

RESEARCH EXCEPTIONS IN COMPARATIVE COPYRIGHT

*Sean M. Fiil-Flynn*¹

*with Luca Schirru, Michael Palmedo, Andres Izquierdo*²

April 2022

ABSTRACT

This Article categorizes the world's copyright laws according to the degree to which they provide exceptions to copyright exclusivity for research uses. We classify countries based on the degree to which they have a research exception in their law that is sufficiently open to be able to permit reproduction and communications of copyrighted work needed for academic (i.e. non-commercial) text and data mining (TDM) research. We show that nearly every copyright law has at least one exception that promotes uses for research purposes. We find six different approaches to the provision of research exceptions that implicate application to TDM. Notably, not all recent exceptions passed specifically to enable TDM receive the most open ranking in our typology. And a significant number of countries, marked red in our maps, do not provide a research exception or limit uses only to quotations. This report may be useful in helping countries find models for domestic copyright reform as well for consideration of guidelines or norms for harmonization between countries.

¹ Sean Flynn is Professorial Lecturer and Director, Program on Information Justice and Intellectual Property, American University Washington College of Law, and principal investigator of the Project on the Right to Research in International Copyright, supported by Arcadia, a Charitable Fund of Lisbet Rausing and Peter Baldwin. Professor Flynn is the “author” of this report to the extent he is responsible for the creative and analytical decisions contained within it. But the research and analysis is a product of collective work, as indicated below.

² This version of the study was carried out by Professor Flynn with assistance in writing and research by Luca Schirru, the Arcadia Post-doctorate Fellow in International Copyright at PIJIP. This version updates *Research Exceptions in Comparative Copyright Law* (2021) PIJIP/TLS Research Paper Series no. 72. <https://digitalcommons.wcl.american.edu/research/72>, which was assisted primarily by Mike Palmedo, PIJIP's Research Director, and Andres Izquierdo, PIJIP's Senior Research Analyst. Numerous PIJIP Research Assistants and Fellows contributed to the research.

ABSTRACT	1
INTRODUCTION	3
I. BACKGROUND AND STUDY DESIGN	5
<i>A. Copyright Exceptions and Text and Data Mining Research</i>	5
<i>B. Research Exceptions in International Copyright Law</i>	6
1. Quotation and Reproduction	7
2. Communication to the Public	9
<i>C. Study design</i>	10
1. Creating a typology of research exceptions	10
2. Limitations of the Typology	14
II. A TYPOLOGY OF RESEARCH EXCEPTIONS	15
<i>A. Open Research Exceptions</i>	16
<i>B. Closed Exceptions</i>	20
<i>C. Restrictions on Uses, Works and Users</i>	22
1. Reproduction Only	23
2. Restrictions to Institutional Users	23
3. Restrictions to Individual Users	25
4. Restrictions on Types of Works	26
III. A TYPOLOGY OF TEXT AND DATA MINING EXCEPTIONS	28
<i>A. Open TDM Exceptions</i>	29
<i>B. Reproduction Only</i>	30
<i>C. Restrictions to Institutional Users</i>	31
<i>D. Restrictions on Works</i>	31
<i>E. Safe harbor-only</i>	32
CONCLUSION	35

INTRODUCTION

Promoting “learning” and “science” were among the first purposes of early copyright laws.³ The granting of exclusive rights to authors was a means to these ends by creating the conditions for vibrant publishing markets. But exceptions to exclusive rights were also seen as necessary to permit copies for private study and classroom uses. Reflecting these commitments, the Berne Convention for the Protection of Literary and Artistic Works of 1886 included a specific provision permitting countries to allow uses of works “destined for educational or scientific purposes.”⁴

While most copyright laws today contain exceptions at least for limited copies made in the course of research,⁵ modern research methods often require more flexible uses than many of these exceptions permit. This Article focuses specifically on the practice of text and data mining (TDM) research, which uses computational methods to analyze information in books, articles, databases, and other sources.⁶ TDM is helping address some of the world’s greatest challenges from tracking disease outbreaks⁷ to examining hate speech and disinformation on social media,⁸ and beyond.⁹ But to engage in TDM, researchers need to make and share with others reproductions of whole

³ See 8 Anne, c. 19 (1710) (Gr. Brit), “The Statute of Anne” (stating its purpose as “the Encouragement of learning”); U.S. Const. art. 1, § 8, cl. 8 (granting the power to Congress to enact copyright “[t]o promote the progress of science”).

⁴ Berne Convention for the Protection of Literary and Artistic Works, art. 8 (1886).

⁵ PAUL GOLDSTEIN & BERNT HUGENHOLTZ, INTERNATIONAL COPYRIGHT: PRINCIPLES, LAW, AND PRACTICE (4th Ed. 2019) (finding that most copyright laws contain an exception for private reproductions for a research purpose).

⁶ ⁷ See Marc Prosser, How AI Helped Predict the Coronavirus Outbreak Before it Happened, SINGULARITY HUB (Feb. 05, 2020), <https://singularityhub.com/2020/02/05/how-ai-helped-predict-the-coronavirus-outbreak-before-it-happened/> (reporting on discovery of COVID-19 by Canadian TDM firm); Will Knight, Researchers Will Deploy AI to Better Understand Coronavirus, WIRED (Mar. 17, 2020), <https://www.wired.com/story/researchers-deploy-ai-better-understand-coronavirus/> (reporting on use of TDM in vaccine research).

⁸ See, e.g., Al-Makhadmeh, Z., Tolba, A. *Automatic hate speech detection using killer natural language processing optimizing ensemble deep learning approach*. COMPUTING 102, 505 (2020) (describing a machine learning tool to “help accurately predict hate speech”); E. Raisi and B. Huang, Cyberbullying detection with weakly supervised machine learning, in PROC. IEEE/ACM INT. CONF. ADV. SOCIAL NETW. ANAL. MINING 409 (Jul. 2017) MacAvaney S, Yao H-R, Yang E, Russell K, Goharian N, Frieder O *Hate speech detection: Challenges and solutions*, 14(8): e0221152 PLOS ONE 8-9 (2019) (describing machine learning tool for “hate speech detection”); Rui Cao, Roy Ka-Wei Lee, Tuan-Anh Hoang, *DeepHate: Hate Speech Detection via Multi-Faceted Text Representations*, ARXIV (2021) <https://arxiv.org/abs/2103.11799> (proposing “a novel deep learning framework known as DeepHate, which utilized multi-faceted text representations for automatic hate speech detection”).

⁹ See Bernt Hugenholtz, The New Copyright Directive: Text and Data Mining (Articles 3 and 4), Kluwer Copyright Blog, WOLTERS KLUWER (Jul. 24, 2019) (“In the industrial and commercial realm TDM has become even more pervasive. Text and data mining is nowadays standard practice in pharmaceutical research, journalism, information retrieval, search, and consumer information – to name just a few areas. TDM is also an essential tool in developing intelligent applications that require vast volumes of raw text and data to ‘self-learn’ complex tasks such as translation or speech recognition. Much of the current and future development in artificial intelligence, therefore, depends on TDM), <http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/>); R. Ducato & A. Strowel, *Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to “Machine Legibility”* 3 (CRIDES Working Paper Series, 2018) (“Several of these studies highlight that the beneficial uses of TDM are not limited to scientific research, but take place in other contexts, including consumer information and protection.” [...] “The automated analysis of contracts and privacy policy for enhancing the awareness of consumers and, ultimately, ensuring consumer empowerment, is a perfect lab to test TDM’s pierres d’achoppement.”); TDM Stories”. OPENMINTED, <http://openminted.eu/blog/> (last visited Mar. 29, 2022) (collecting examples).

works of all kind mu a multiplicity of users. A few countries have responded with specific copyright exceptions for “computational” or “information” analysis, or “text and data mining.”¹⁰ But despite numerous studies of copyright limitations and exceptions by the World Intellectual Property Organization¹¹ and by academics,¹² this is the first analysis to compare copyright exceptions around the world for their ability to enable TDM uses.

We find that only a minority of copyright research exceptions are fully open to TDM uses.¹³ Most countries have research exceptions, including for “private” or “personal” uses, that are restricted to some uses, by defined users, of limited kinds of works – effectively limiting application to many or all TDM projects. And a significant number of countries still provide only the minimum exceptions required by the Berne Convention – which extend only to the quotation of excerpts of works – effectively excluding all TDM uses.

This Article presents the results of our research in three parts. Part I describes the background and design of our study. Part II uses the data we collected to identify six major classifications of research exceptions in copyright law mapped from the most open to TDM (green in our map) to the most restrictive (red). Part III uses this same typology to analyze TDM-specific exceptions.

¹⁰ See Part III, *infra*.

