regimes, especially when it comes to scientific and technological research.

Therefore, it is imperative to reconsider the current state of many copyright regimes, especially in the Global South. While many countries have taken decisive steps in recent years to ensure that national IP regulations can better act as enablers of innovation, most of the reforms seen up to this point seem to be concentrated in developed nations. If the Brazilian cases observed in this study are of any indication, however, not only do developing countries struggle with the same issues but they can also bring important contributions to the resolution of such problems. Bridging the institutional divide in openness between North and South is, thus of the utmost importance if one seeks to truly obtain satisfactory results in the combat against several global crises – whether they be a pandemic of the body or of the mind.

7. CONCLUSIONS

This paper focused on the use and importance of TDM tools for research and innovation and how current copyright database protection and corresponding limitations and exceptions impact on innovation and the struggle against the pandemic. The paper looked at the Brazilian experience, namely its early research efforts against COVID-19 and national initiatives against disinformation so as to stress the importance of said technologies for developing countries and the necessity

to address TDM regulations beyond the context of developed nations.

From the outset, the paper presented some key applications of data-intensive technologies and their social value, particularly in Brazil, where data-driven genetic research has provided valuable assistance during the COVID-19 pandemic, having also been employed in fact-checking and in combatting disinformation. The second part addressed the content of Text and Data Mining and the role of AI in supporting the processing of vast amounts of data to find patterns of interest and to extract knowledge out of relational or non-relational resource pools, as well as the effective embeddedness of AI technologies in TDM. We went on to focus on the insertion of databases and their content into an institutional-technical system of IP rights protection that operates, according to Estelle Derclaye, in three main 'layers': copyright, technological protection measures, and anti-circumvention measures. The copyright protection granted to databases, the risk of facing infringement claims in countries where there are no suitable L&Es in copyright law, and the importance of enabling access to data to conduct scientific research was then examined. We stress that it is imperative to reconsider the current state of many copyright regimes, especially in the Global South, considering that most of the reforms seen up to this point seem to be concentrated in developed nations.

The analysis concludes that, in a context where scientific and journalistic activity increasingly depends on access to data to flourish and provide better results, the Brazilian cases have provided strong evidence not only of the role of data-intensive activities in the resolution of globalized problems, but also of the importance of an institutional and regulatory environment that works as an enabler and promoter of research and innovation. Allowing such a gap to continue means more than just imposing renewed obstacles to development; it also means prolonging health crises and creating difficulties for the maintenance

of healthy political discourse and the improvement of social welfare throughout the world.

As a final suggestion, we take the opportunity to advise developing countries to incorporate, as soon as possible, friendlier and more expansive limitations and exceptions to copyright that could ensure the proper balance under current circumstances as well as promote one's ability to innovate, such as permitted uses for research purposes and TDM activities. Furthermore, as has become clear, this is not simply a local and national problem, but an international one, as such activities are increasingly carried under cross-border cooperation. This issue may call for international agencies such as World Intellectual Property Organization to take a fresh look into the need to promote worldwide research and text and data mining limitations and exceptions, including by means of discussion and adoption of legal instruments, model legislation and guidance documents

BIBLIOGRAPHY

Abhari RS, Marini M, Chokani N, 'COVID-19 Epidemic in Switzerland: Growth Prediction and Containment Strategy Using Artificial Intelligence and Big Data' (medRxiv 2020).

Alvarenga MB, 'Mineração de dados, Big Data e Direitos Autorais no Brasil' (Masters Thesis, Federal University of Rio de Janeiro 2019).

Aos Fatos, 'Radar' <<https://www.aosfatos.org/radar/#!/>> accessed 9 June 2021.

Aos Fatos, 'Nosso Método' <[https://web.archive.org/web/20210510104827/https://www.aosfatos.org/nosso-m\[%C3%A9\]todo/](https://web.archive.org/web/20210510104827/https://www.aosfatos.org/nosso-m[%C3%A9]todo/)> accessed 9 June 2021.

Aos Fatos, 'Fátima' <<https://twitter.com/fatimabot>> accessed 5 June 2021.

Ascensão JO, 'Direito da Internet e da sociedade da informação: estudos' (Forense 2002).

Banterle F, 'Data ownership in the data economy: a European dilemma. EU Internet Law in the digital era (edited volume based on the REDA 2017 conference)' (Springer 2018, forthcoming).