¹¹ See Monica Torres & Raquel Xalabander, *Report on Practices and Challenges in Relation to Online Distance Education and Research Activities*, WIPO, SCCR/ 39/6, Thirty-Ninth Session (2019). Kenneth Crews, *Study on Copyright Limitations and Exceptions for Libraries and Archives*, WIPO, SCCR/17/2, Seventeenth Session (2008). See also Raquel Xalabarder, *Study on Copyright Limitations and Exceptions for Educational Activities in North America, Europe, Caucasus, Central Asia and Israel*, WIPO, SCCR/19/8, Nineteenth Session (2009). Sam Ricketson, *WIPO Study on Limitations and Exceptions of Copyright and Related Rights in the Digital Environment*, WIPO, SCCR/9/7 (2003). Daniel Seng, *Updated Study and Additional Analysis of Study on Copyright Limitations and Exceptions for Educational Activities*, WIPO, SCCR/35/5/REV, Thirty-Fifth Session (2017). Joseph Fometeu, *Study on Limitations and Exceptions for Copyright and Related Rights for Teaching in Africa*, WIPO, SCCR/19/5, Nineteenth Session (2009). Victor Nabhan, *Study on Limitations and Exceptions for Copyright for Educational Purposes in the Arab Countries*, WIPO, SCCR/19/6, Nineteenth Session (2009)

¹² For articles comparing exceptions and limitations from different countries, see, e.g. RETO HILTY AND SYLVIE NERISSON, *BALANCING COPYRIGHT A SURVEY OF NATIONAL APPROACHES* (2012) (analyzing common elements in limitations and exceptions from 38 countries); Sean Flynn and Mike Palmedo, *The User Rights Database: Measuring the Impact of Copyright Balance* (Joint PIJIP/TLS Research Paper Series no. 2018-01) (examining the openness of limitations and exceptions over time in 21 countries); Jonathan Band & Johnathan Gerafi, *The Fair Use/Fair Dealing Handbook* (March 2015) <http://infojustice.org/wp-content/uploads/2015/03/fair-use-handbook-march-2015.pdf> (compiling fair use and fair dealing exceptions); Christian Handke et al., *Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research*, (June 7, 2015) (classifying the exceptions of 40 countries on the degree of application to TDM uses). For recent work analyzing elements of text and data mining exceptions, see Sean Flynn, Christopher Geiger, João Quintais, Thomas Margoni, Matthew Sag, Lucie Guibault, Michael W. Carroll, *Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action* (Joint PIJIP/TLS Research Paper Series, No. 3, 2020), 7 EIPR (2020); Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66. JOURNAL OF THE COPYRIGHT SOCIETY OF THE USA 1, 34 (2019); Christophe Geiger, Giancarlo Frosio, Oleksandr Bulayenko. *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU*. In: "Propiedad intelectual y mercado único digital europeo" (Concepción Saiz Garcia & Raquel Evangelio Llorca eds., 2019), Tirant lo blanch, 27, Centre for International Intellectual Property Studies (CEIPI) Research Paper No. 2019-08; Sergey Filippov, *Mapping Text and Data Mining in Academic and Research Communities in Europe*, The Lisbon Council (2014); Ian Hargreaves et al., *Report from the Expert Group on Standardisation in the Area of Innovation and Technological Development, Notably in the Field of Text and Data Mining*, European Commission (2014); Jerome Reichman & Ruth Okediji, *When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale*, 96 MINN. L. REV. 1363, (2012). Patricia Aufderheide & Peter Jaszi, *Reclaiming Fair Use: How to Put Balance Back in Copyright*, University of Chicago Press (2018)

¹³ See Part IIA below, defining the green classification in our maps.

We conclude with reflections on possible next steps in our research as well as how researchers and policy makers may utilize the information we present thus far.

I. BACKGROUND AND STUDY DESIGN

We explain the legal and factual background and design of the study in this part, before reporting our results in Parts II and III.

A. Copyright Exceptions and Text and Data Mining Research

This Article contributes to the literature examining the openness of copyright exceptions for various uses.¹⁴ We use the terminology originally explained by Flynn and Palmedo. Specifically:

- We refer to an exception as being *open* in one or more dimensions if its terms permit application to all protected uses (e.g. reproduction, communication, etc.), of any work (e.g. literary, artistic, audiovisual, etc.), by any user (e.g. individuals, institutions, etc.), or for any purpose.¹⁵ We call an exception “fully open” if it is open on each dimension – uses, works, users, purpose.
- We refer to an exception as *general* if it can apply to uses for multiple purposes (e.g. research, education, criticism and review, etc.).¹⁶
- We refer to an exception as *flexible* if application of the exception turns on a case-by-case proportionality standard (e.g. consistent with fair practice) rather than a specifically delineated rule (e.g. “less than 10%).¹⁷

For example, using this terminology, the U.S. fair use exception is a fully open, general, flexible exception because it applies to any use, of any work, by any user, for any purpose subject to a 4-part proportionality test. At the opposite side of a spectrum may be a country like Argentina, which provides an exception only for the reproduction of extracts of a specific lengths (not all uses), of “intellectual works” (not all works), by a “person” (not all users), for “for didactic or scientific purposes, comments, criticisms or notes” (not all purposes).¹⁸

Text and datamining is one component of the digital revolution in scientific research that can collide with narrowly drafted copyright exceptions.¹⁹ The process for engaging in TDM can be

¹⁴ Sean Flynn & Mike Palmedo, *The User Rights Database: Measuring the Impact of Copyright Balance*, 8 (Joint PIJIP/TLS Research Paper Series no. 2018-01), http://infojustice.org/wp-content/uploads/2018/09/WCL_PIJIP_WhitePaper-UserRightsDatabase.pdf (“Openness: the user right can be applied to any purpose, use (aka activity), work or user;”)

¹⁵ *Id.* (“The U.S. fair use right in Section 107 is open (in each dimension), flexible, and general. The UK fair dealing clause is a flexible, general exception – but it is not open to any purpose. The South African quotation right is open to any purpose and is flexible, but is not open to any kind of use and is not general.”)

¹⁶ *Id.* (“*Generality*: the exception applies a single test to a group of permitted activities.”)

¹⁷ *Id.* (“*Flexibility*: the user right is applied through a flexible proportionality test that balances the interests of the rights holder with those of the user and general public;”)

¹⁸ Copyright Law, 1933 (Law No. 11.723 of September 28, 1933, on Legal Intellectual Property Regime, Copyright Law, as amended up to Law No. 26.570 of November 25, 2009) (Arg.) (“Article 10. Any person may publish, for didactic or scientific purposes, comments, criticisms or notes referring to intellectual works, including up to 1,000 words for literary or scientific works, or eight bars in musical works and, in all cases, only the parts of the text essential for that purpose. This provision shall cover educational and teaching works, collections, anthologies and other similar works. Where inclusions from works by other people are the main part of the new work, the courts may fix, on an equitable basis and in summary judgment, the proportional amount to which holders of the rights in the works included are entitled.”)

¹⁹ See, e.g., Jerome H. Reichman & Ruth L. Okediji, *When Copyright Law and Science Collide: Empowering*

described as involving several stages where copyrighted works may be reproduced or communicated to others.²⁰ First a researcher must access the materials to be mined – which could be published or unpublished articles, books, pictures, web pages, data sets, etc. Second, the researcher must extract (copy) the materials into a “corpus” that can be prepared for the research. Third, the researcher engages in the mining process, which involves analytical processing using computers and algorithms, as well as internal and external verification and validation. Finally, the researcher must disseminate the results of the research, which often requires disclosing all or part of the underlying sources for purposes of illustration and validation to others.²¹ At each of these stages, a researcher may be collaborating with others, including across borders.²²

Flynn and Palmedo analyze the openness of research and other exceptions in a basket of countries, as do Handke et al.²³ These studies suggest that more open research exceptions contribute to increased production of citable works of scholarship and increased academic use of TDM methodologies.²⁴ Both of those studies analyzed the changes in copyright laws in a several dozen countries over time. This Article is the first to apply a similar analysis of the openness of research exceptions in nearly every copyright law in the world. In a subsequent version, we intend to add a time dimension – tracking the changes in copyright laws toward more openness.

B. *Research Exceptions in International Copyright Law*

Authorizing research uses through exceptions to copyright’s exclusive rights has always been allowed, but never required, by international copyright treaties. As described below, the Berne Convention permits exceptions for both quotation and reproduction rights for research uses while

Digitally Integrated Research Methods on a Global Scale, 96 MINN LAW REV. 1362 (2012) (citing as motivation for the article the growth of information technology that is “transforming fields,” and has “revolutionized both how basic research is conducted and how the resulting knowledge is preserved and disseminated”); Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66 J. COPYRIGHT SOC’Y OF THE U.S.A. 291 (2019); Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining is Lawful*, 53 U.C. DAVIS L. REV. 893 (2019); ; Thomas Margoni, *Text and Data Mining in Intellectual Property Law: Towards an Autonomous classification of Computational Legal Methods*, (CREATE working paper 01/2020) in Irene Calboli & Maria Lilla Montagnani, *Handbook on Intellectual Property Research* (OUP, forthcoming 2020); Josef Drexl et al., *Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective* (Max Planck Institute for Innovation & Competition Research Paper No. 19-13, 2019).

²⁰ Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66. J. COPYRIGHT SOC’Y OF THE U.S.A. 1, 34 (2019).

²¹ Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66. J. COPYRIGHT SOC’Y OF THE U.S.A.1, 34 (2019).

²² See, e.g., Fangli Su, *Cross-national digital humanities research collaborations: structure, patterns and themes*. 76(6) THE JOURNAL OF DOCUMENTATION, 1295 (2020).

²³ Christian Handke et al., *Is Europe Falling Behind in Data Mining? Copyright’s Impact on Data Mining in Academic Research*, in *New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science: FfScale, Openness and Trust* 120–130 (Brigit Schmidt & Milena Dobreva eds., 2015)

²⁴ See Christian Handke et al., *Is Europe Falling Behind in Data Mining? Copyright’s Impact on Data Mining in Academic Research*, in *New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science: FfScale, Openness and Trust* 120–130 (Brigit Schmidt & Milena Dobreva eds., 2015), (“We demonstrate that countries in which data mining for academic research requires the express consent of rights holders, data mining makes up a significantly smaller share of total research output.”); Sean Flynn & Mike Palmedo, *The User Rights Database: Measuring the Impact of Copyright Balance* (Joint PIJIP/TLS Research Paper Series no. 2018-01) (finding that more open limitations and exceptions are correlated with higher research and development investments and scholarship output); Mike Palmedo, *The Impact of Copyright Exceptions for Researchers on Scholarly Output*, *Efil Journal of Economic Research*, 2(6), 114-39 (2019) (finding that “scientists residing in countries that implement more robust research exceptions publish more papers and books in subsequent years”).

the more recent World Copyright Treaties clarify that exceptions may also be extended to the communication to the public right, specifically recognizing the goal of promoting uses for “research” and “access to information.”

1. Quotation and Reproduction

The debates that produced the first multilateral copyright treaty – the Berne Convention of 1886 – included substantial attention to the issue of research and educational uses of copyrighted works.²⁵ The topic was championed by Germany who referred to “a principle recognized not only in practically all earlier conventions but also, specifically, by the French Government in the Franco-German Convention of 1883, the purpose of which is to provide education and scholarship with the means of drawing, to a limited extent, on the literature of other countries without having to resort to the author's authorization.”²⁶ Germany circulated an early survey of countries including the question:

In line with what has been accepted for practically all literary conventions at present in force, would it not be appropriate to establish, for the whole Union, the reciprocal right: "(a) To reproduce, without the author's consent, for scientific or teaching purposes, excerpts or whole sections of a work, subject to certain conditions?"²⁷

Germany later proposed text that would become Article 8 of the original Berne Convention, titled: “Lawful reproduction of protected works in scientific or educational works.” The draft article provided:

The publication in any of the countries of the Union of excerpts, fragments, or whole passages of a literary or artistic work that has appeared for the first time in any other country of the Union shall be lawful, provided that the publication is specially designed and adapted for education, or has scientific character. The reciprocal publication of chrestomathies consisting of fragments of works by various authors shall also be lawful, as shall the insertion in a chrestomathy or in an original work published in one of the countries of the Union of the whole of a short writing published in another country of the Union. It is understood that the name of the author from whom, or of the source from which, the excerpts, passages, fragments or writings referred to in the above two paragraphs have been borrowed shall always be mentioned. The insertion of musical compositions in collections intended for schools of music shall be considered unlawful reproduction, however.²⁸

The German delegate explained the provision as being needed to promote “a universal interest in certain borrowings from authors” for educational and scientific purposes,²⁹ and threatened to

²⁵ The historical research in this section was provided by PIJIP Fellow Lokesh Vyas.

²⁶ Minutes of the Fifth Meeting of the Conference for the Protection of Authors' Rights (Sept. 17, 1884), Draft Convention Concerning the Creation of a General Union for the Protection of Authors' Rights, in , International Bureau of Intellectual Property, The Berne Convention for the Protection of Literary and Artistic Works from 1886 to 1986, Reports of the Various Diplomatic Conferences, WIPO Publication No. 877, at 98 (delegate from Germany commenting in 1884).

²⁷ Minutes of the Second Meeting of the Conference for the Protection of Authors' Rights (Sept. 9, 1884) in International Bureau of Intellectual Property, The Berne Convention for the Protection of Literary and Artistic Works from 1886 to 1986, Reports of the Various Diplomatic Conferences, WIPO Publication No. 877 at 90-91 (discussing German question of delegates).

²⁸ Minutes of the Fifth Meeting of the Conference for the Protection of Authors' Rights (Sept. 17, 1884) in International Bureau of Intellectual Property, The Berne Convention for the Protection of Literary and Artistic Works from 1886 to 1986, Reports of the Various Diplomatic Conferences, WIPO Publication No. 877 at 98.

²⁹ Minutes of the Fifth Meeting of the Conference for the Protection of Authors' Rights (Sept. 17, 1884), Draft Convention Concerning the Creation of a General Union for the Protection of Authors Rights, in , International Bureau of Intellectual Property, The Berne Convention for the Protection of Literary and Artistic Works from 1886 to 1986,

pull out of the treaty negotiations if it was not included.³⁰ Germany described its proposal not as authorizing exceptions to exclusive rights but rather as putting in place a “restriction on copyright”³¹ that would create “the right” to make quotations for educational and scientific uses.³² The Committee considering the language reported that it “had acknowledged the existence of that [universal] interest” in permitting educational and scientific uses and “considered it preferable to provide for the reproduction right concerned in the General Convention rather than leave it to special conventions and the domestic legislation of each country.”³³

Article 8 of the final 1886 Convention permitted countries to protect “the liberty of extracting portions from literary or artistic works for use in publications destined for educational or scientific purposes, or for chrestomathies.”³⁴ The focus on use in publications, akin to the modern quotation right, rather than reproductions in the research process itself, was an artifact of the limited exclusive rights covered in the original Berne Convention. The first Berne Convention did not require an exclusive right of reproduction, but rather only addressed the rights to translation and performance. A right of reproduction was not added to the Berne Convention until the 1967 Stockholm revision. And thus the only revision to Article 8 before 1967, through the Brussels Act of 1948, continued to focus on “the right to include excerpts ... in educational or scientific publications,”³⁵ rather than a right to make copies in the research process. As described in part II below, some countries still only provide an exception for this limited form of research use of a copyrighted work through publication of quotations.

The 1967 revision of the Berne Convention to require a right of reproduction prompted

Reports of the Various Diplomatic Conferences, WIPO Publication No. 877, at 98.

³⁰ Minutes of the Fifth Meeting of the Conference for the Protection of Authors’ Rights (Sept. 17, 1884), Draft Convention Concerning the Creation of a General Union for the Protection of Authors’ Rights, in , International Bureau of Intellectual Property, The Berne Convention for the Protection of Literary and Artistic Works from 1886 to 1986, Reports of the Various Diplomatic Conferences, WIPO Publication No. 877, at 98 (Mr. Reichard from Germany in 1884: “I therefore hope that Mr. Lavollée’s intention is merely to state a way of thinking, and not to bring about a vote on Article 8 of the draft, the rejection of which would very probably place the German Government in the position of having to renounce the projected Union completely.”).

³¹ Minutes of the Fifth Meeting of the Conference for the Protection of Authors’ Rights (Sept. 17, 1884), Draft Convention Concerning the Creation of a General Union for the Protection of Authors Rights, in , International Bureau of Intellectual Property, The Berne Convention for the Protection of Literary and Artistic Works from 1886 to 1986, Reports of the Various Diplomatic Conferences, WIPO Publication No. 877, at 98 (“Mr. Reichard (from Germany comment) in 1884: “This is, therefore, one of the most universal principles, and one whose inclusion in the General Convention Germany will never renounce, because through the application of the laws of the country of origin, provided for in Article 2 of the draft Convention, the deletion of Article 8, which introduces a restriction on copyright, would make all provisions comparable to Article 8 contained in existing conventions void by virtue of the Additional Article.”).

³² See Minutes of the Third Meeting of the Conference for the Protection of Authors’ Rights (Sept. 8, 1885), Draft Convention Concerning the Creation of a General Union for the Protection of Authors Rights, in , International Bureau of Intellectual Property, The Berne Convention for the Protection of Literary and Artistic Works from 1886 to 1986, Reports of the Various Diplomatic Conferences, WIPO Publication No. 877, at 114 (“Dr. Dambach (from Germany) in 1885: pointed out that the case law and legislation of the various countries could vary, and that consequently it seemed preferable to retain Article 8 and to specify in the Convention itself the right to make quotations, etc.”).

³³ Minutes of the Fifth Meeting of the Conference for the Protection of Authors’ Rights (Sept. 17, 1884), Draft Convention Concerning the Creation of a General Union for the Protection of Authors Rights, in , International Bureau of Intellectual Property, The Berne Convention for the Protection of Literary and Artistic Works from 1886 to 1986, Reports of the Various Diplomatic Conferences, WIPO Publication No. 877, at 98.

³⁴ Berne Convention for the Protection of Literary and Artistic Works, art. 8 (1886).

³⁵ Berne Convention for the Protection of Literary and Artistic Works art. 10 (2), Sep. 9, 1886, as revised at Brussels on June 26, 1948.

substantial redesign of the Convention’s provisions on exceptions. The act of quotation, which had been the focus of Article 8 until then, was broken out and generalized into Article 10(1) – requiring all countries to allow quotation for any purpose “provided that their making is compatible with fair practice, and their extent does not exceed that justified by the purpose.”³⁶ Using the terminology described above – the quotation exception in Berne is open to purpose, works and users (but not use), prompting Alpin and Bentley to provocatively refer to it as “global, mandatory fair use.”³⁷

Article 10(2) added a permissive exception for educational uses – allowing utilization “by way of illustration in publications, broadcasts or sound or visual recordings for teaching.” Article 10(2) is open to uses that implicate any exclusive right by virtue of the use of the word “utilization”, but applies only to the limited purposes of “illustration ... for teaching.”

With the reproduction right now protected separately in Article 9(1), Article 9(2) was added granting general authority to every member “to permit the reproduction of such works in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author.” Article 9(2) only refers to the reproduction right (it is not open to uses), but is otherwise an open, general, flexible exception – it permits exceptions for any work, by any user, for any purpose, subject to a “3-step” proportionality analysis.³⁸

Although the amended Convention of 1967, which remains the operative text for these purposes today, no longer specifically mentioned research or scientific uses, the openness of the quotation exception in 10(1) and of the reproduction exception in 9(2) were seen as including, without being limited to, authorization permit research uses of works.³⁹

2. *Communication to the Public*

The 1996 WIPO Copyright Treaty (WCT) was the first to require copyright laws to protect an exclusive right of “communication to the public of their works, by wire or wireless means, including the making available to the public of their works in such a way that members of the public may access these works from a place and at a time individually chosen by them.” The communication right was specifically intended to reach the digital sharing of access works, for example through an Internet-enabled platform. Thus, depending on how it is locally implemented, the right of communication may be triggered by researchers sharing access to works with their colleagues and collaborators, for example through a cloud storage tool.⁴⁰

³⁶ Berne Convention for the Protection of Literary and Artistic Works art. 9 (2) and art. 10 (1), Sep. 9, 1886, as revised at Stockholm on July 14, 1967.

³⁷ TANYA APLIN & LIONEL BENTLY, *GLOBAL MANDATORY FAIR USE: THE NATURE AND SCOPE OF THE RIGHT TO QUOTE COPYRIGHT WORKS* (Cambridge University Press) (2020).