Barbosa DB, 'A noção de Originalidade e os Títulos de Obra, em particular, de Software' (2005) <<http://denisbarbosa.addr.com/originalidade.pdf>> accessed 6 September 2021.

Batista AFM, *et al.*, 'COVID-19 diagnosis prediction in emergency care patients: a machine learning approach' (medRxiv 2020).

Bechtold S, 'Digital Rights Management in the United States and Europe' (2004) 52 *The American Journal of Comparative Law* 323.

Branco SV, 'O Domínio Público no Direito Autoral Brasileiro – Uma Obra em Domínio Público' (Lumen Juris, 2011).

Brazilian Federal Senate, 'Projeto de Lei No. 21 de 2020. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências' (2022) <<https://legis.senado.leg.br/sdleg-getter/documento?dm=9063365&ts=1651257291876&disposition=inline>> accessed 11 May 2022.

Brazilian Ministry of Citizenship (Special Secretariat of Culture), 'Consulta pública sobre Lei de Direitos Autorais é prorrogada até 15 de setembro de 2019' (2019) <<http://cultura.gov.br/consulta-publica-sobre-lei-de-direitos-autorais-e-prorrogada-ate-15-de-setembro/>> accessed 9 June 2021.

Brazilian Ministry of Health, 'Sequenciamento do coronavírus possibilita o desenvolvimento de vacinas'

Blog da Saúde (2020)
<<https://web.archive.org/web/20210309015721/http://www.blog.saude.gov.br/index.php/perguntas-e-respostas/54104-confira-a-entrevista-sobre-o-sequenciamento-do-coronavirus>> accessed 9 June 2021.

Brown K, 'Digital Rights Management: Trafficking in Technology That Can Be Used to Circumvent the Intellectual Property Clause' (*Houston Law Review* 803 2013) 804.

Carroll MW, 'Copyright and the Progress of Science: Why Text and Data Mining is Lawful' (Washington College of Law Research Paper No. 2020-15, American University Washington College of Law 2020).

Coriat B, 'From Natural-Resource Commons to knowledge Commons: Common Traits and Differences (International Seminar on Property and Commons: new issues of shared access and innovation, Sorbonne 2013).

Chen H, Chiang R, Storey V, 'Business Intelligence and Analytics: from Big Data to Big Impact' (*MIS Quarterly: Management Information Systems* 36 (4) 2012).

Dean J, *Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners* (ProQuest Ebook Central, Wiley 2018).

Derclaye E, *The Legal Protection of Databases: A Comparative Analysis* (Edward Elgar 2008).

Drexl J, *et al.*, 'Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective' Max Planck Institute for Innovation and Competition Research Paper Series – Research Paper No. 19-13 (Research Group on the Regulation of the Digital Economy 2019).

European Copyright Society, 'General Opinion on the EU Copyright Reform Package' (2017).

Flynn S, *et al.*, 'Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action' (PIJIP/TLS Research Paper Series, 48 2020).

Flynn S, Palmedo M, 'The User Rights Database: Measuring the Impact of Copyright Balance' (2017) 14-17.

Flynn S, 'Statement of Sean Flynn, Principal Investigator, Global Expert Network on Copyright User Rights' (Second Session of the WIPO Conversation on IP and AI 2020).

Garvey EG, 'facts and FACTS': Abolitionists' Database Innovations' in GITELMAN L, *Raw Data is an Oxymoron* (The MIT Press 2013).

Geiger C, Frosio G, Bulayenko O, 'The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects' (Centre for International Intellectual Property Studies Research Paper No. 2018-02 2018).

Gervais D, 'Exploring the Interfaces Between Big Data and Intellectual Property Law' *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* (Vol. 10 (1) 2019).

GISAID, 'Terms of Use'
<<https://www.gisaid.org/registration/terms-of-use/>>
accessed 20 June 2021.

Globo, G1 'G1 lança Fato ou Fake: novo serviço de checagem de conteúdos suspeitos'
<<https://g1.globo.com/fato-ou-fake/noticia/2018/07/30/g1-lanca-fato-ou-fake-novo-servico-de-checagem-de-conteudos-suspeitos.ghtml>>
accessed 9 June 2021.

Han J, Pei J, Kamber M, *Data mining: concepts and techniques* (Elsevier 2011).