³⁸ Geiger, Christophe; Gervais, Daniel; and Senftleben, Martin. *The Three-Step Test Revisited: How to Use the Test’s Flexibility in National Copyright Law*. (2013). PIJIP Research Paper no. 2013-04.

³⁹ See Monica Torres & Raquel Xalabander, *Report on Practices and Challenges in Relation to Online Distance Education and Research Activities*, WIPO, SCCR/ 39/6, Thirty-Ninth Session (2019) at 10.

⁴⁰ Jane C. Ginsburg, *The (New?) Right of Making Available to the Public*, *INTELLECTUAL PROPERTY IN THE NEW MILLENNIUM: ESSAYS IN HONOR OF WILLIAM R. CORNISH, DAVID VAVER & LIONEL BENTLY*, EDS., CAMBRIDGE UNIVERSITY PRESS, 2004 (Columbia Public Law Research Paper No. 04-78, 2004). (“In considering the scope of the communication and making available rights under the Berne Convention and WIPO Treaties, it is necessary to recognize that neither the Berne Convention nor the WIPO Treaties define the “public” in “communication to the public” or in “making available to the public,” though it may be implicit that any group

The WCT also requires several notable expansions in the subjects of protection that are relevant to TDM research. Article 4 requires that computer programs be protected as literary works within the meaning of Article 2 of the Berne Convention. Article 5 requires that copyright protection extend to “[c]ompilations of data or other material, in any form, which by reason of the selection or arrangement of their contents constitute intellectual creations.”

The WCT contains clarifications on limitations and exceptions to exclusive rights that can be used to meet the purpose stated in its Preamble to “maintain a balance between the rights of authors and the larger public interest, particularly education, research and access to information.”⁴¹ Article 2 clarifies the limits of copyright protection, providing: “Copyright protection extends to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such.” Article 10(1) mirrors Article 9(2) of Berne, providing that “Contracting Parties may, in their national legislation, provide for limitations of or exceptions to the rights granted to authors of literary and artistic works under this Treaty in certain special cases that do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the author.” An agreed statement makes clear that Contracting Parties may “carry forward and appropriately extend into the digital environment limitations and exceptions in their national laws which have been considered acceptable under the Berne Convention.”

At bottom, international copyright treaties generally require that members protect reproductions and communications of works of the kind that are often needed in text and data mining and other research projects. Reproductions are made, for example, when a researcher creates a corpus to be mined. And communications are common in cases where multiple researchers collaborate as well as in the process of validating and disseminating results. International copyright treaties permit exceptions for these uses, but do not require them. Accordingly, providing exceptions for research purposes is an area one may expect a high level of diversity between countries, which is the subject of our study.

C. Study design

A number of studies, including those commissioned by the World Intellectual Property Organization’s Standing Committee on Copyright and Related Rights, have collected and compared copyright exceptions for various purposes from select countries.⁴² This study is most comparable to recent SCCR studies that have produced a “typology” of exceptions for specific purposes for library uses, completed by Kenneth Crews, and for educational uses, completed by Daniel Seng.⁴³ Although both of those studies includes some analysis of private use and other exceptions that could be used for research uses, neither produced a typology of research exceptions as we undertake here.

1. Creating a typology of research exceptions

Like the Crews and Seng studies, we created our typology by examining the common features

comprising the “non public” (for example, the traditional “family circle”) should be economically insignificant.”)

⁴¹ WIPO Copyright Treaty, Dec. 20, 1996. S. Treaty Doc. No. 105-17 (1997); 2186 U.N.T.S. 121; 36 I.L.M. 65 (1997) Preamble.

⁴² See, e.g. Reto Hilty and Sylvie Nerisson, *Balancing Copyright- A survey of National Approaches* (2012) (publishing scholar surveys from 38 countries).

⁴³ Kenneth Crews, *Study on Copyright Limitations and Exceptions for Libraries and Archives*, WIPO, SCCR/17/2, Seventeenth Session (2008). Daniel Seng, *Updated Study and Additional Analysis of Study on Copyright Limitations and Exceptions for Educational Activities*, WIPO, SCCR/35/5/REV, Thirty-Fifth Session (2017).

of categories of exceptions in every copyright law in the WIPO Lex database.⁴⁴ Where available, we used the official English translation provided by WIPO. Where we were not able to find an official translation, we made our own unofficial translation using Google’s translation tool.⁴⁵ We also followed the other studies in augmenting a search of WIPO Lex by checking public sources – most commonly the country’s copyright office webpage – for newly enacted laws.⁴⁶ From each law, we collected excerpts of exceptions that specifically mention “research” as a permitted purpose, those for libraries and for private uses even if they did not specifically reference “research,” open general exceptions that can be applied to a use for any purpose, and exceptions specifically for text and data mining.

The methodology we used to analyze each law consisted of collecting all relevant exceptions and reviewing them to find patterns that could be used to note differences and similarities between them. We coded each exception separately based on how open it was any research use (i.e. implicating any exclusive right), of all works, by all users. We also noted whether an exception was “general” in the sense of applying to all “research” purposes, or more specifically applied to a subset of research uses, including for computational research.

Where there were more than one research exception, we focused the analysis on the one(s) that would be most likely to permit a TDM project. For this purpose, we judged the most important factor to be the allowance to make a research copy of a whole work. If more than one exception permitted a research copy of a whole work, we included the most open exception in our study in the following order from least restrictive (1) to most restrictive (6). We associate each point along the spectrum with a color for the 3-color and 6-color maps in figures 1 and 2, below.

Classification	6-Color	3-Color
(1) fully open to all uses, works, users;	Green	Green
(2) reproduction (not other uses) of all works by all users;	Blue	Yellow
(3) private reproduction or use by individual users;	Light Blue	Yellow
(4) institutional uses;	Purple	Yellow
(5) restrictions on types of works subject to exception;	Orange	Yellow
(6) excerpts only (no TDM).	Red	Red

If an exception had more than one restriction within the same exception, then we classified it according the most restrictive element. For example, Oman limits the users of its research exception to libraries. Without more, Oman would be coded purple to signal that institutional TDM

⁴⁴ WIPO Lex Database, <https://wipo.lex.wipo.int/en/main/legislation> (last visited Sep. 7. 2021).

⁴⁵ For analysis of the use of Google’s translation tool in academic study, see Erik De Vries, Martijn Schoonvelde & Gijs Schumacher, *No longer lost in translation: Evidence that Google Translate works for comparative Bag-of-Words Text Applications*, 26 (4) POLITICAL ANALYSIS 417 (2018), <https://www.cambridge.org/core/journals/political-analysis/article/no-longer-lost-in-translation-evidence-that-google-translate-works-for-comparative-bagofwords-text-applications/43CB03805973BB8AD567F7AE50E72CA6>. On the limitations of translations, see Kenneth Crews, *Study on Copyright Limitations and Exceptions for Libraries and Archives*, WIPO, SCCR/17/2, Seventeenth Session (2008) at 17 (“Translations are often inaccurate and include misinterpretations; those deficiencies are one of the inherent limitations of a study of worldwide copyright law.”).

⁴⁶ Cf. Victor Nabhan, *Study on Limitations and Exceptions for Copyright for Educational Purposes in the Arab Countries*, WIPO, SCCR/19/6, Nineteenth Session (2009), at 8 (“Of the 19 countries selected, we are not able to obtain copies of the laws of Yemen and Mauritania, so these two laws are unfortunately not analyzed in this report.”)

may be permitted. But Oman also limits the types of works those libraries can use to “a published article or short work,” and therefore has a restriction on works that we normally code orange.⁴⁷ In such cases where two restrictions apply within the same exception, we listed the applicable color as the most restrictive (toward 6) of the two. Under this logic, we rate Oman as orange because it limits library uses to some types of works.

There are several common features of research exceptions we do not consider in our classification in this Part and in Figures 1 and 2. First, we do not alter the classification if a country restricts research uses to “non-commercial” or “non-profit” use. About half of the exceptions we analyze have such restrictions, and non-commercial restrictions exist in some countries in every color category (including green). In Part III, analyzing TDM exceptions, we include identifications of which permit commercial uses.⁴⁸ Many countries restrict research uses to a single⁴⁹ or small number,⁵⁰ of copies. But we did not consider these relevant to their applicability to a TDM project

⁴⁷ Law on Copyright and Neighboring Rights, 2008 (promulgated by Royal Decree No. 65/2008) (Oman) Art 20(3) (restricting reproductions of works “by public libraries, non-commercial documentation centers, educational establishments and scientific and cultural institutions” to “a published article or short work”)

⁴⁸ See Sean Flynn, Christopher Geiger, João Quintais, Thomas Margoni, Matthew Sag, Lucie Guibault, Michael W. Carroll, Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action, 48 (Joint PIJIP/TLS Research Paper Series, No. 3, 2020), 8 EIPR (2020) (finding it “tempting, but harmful, to restrict TDM rights to non-commercial uses”).

⁴⁹ Countries allow for a single copy in different ways, as it may be seen in Albania, Belarus, Belize, Botswana, Denmark, Dominican Republic, Eritrea, Finland, Greenland, Honduras, Malawi, Morocco, Poland, Saint Lucia, Singapore, Syria, Algeria, Andean Decision, Andorra, Azerbaijan, Bahrain, Bhutan, Brazil, Cambodia, Colombia, Costa Rica, Croatia, Dominica, Egypt, El Salvador, Ethiopia, Gambia, Jordan, Kazakhstan, Kyrgyzstan, Lebanon, Libya, Maldives, Mexico, Morocco, Oman, Palau, Papua New Guinea, Rwanda, Seychelles, Sierra Leone, Slovakia, Tonga, Trinidad and Tobago, Turkmenistan, United Arab Emirates, Venezuela. Some of them expressly mention “single copy”, “individual copy”, and “sole copy”. See, e.g., Law on Copyright and Related Rights, 1982 (Ley N° 6683, de 14 octubre de 1992, sobre el Derecho de Autor y Derechos Conexos, así reformada por la Ley N° 8834 de 3 de mayo de 2010) (Costa Rica) Article 74 (“This reproduction must be made in a single copy”); Consolidated Act on Copyright 2014 (Consolidate Act No. 1144 of October 23, 2014, on Copyright) (Den.) Section 12. (“Anyone is entitled to make or have made, for private purposes, single copies of works”); Law Concerning Copyrights and Neighboring Rights, 2002 (Federal Law No. 7 of 2002) (U.A.E.) “Art 22 (authorizing “a sole copy from the work for the merely personal and non-commercial or professional but personal use”). We also considered those that mention “a copy” and those which do not authorize multiple copies for the same work. See, e.g. Copyright Act (Cap. 252, Revised Edition 2000) (Belize) (“Article 56. Research and Private Study (1) Subject to subsection (2) and section 58, fair dealing with a protected work for the purposes of research or private study does not infringe copyright in the work. (2) Copying by a person other than the researcher or student himself is not fair dealing if - (a) in the case of a librarian, or a person acting on behalf of a librarian, he does anything which Regulations under section 66 would not permit to be done under section 67 or 68 (articles or parts of published works; restriction on multiple copies of same material)”). Law on Copyright and Related Rights, 1999 (approved by Decree No. 4-99-E, as amended by Decree No. 16-2006) (Hond.) (“Regarding copies of lawfully acquired works by a person, it is allowed without authorization of the author or remuneration, reproduction of a copy of the work for the personal and exclusive use of that person, made by him, with his own means, provided that happens in special cases, that does not attempt the normal exploitation of the work or cause unjustified damage to the author's legitimate interests.”)

⁵⁰ There were multiple terms used by countries including, but not limited to, “some”, “not big number of works”, “few”, “small number”, “limited number”, “small quantity”. See, e.g., Copyright Law, 2010 (Copyright Law of the People's Republic of China of February 26, 2010, amended up to the Decision of February 26, 2010, by the Standing Committee of the National People's Congress on Amending the Copyright Law of the People's Republic of China) (“Article 22....In the following cases, a work may be exploited without permission from, and without payment of remuneration to, the copyright owner, provided that the name of the author and the title of the work shall be mentioned and the other rights enjoyed by the copyright owner by virtue of this Law shall not be prejudiced: ... (6) translation, or reproduction in a small quantity of copies, of a published work for use by teachers or scientific researchers, in classroom teaching or scientific research, provided that the translation or reproduction shall not be published or

and so do not alter the classification of a country with a restriction on a number of copies that can be made. Finally, we did not alter the classification of a country based on whether the exception is subject to “equitable remuneration,” for example through a statutory license or equipment levy.⁵¹ For our purposes, the important aspect of the exception is that it not require individual negotiation with a rights holder, not whether a rights holder might be compensated through some other scheme.⁵²

distributed;”). Copyright Act, 1912 (Act of September 23, 1912, containing New Regulation for Copyright, as amended up to September 1, 2017) (Neth.) Art. 16(b) (authorizing “a few copies intended exclusively for personal practice, study or use by the natural person”). Copyright Act, 2018 (Act No. 40 of June 15, 2018, Relating to Copyright in Literary, Scientific and Artistic Works, consolidated version, status as at December 20, 2018) (Nor.) (“Sec 26. When this is not done for commercial purposes or on the basis of a work reproduced in violation of this Act,”). Intellectual Property Code of the Philippines (Republic Act No. 8293, 2015 Edition) (Phil.) (“Sec 185.1. The fair use of a copyrighted work for criticism, comment, news reporting, teaching including limited number of copies for classroom use, scholarship, research, and similar purposes is not an infringement of copyright.”). Law on Copyright and Neighboring Rights, 1993 (SG No. 56/1993, as amended up to December 13, 2019) (Bulg.) (“Article 24...The following shall be permissible without the consent of the copyright holder and without payment of compensation:...3. use of parts of published works or of not big number of works in other works in amount, necessary for analysis, commentary or other kind of scientific research; such use shall be permissible only for scientific and educational purposes, indicating the source and the name of the author, unless impossible;”). Slovenia is a particular case where there is a specific limit of copies allowed. Copyright and Related Rights Act, 1995 (as amended up to October 22, 2016) (Slovn.) (“Art 50 - (1) Subject to Article 37 of this Act, the reproduction of an already published work is free if it is performed in a maximum of three copies and if the conditions referred to in the second or third paragraph of this Article are met.”)

⁵¹ Copyright Law, 2000 (Law No. 65-00 on August 21, 2000) (Dom. Rep.). (“Article 37. ...It shall be lawful to reproduce once and in a single copy a literary or scientific work for personal use and not for profit-making purposes, without prejudice to the right of the right holder to obtain equitable remuneration for the reprographic reproduction or for the private copying of a sound or audiovisual recording, in the manner established under the Regulations. Computer programs shall be governed by the guidelines expressly established in the special provisions of this Law relating to such works.”). Law on Copyright and Neighboring Rights, 1993 (SG No. 56/1993, as amended up to December 13, 2019) (Bulg.): “Art. 25. (Amended, SG No. 77/2002, effective 01.01.2003) (1) (Amended, SG No. 99/2005, effective 10.01.2006)) Without the consent of the copyright holder, but upon payment of a fair compensation, the following shall be admissible: ; 2. the reproduction of works, regardless of any medium, by a natural person for his personal use, provided that it is not carried out for commercial purposes. (2) The provision of para. 1, item 2 shall not apply to computer programs and architectural works.”. Intellectual Property Law, 1996 (Texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes (aprobado por el Real Decreto Legislativo N° 1/1996 de 12 de abril de 1996, y modificado hasta el Real Decreto-Ley N° 26/2020 de 7 de julio de 2020) (Spain) (“Art 31(2) Without prejudice to the equitable compensation provided for in article 25, the reproduction, in any medium, without the assistance of third parties, of works already disclosed, does not need authorization from the author, when the following circumstances concur simultaneously, constituting the legal limit of private copying: a) That it is carried out by a natural person exclusively for their private, non-professional or business use, and without direct or indirect commercial purposes. b) That the reproduction is made from a legal source and that the conditions of access to the work or service are not violated. c) That the copy obtained is not the object of collective or lucrative use, or of distribution by price”). *See generally* Council Directive 2001/29 on the Harmonization of Certain Aspects of Copyrights and Related Rights in the Information Society, art. 5(2)(b), 2001 O.J. (L 167) 10, 16 (EU) (permitting exceptions or limitations “in respect of reproductions on any medium made by a natural person for private use and for ends that are neither directly nor indirectly commercial, on condition that the right holders receive fair compensation which takes account of the application or non-application of technological measures referred to in Article 6 to the work or subject-matter concerned.”).

⁵² Requiring individual negotiation can prohibit many TDM projects because of the massive amounts of works required for many such projects. A digital humanities project, for example, may seek to use every work of literature – implicating numerous publishers across a range of databases making individual negotiation very difficult. Some large library systems do engage in such negotiations, sometimes with considerable success.

2. *Limitations of the Typology*

It was not our purpose to create a guide for where TDM is lawful. A number of important limitations to our study prevents its use for this purpose.

Like other SCCR studies of limitations and exceptions to date, this is a study of *copyright statutes*, not the application or interpretation of those statutes in courts or enforcement bodies.⁵³ There may be interpretations of the law that broaden or narrow the application of the exceptions we classify in this Article. Where we know of such interpretations, we note them in our citations. But we did not do extensive case law or other legal research to investigate interpretation norms in every country we analyze, and did not take into account any known interpretations of the law in our classifications.⁵⁴

This is a study of copyright *exceptions*, not necessarily of other forms of limitation of the scope of rights that might impact TDM.⁵⁵ For example, most laws define copyright to cover only expression, excluding underlying facts or information that the expression can reveal, as is required among WCT members. Many TDM uses may be authorized by this expression/fact distinction.⁵⁶ But we do not analyze the framing of such exclusions.

We also analyze only *copyright* exceptions. TDM research may be affected by other forms of regulation, including “related rights” that govern performances, phonograms and broadcasts, as well as data privacy rules and *sui generis* rules for the protection of databases. We do not classify countries based on such regulations or exceptions to them in this study.

Based on these and other limitations, this study is emphatically not legal advice. It does not tell any researcher what they can or cannot do in a TDM or other research project in any country that we classify. With that important qualification in place, the next Part analyzes the results of our

⁵³ See Joseph Fometeu, *Study on Limitations and Exceptions for Copyright and Related Rights for Teaching in Africa*, WIPO, SCCR/19/5, Nineteenth Session (2009) at 4 (“One should bear in mind that even in the field of intellectual property, which is strongly influenced by a host of multilateral agreements, national texts inevitably remain immersed in domestic socio-juridical contexts which justify some of their provisions, making it somewhat difficult for an outsider to understand them fully. Accordingly, the author wishes to beg the reader’s indulgence for any misinterpretation of a domestic law.”).

⁵⁴ This methodology differs from other studies such as Flynn and Palmedo, *supra* (basing the rating of countries on expert surveys of change in statutes or interpretations by courts or other authorities); Hilty and Nerisson, *supra* (analyzing the exceptions of countries including matters of interpretation without ranking countries based on that analysis).

⁵⁵ By “exceptions” this study will use the definition proposed by Professor Sam Ricketson: “Provisions that allow for the giving of immunity (usually on a permissive, rather than mandatory, basis) from infringement proceedings for particular kinds of use, for example, where this is for the purposes of news reporting or education, or where particular conditions are satisfied. These can be termed “permitted uses,” or exceptions to protection, in that they allow for the removal of liability that would otherwise arise.” Sam Ricketson, *WIPO Study on Limitations and Exceptions of Copyright and Related Rights in the Digital Environment*, WIPO, SCCR/9/7 (2003), at 3-4.