- Hilty R, Richter R, 'Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernization of European Copyright Rules Part B Exceptions and Limitations (Article 3 – Text and Data Mining)' (14 January 2017)". (Max Planck Institute for Innovation & Competition Research Paper No. 17-02 2017).
- Ho D, 'Addressing COVID-19 Drug Development with Artificial Intelligence' *Advanced Intelligent Systems* (Vol. 2. 5, WILEY-VCH Verlag GmbH & Co. KGaA 2020).
- Hugenholtz B, 'Data property: Unwelcome Guest in the house of IP' in REDA J, *Better Regulation for Copyright: Academics meet Policy Makers* (TheGreens/EFA 2017).
- Jesus G, *et al.*, 'Importation and early local transmission of COVID-19 in Brazil' (Journal of the São Paulo Institute of Tropical Medicine 2020).
- Kaplan A, Haenlein M, 'Siri, Siri, in my hand: Who's the fairest of the land? On the interpretations, illustrations, and implications of artificial intelligence' (2018) 62 *Business Horizons*, 15-25.
- Kelleher J, Tierney B, *Data Science* (MIT Press 2018).
- KROENKE DM *et al.*, *Database Concepts* (8th ed., Pearson 2016).
- Lessig L, *Free Culture: How Big media uses technology and the law to lock down culture and control creativity* (The Penguin Press 2004).
- Martens B, 'The importance of data access regimes for artificial intelligence and machine learning' *Digital Economy Working Paper 2018-09, JRC Technical Reports* (European Commission 2018).
- Ostrom E, Hess C, 'Private and Common Property Rights' (Workshop in Political Theory and Policy Analysis, Indiana University 2007).
- Rosenberg D, 'Data before the Fact' in GITELMAN L, *Raw Data is an Oxymoron* (The MIT Press 2013).
- Rowley J, 'The wisdom hierarchy: representations of the DIKW hierarchy' (*Journal of Information Science*, Vol. 33 (2) 2007).
- Russel S, Norvig P, *Artificial Intelligence* (3rd ed, Elsevier 2013).
- Samuel AL, 'Some Studies in Machine Learning Using the Game of Checkers' (2000) 4 (1/2) *IBM Journal of Research and Development*, 207-226.
- Schirru L, 'Direito Autoral e Inteligência Artificial: Autoria e Titularidade nos Produtos da IA' PhD Dissertation (Federal University of Rio de Janeiro, 2020).
- Souza AR, *A Função Social dos Direitos Autorais: Uma Interpretação Civil-Constitucional dos limites da proteção jurídica. Brasil: 1988-2005* (Coleção José do Patrocínio, v.4, ed. Faculdade de Direito de Campos, 2006).
- Souza AR, Schirru L, 'Os direitos autorais no marco civil da internet' (Liinc em Revista, v. 12, 2016).
- Souza AR, Schirru L, Alvarenga MB, 'Direitos autorais e mineração de dados e textos no combate à COVID-19 no Brasil' (Liinc em Revista, v.16, 2020).
- Souza CAP, 'Abuso do direito autoral' (PhD Dissertation, University of the State of Rio de Janeiro 2009).
- Souza AR, Fairbanks ASP, 'The Marrakesh Treaty Ratification in Brazil: Immediate Effects' (*Panorama of Brazilian Law*, v.4 2017).
- Souza AR, 'Resgatando a função cultural nos direitos autorais' in Vicente DM (org.), *Estudos de Direito Intelectual em homenagem ao Prof. Doutor José de*

Oliveira Ascensão. *50 anos de Vida Universitária* (1ed, v.1, Almedina 2015).

Stucke ME, Grunes AP, 'Debunking the Myths Over Big Data and Antitrust' (CPI Antitrust Chronicle (2) 2015).

Toledo K, 'Sequenciamento identifica genomas diferentes nos dois casos brasileiros de coronavírus' (*USP Newspaper*, (3 December 2020) <<https://jornal.usp.br/ciencias/ciencias-da-saude/sequenciamento-identifica-genomas-diferentes-nos-dois-casos-brasileiros-de-coronavirus/>> accessed 6 September 2021).

Truyens M, Van Eecke P, 'Legal aspects of text mining' (*Computer Law & Security Review* 30 2014).

Wang S, *et al.*, 'A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)' (*medRxiv* 2020).

Wu T, 'Law and Economics of Information' in PARISI F, *The Oxford Handbook of Law and Economics* (Vol. 2, Oxford Handbooks 2017).