⁵⁶ Many scholars argue that the act of analyzing text or data is not itself a copyright protected activity, and therefore should not need the operation of an exception. Rossana Ducato & Alain Strowel, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*. CRIDES Working Paper Series no. 1/2021; forthcoming in 43 EIPR 2021/5, 322 (2021); Christophe Geiger, Giancarlo Frosio and Oleksandr Bulayenko, *Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?*, 49(7) IIC - INT’L REV. OF INTELLECTUAL PROP. & COMPETITION L. 814, 817 (2018) (explaining that requiring permission to analyze information could be considered an unlawful “restriction of freedom of expression and information as protected by e.g., the European Court of Human Rights (ECHR) and the Charter of Fundamental Rights of the European Union”). See generally Jessica Litman, *The Exclusive Right to Read*, 13 CARDOZO ARTS & ENT. L.J. 29 (1994).

typology.

II. A TYPOLOGY OF RESEARCH EXCEPTIONS

The results of our classifications of research exceptions through the methodology described in Part I are presented below. Figures 1 and 2 present the results in the form of two maps of the world. Figure 1 presents a 3-color map where green represents “go” (open to TDM uses); yellow urges “caution” (there are restrictions on the uses, works, or users subject to the exception); and red means “stop” (exceptions permit uses only of excerpts of works).

Figure 1.

Research Exceptions in Comparative Law: Three Color

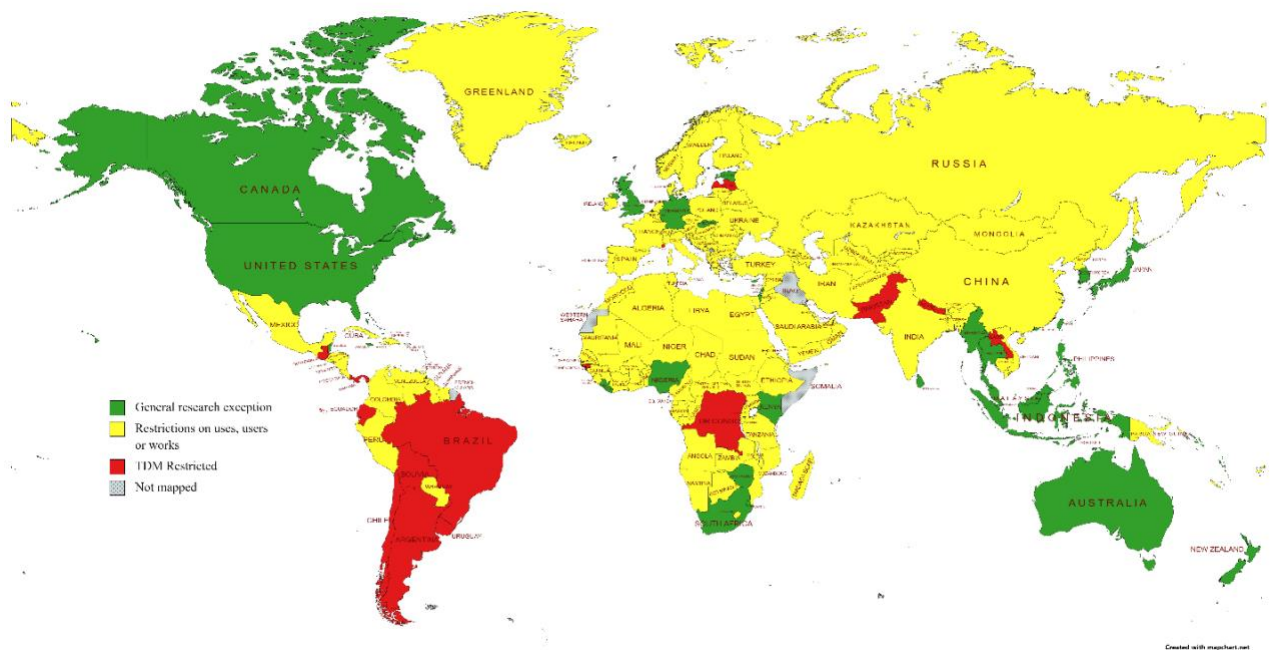
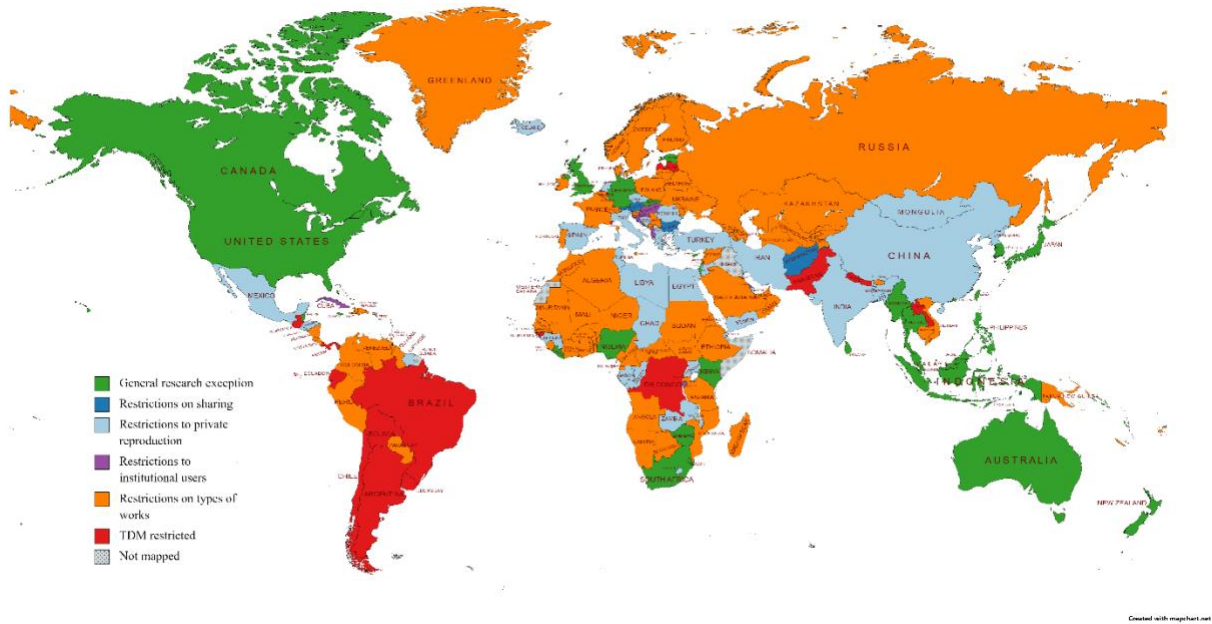


Figure 2 breaks the yellow / caution category into its elemental parts, showing which countries have restrictions on uses (blue), users (purple), or the works (orange) that may be used. Figure 2 also identifies countries (in light blue) that limit research uses to individuals, such as through “private” or “personal” use exceptions.

Figure 2.

Research Exceptions in Comparative Copyright: Six Color



The remainder of this Part describes each of the categories we used to classify exceptions in more detail, with illustrative examples from statutes we reviewed and tables identifying all the countries we classified into each category.

A. Open Research Exceptions

The most open research exceptions are labeled green in our maps. All of these countries have either a general exception for “research”, an open general exception for use for any purpose, or a specific TDM exception that can apply to reproductions and communications of all copyrighted works by any user.⁵⁷ The countries in the green category in our study are listed below in Table 1.

Table 1: Countries/Regions Coded Green (42)

Antigua and Barbuda	Australia	Bahamas	Barbados	Belize
Brunei Darussalam	Canada	Cyprus	Estonia	Eswatini (Swaziland)

⁵⁷ As discussed above, we use the term “general” to refer to an exception that apply to more than one purpose of use. In *The User Rights Database* we classified exceptions for research as specific, not general, because they list one purpose. An exception for research, private use, and criticism and review would be general because it lists several different purposes. Here we find that generality is a question of degree. As compared to a TDM exception, an exception for any form of research, including TDM, appears general.

Germany	Grenada	Hong Kong SAR	Indonesia	Israel
Jamaica	Japan	Kenya	Liberia	Liechtenstein
Malaysia	Mauritius	Myanmar	Nauru	New Zealand
Nigeria	Philippines	Republic of Korea	Saint Kitts and Nevis	Saint Lucia
Saint Vincent and The Grenadines	Singapore	Slovakia	Solomon Islands	South Africa
Sri Lanka	Taiwan	Thailand	United Kingdom	United States of America
Vanuatu	Zimbabwe			

Open exceptions for research are common, but not universal, in countries with a fair use or fair dealing exception influenced by British law.⁵⁸ Belize provides an example of a commonly occurring “fair dealing” exception that applies to any protected use (aka “dealing”) with any “work” by any user for a research purpose:

Article 56. Research and Private Study

⁵⁸ These countries are referred to in different ways in the literature. Many distinguish between “common law” (as opposed to civil law) or “copyright” (as opposed to authors rights) countries. See Goldstein and Hugenholtz, 14-21 (defining copyright legal traditions). Countries in the open research exception category that have a fair use or fair dealing exception include: Copyright Act, 2003 (Act. no. No. 22 of 2003) (Ant. & Barb.). Copyright Act, 1968 (Act No. 63, 1968, consolidated as of January 1, 2019) (Austl.). Copyright Act, 1998 (CH.323) (Act No. 2 of 2004, as amended by the Copyright (Amendment) Act,) (Bah). Copyright Act, 1998 (Chapter 300, as revised up to 2006) (Barb.). Copyright Act (Cap. 252, Revised Edition 2000) (Belize). Copyright Act, 1999 (Emergency (Copyright) Order, 1999, Constitution of Brunei Darussalam, Order under section 83(3)). Copyright Act, 1985 (R.S.C., 1985, c. C-42, as amended up to June 22, 2016) (Can.). Copyright Act, 1912 (The Copyright Act) (Eswatini). Copyright Ordinance - CAP 528 (An Ordinance to make provisions in respect of copyright and related rights and for connected purposes, amended 15 of 2007 s. 3) (H.K.). Copyright Act, 2007 (as amended on July 28, 2011) (Isr.). Copyright Act, 1993 (Act 5 of 1993), (as amended by The Copyright (Amendment) Act no. 13, 2015) (Jam.). Copyright Act, 2001 (Act No. 12 of 2001, as amended up to Act No. 11 of 2017) (Kenya). Liberia Intellectual Property Act, 2016 (An Act to Repeal an Act Adopting a New Copyright Law of the Republic of Liberia approved July 23, 1997; and the Industrial Property Act of Liberia approved March 20, 2003, constituting Title 24 of the Liberian Code of Laws Revised, and to enact in their stead a New Title 24 to be known as the “Liberia Intellectual Property Act, 2016”) (Liber.). Copyright Act, 1987 (Act 332, as at 1 July 2012) (Malay.). Copyright Act, 1911 (The Burma Copyright Act. India Act III, 1914) (Myan.). Copyright Act, 1994 (Public Act 1994 No 143, reprint as at 4 January 2020) (N.Z.). Copyright Act, C. 28 (as codified 2004) (Nigeria). Intellectual Property Code of the Philippines (Republic Act No. 8293, 2015 Edition) (Phil.). Copyright Act, 2000, c. 18.08 (Copyright Act, Rev. ed. showing the law as at 31 December 2002) (St. Kitts & Nevis). Copyright Act 1995, c. 13.07 (Act no. 10 of 1995, as amended by Act No. 7 of 2000) (St. Lucia). Copyright Act, 2003 (Act n. 21 of 2003) (St. Vincent and The Grenadines). Copyright Act 2021 (Revised Edition 2020, Act No. 22 of 2021) (Singapore). Copyright Act 1987 (Cap 138) (Solomon Islands). Copyright Act, 1978 (Act No. 98 of 1978, as amended up to Copyright Amendment Act 2002) (S. Afr.). Intellectual Property Act, 2003 (Intellectual Property Act No. 36 of 2003) (Sri Lanka). Copyright, Designs and Patents Act 1988, c. 48 (incorporating amendments up to the Digital Economy Act 2017) (UK.). Copyright Law of the United States and Related Laws Contained in Title 17 of the United States Code, 17 U.S.C. § 107 (2020). Copyright and Related Rights Act n. 42 of 2000 (Vanuatu.). Copyright and Neighbouring Rights Act 2000, c. 26:05 (Act 11/2000, as amended up to Act No. 32 of 2004) (Zim.)

(1) Subject to subsection (2) and section 58, fair dealing with a protected work for the purposes of research or private study does not infringe copyright in the work.⁵⁹

In other countries, including, but not limited to, the United States, the general copyright exception authorizes a “use” or “utilization” rather than “dealing.” But in this limited sense, fair use and fair dealing are the same thing. Both apply to any activity with a work that may be protected by copyright.

The U.S. fair use exception, which specifically mentions uses for “scholarship, or research” among its illustrative purposes, has been interpreted to allow instances of TDM.⁶⁰ But we do not require such interpretations to include a country in the green category. Rather, as explained above, we classify each country based on the face of the statute alone.

Many broad categories we often use to categorize limitations and exceptions do not hold in our analysis. Scholars and analysts often distinguish between “common law” countries with “fair use” and “fair dealing” exceptions and “civil law” countries with “closed lists” of exceptions. The implication is often that fair use and fair dealing exceptions are more flexible and permissive than the closed lists. But this characterization does not hold for all the countries in our analysis.

We do not find that all exceptions for research that use the terms “fair use” or “fair dealing” are more open than all civil law exceptions. Some fair use and fair dealing exceptions are only applicable to a “private” or “personal” use – drawing the same classification (light blue in Figure 2) as many civil law countries.⁶¹ Other countries with British influenced copyright law, including Guyana, Ireland, and Namibia, limit application of their fair dealing exceptions to limited classes of works, and thus are coded orange in Figure 2.⁶² Pakistan appears red in our maps because its exception for fair dealing for “research or private study” is qualified by an “explanation” that appears to limit uses to extracts.⁶³ Finally, many “common law” countries don’t have general

⁵⁹ Copyright Act (Cap. 252, Revised Edition 2000) (Belize)

⁶⁰ Copyright Law of the United States and Related Laws Contained in Title 17 of the United States Code, 17 U.S.C. § 107 (2020). See *Authors Guild v. Google, Inc.*, 804 F.3d 202, 215 (2d Cir. 2015); *Authors Guild, Inc. v. HathiTrust* 755 F.3d 87, 105 (2d Cir. 2014).

⁶¹ India, Bangladesh, Uganda, and Zambia, for example, are coded in light blue because they permit a “fair dealing” or “fair use” only if “private or personal”. Copyright Act, 1957 (Act No. 14 of 1957, as amended up to Act No. 27 of 2012) (India) Art. 52(1) (permitting “a fair dealing with any work, not being a computer programme, for the purposes of— (i) private or personal use, including research”); Copyright Act, 2000 (Act No. 28 of 2000, as amended up to 2005) (Bangl.) Art. 72 (permitting “Fair use of a literary, dramatic, musical or artistic work for the purpose of- (i) private study or private use including research”); The Copyright and Neighbouring Rights Act, 2006 (Uganda) Art 15(1) (permitting “fair use of a protected work ... where— (a) the production, translation, adaptation, arrangement or other transformation of the work is for private personal use only”); The Copyright and Performance Rights Act, 1994 (Act No. 44 of 1994) (Zam.) Art 21(1) (permitting “fair dealing with a work for private study or for the purposes of research done by an individual for his personal purposes, otherwise than for profit”).

⁶² Copyright Act 1956, 4 & 5 Eliz. 2, c. 74 (Guy.) Section 6 (applying the fair dealing exceptions only to “a literary, dramatic or musical work for purposes of research,” while protecting other works); Copyright and Related Rights Act, 2000 (Act No. 28 of 2000) (Ir.) Article 50 (applying fair dealing exception for research to works including a “non- electronic original database,” thus excluding application to an electronic original database); Copyright and Neighbouring Rights Protection Act, 1994 (Act No. 6 of 1994) (Namib.) Art 15.(1) (applying fair dealing exception only to “a literary or musical work” and not other forms of protected works).

⁶³ The provision could also be read as a safe harbor that might not restrict a different fair dealing: “Explanation.—For the purposes of clause (a) or clause (b) of this subsection—(i) in relation to a literary or dramatic work in prose, a single extract up to four hundred words, or a series of extracts (with comments interposed) up to a total of eight hundred words with no one extract exceeding three hundred words; and (ii) in relation to a literary or dramatic work in poetry, an extract or extracts up to a total of forty lines and in no case exceeding one-fourth of the whole of any poem may be deemed to be fair dealing with such work: Provided that in a review of a newly published work reasonably longer extracts may be

exceptions for fair use or fair dealing at all, leaving them functionally indistinguishable from a civil law closed list system.⁶⁴

On the other hand, fair use and fair dealing countries are not the only ones with open research exceptions. The EU's 2001 "InfoSoc Directive" permits countries to adopt an open exception for research within European "closed list" systems of exceptions.⁶⁵ Liechtenstein's law, for example, provides an open scientific research exception applying to any "use" of any work by any user "insofar as this is justified for the pursuit of non-commercial purposes and if possible the source and the name of the author are given."⁶⁶ This civil law research exception is thus more open than many exceptions for fair use or fair dealing.

Fair use and fair dealing countries are also not the only ones with fully open, general, flexible exceptions – i.e. exceptions like U.S. fair use that can apply to any use, of any work, by any user, *for any purpose*. Thailand provides similar openness without using the fair use or fair dealing framing by permitting any use that complies with the terms of the international 3-step test.⁶⁷ Japan's Article 30-4 makes it permissible to "exploit" any work by any user that "is not for enjoying or causing another person to enjoy the ideas or emotions expressed in such work" – a general standard that includes, but is not limited to, "exploitation for using the work in a data analysis."⁶⁸

deemed fair dealing with such work."

⁶⁴ Some countries with a British-influenced legal tradition, including Botswana, Malawi, Ghana, and Tanzania, do not have fair dealing or fair use exceptions at all, making their exceptions appear closer to many civil law countries from mainland Europe. Copyright and Neighboring Rights Act 2000, c. 68:02 (Act No. 8 of 2000, as amended by Act No. 6 of 2006) (Bots.). Copyright Act, 2016 (Act No. 26 of 2016) (Malawi). Copyright Act, 2005 (Act 690, 2005) (Ghana). Copyright and Neighbouring Rights Act, 1999 (Act No. 7 of 1999) (Tanz.).

⁶⁵ Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (authorizing exceptions for "use" of any work for "illustration for teaching or scientific research ... to the extent justified by the non-commercial purpose to be achieved").

⁶⁶ Law on Copyright and Neighboring Rights, 1999 (consolidated version of December 19, 2006) (Liech.): Article 22. Privileged uses of the work.

1) Published works may be used for special purposes. A special purpose is:

a) any use of the work in the personal sphere and in the circle of persons who are closely related, such as relatives or friends;

b) the use of the work for illustration in class or for scientific research insofar as this is justified for the pursuit of non-commercial purposes and if possible the source and the name of the author are given;

c) the reproduction of the work on paper or a similar medium by means of photomechanical processes or other processes with a similar effect for educational purposes, for scientific research or for internal information and documentation in companies, public administrations, institutes, commissions and similar institutions;

d) digital reproduction for educational purposes and for scientific research without any direct or indirect economic or commercial purpose."

⁶⁷ Copyright Act, 1994 Section 32 (Copyright Act B.E. 2537, 1994) (Thai.):

32. An act against a copyright work under this Act of another person which does not conflict with normal exploitation of the copyright work by the owner of copyright and does not unreasonably prejudice the legitimate rights of the owner of copyright shall not be deemed an infringement of copyright.

Subject to the provision in the first paragraph, the following acts in relation to a copyright work shall not be deemed an infringement of copyright:

(1) research or study of the work which is not for profit;

(2) use for personal benefit or for the benefit of the user and his family members or close relatives;

(3) comment, criticism or introduction of the work with an acknowledgment of the ownership of copyright in such work

⁶⁸ Copyright Act, 1970 (Act No. 48 of May 6, 1970, as amended up to Act No. 72 of July 13, 2018) (Japan): Article 30-4. Exploitations not for enjoying the ideas or emotions expressed in a work.

It is permissible to exploit work, in any way and to the extent considered necessary, in any of the following cases or

One close call in our classification is presented by the case of Germany. Germany’s general research exception would merit an orange classification because it only allows uses “of illustrations, isolated articles from the same professional or scientific journal, other small-scale works and out-of commerce works.”⁶⁹ But we classify Germany as green in our map based on its specific text and data mining exception which applies to reproductions and communications of any work “to enable the automatic analysis of large numbers of works (source material) for scientific research.”⁷⁰

At the opposite end of the spectrum from the green countries are those whose research exceptions extend only to the use of quotations or excerpts.

B. Closed Exceptions

We color in red in Figures 1 and 2 those countries that specifically limit research exceptions to uses of excerpts of works, which is the minimum standard for limitations and exceptions required by Article 10(1) of the Berne Convention. TDM research must analyze whole works. The quotation right may be useful in sharing results of the research to the public. But to find the facts and data that TDM projects seek, whole works are normally reproduced into the corpus to be mined. Accordingly, quotation-only exceptions are inadequate to enable the activity.⁷¹ Table 2 lists the countries in this category.

other cases where such exploitation is not for enjoying or causing another person to enjoy the ideas or emotions expressed in such work; provided, however that this does not apply if the exploitation would unreasonably prejudice the interests of the copyright owner in light of the natures and purposes of such work, as well as the circumstances of such exploitation:

...

(ii) exploitation for using the work in a data analysis (meaning the extraction, comparison, classification, or other statistical analysis of language, sound, or image data, or other elements of which a large number of works or a large volume of data is composed; the same applies in Article 47-5, paragraph (1), item (ii).

⁶⁹ Act on Copyright and Related Rights, 1965 (Copyright Act, as amended up to Act of September 1, 2017) (Ger.)

Section 60c. Scientific research (1) Up to 15 per cent of a work may be reproduced, distributed and made available to the public for the purpose of non-commercial scientific research 1. for a specifically limited circle of persons for their personal scientific research and 2. for individual third persons insofar as this serves the monitoring of the quality of scientific research. (2) Up to 75 per cent of a work may be reproduced for personal scientific research. (3) In derogation from subsections (1) and (2), full use may be made of illustrations, isolated articles from the same professional or scientific journal, other small-scale works and out-of commerce works. (4) Subsections (1) to (3) do not authorise the recording of the public recitation, performance or presentation of a work onto a video or audio recording medium and the subsequent making available to the public of that recording.

⁷⁰ Act on Copyright and Related Rights, 1965 (Copyright Act, as amended up to Act of September 1, 2017) (Ger.) (“Section 60d. Text and data mining. (1) In order to enable the automatic analysis of large numbers of works (source material) for scientific research, it shall be permissible: 1. to reproduce the source material, including automatically and systematically, in order to create, particularly by means of normalisation, structuring and categorisation, a corpus which can be analysed and 2. to make the corpus available to the public for a specifically limited circle of persons for their joint scientific research, as well as to individual third persons for the purpose of monitoring the quality of scientific research. In such cases, the user may only pursue non-commercial purposes.”).

⁷¹ It is not necessarily true that a TDM project would be illegal in each of these countries. It may be that a limitation on the scope of copyright protection could enable some TDM uses. See Rossana Ducato et al, *supra*; Thomas Margoni & Martin Kretschmer, *A deeper look into the EU Text and Data Mining exceptions: Harmonisation, data ownership, and the future of technology*, ZENODO (2021) (describing and embracing the argument that TDM uses do not implicate copyright’s exclusive protections).

Table 2: Countries/Regions Coded Red (15)

Argentina	Bolivia	Brazil	Chile	Congo, Dem. Rep.
Ecuador	Guatemala	Guinea-Bissau	Laos PDR	Latvia
Monaco	Nepal	Panama	Pakistan	Uruguay

Perhaps the most restrictive exception in our sample is from Argentina, which authorizes users only to “publish” excerpts of specifically identified works of specific lengths:

Argentina

Article 10. Any person may publish, for didactic or scientific purposes, comments, criticisms or notes referring to intellectual works, including up to 1,000 words for literary or scientific works, or eight bars in musical works and, in all cases, only the parts of the text essential for that purpose.

This provision shall cover educational and teaching works, collections, anthologies and other similar works.

Where inclusions from works by other people are the main part of the new work, the courts may fix, on an equitable basis and in summary judgment, the proportional amount to which holders of the rights in the works included are entitled.⁷²

Argentina and some other countries in our red category appear to be modeled on pre-1967 versions of the Berne Convention and therefore only grant permission to “publish” a quotation, offering no permission for reproductions in the research process.⁷³ Other laws in our red category, such as Brazil, authorize reproductions for research uses, but limit those uses to excerpts.⁷⁴ Guatemala and Uruguay authorize reproduction of excerpts only “for teaching,” lacking any exception for a research or personal use.⁷⁵

⁷² Law No. 11.723 of September 28, 1933, on Legal Intellectual Property Regime (Argentina), as amended up to Law No. 26.570 of November 25, 2009.

⁷³ *Accord* Law on the Protection of Literary and Artistic Property, 1948 (Law No. 491 of November 24, 1948, as amended up to Law No. 1.313 of June 29, 2006) (Monaco), <https://wipolex.wipo.int/en/legislation/details/9056>, Article 16. (Quotation) (“Monaco, for example, provides only that “It is permissible to publish borrowings from literary or artistic works, provided that the source and the author are indicated when these publications are of a scientific or academic nature or constitute chrestomathies.”). *See also* Law on Copyright, 1992 (Law No.1322 of April 13, 1992) (Bol.), <https://wipolex.wipo.int/en/text/225957> Article 24. (Quotation) (“An author may be quoted, whereby quotation shall be understood to mean the inclusion within one’s own work of short excerpts from works by others”)

⁷⁴ Law on Copyright and Neighboring Rights, 1998 (Law No. 9.610 of February 19, 1998, as amended by Law No. 12.853 of August 14, 2013) (Braz.), (“article 46. ...”Does not constitute copyright infringementII. reproduction, in a single copy of short excerpts, for the private use of the copyist, provided that it is made by the latter, without profit intention”); Ordinance-Law No. 86-033 of April 5, 1986 on the Protection of Copyright and Neighboring Rights (Dem. Rep. Congo), Art.24 (Quotation) (“It shall be lawful to reproduce quotations or excerpts of protected works for cultural, scientific, teaching, critical or polemic purposes, provided that the source, title and name of the author are mentioned.”); Law on Copyright and Neighboring Rights, 2012 (Law No. 64 of October 10, 2012) (Pan.), Article 68. (“Regarding works, services or productions already lawfully disclosed, the following are allowed without authorization ... 2. The reprographic reproduction of a legitimate example for exclusive personal use provided that it is limited to small parts of the protected work or to sold-out works.”).

⁷⁵ Law on Copyright and Related Rights, 1998 (Decree No. 33-98, as amended up to Decree No. 11-2006 of the Congress of the Republic) (Guat.), Article 64. (Reproduction exceptions) (authorizing “reproduction by reprographic means of articles or brief excerpts from works lawfully published, for teaching or conducting examinations in educational institutions,” by libraries to preserve and replace items in their collection, for judicial or administrative proceedings, and “for personal use of a work of art exhibited permanently in public”); Law on Literary and Artistic

Although most of the countries in this category are developing countries with a civil law legal tradition, two exceptions stand out. Pakistan is a common law country with a fair dealing exception. But is coded red because an explanation in the statute appears to confine its scope to uses of excerpts.⁷⁶ Chile is the rare member of the Organisation for Economic Co-operation and Development that appears on the red list. Chile recently revised its copyright limitations and exceptions and made them more useful for many modern purposes. But its primary exceptions for research uses, contained in its library exceptions, remains restricted to the use of “fragments of works”⁷⁷ or under “conditions guaranteeing that the users are unable to make electronic copies of such reproductions.”⁷⁸

C. Restrictions on Uses, Works and Users

Between the poles of open and closed research exceptions lie the plurality of countries in the world. In these countries, some reproductions of whole works of some types by some users can often be made. This may be sufficient to permit some kinds of TDM projects. But restrictions on communications, on uses by some users, or uses of some kinds of works complicate many needed research activities. We explain each of these categories and the countries within them below.

Property, 1937 (Law No. 9.739 of December 17, 1937, as amended up to Law No. 18.046 of October 24, 2006) (Uru.), Article 45. Exceptions for Reproduction (excepting “publication or broadcast” of “extracts, fragments of poetry and separate articles” “for teaching,” and “transcriptions intended for comments, reviews or polemics”).

⁷⁶ The Copyright Ordinance, 1962 (Act No. XXXIV, as amended by Copyright (Amendment) Ordinance, 2000 dated 29 September 2000) (Pak.) (“57. ... Explanation.—For the purposes of clause (a) or clause (b) of this subsection—(i) in relation to a literary or dramatic work in prose, a single extract up to four hundred words, or a series of extracts (with comments interposed) up to a total of eight hundred words with no one extract exceeding three hundred words; and (ii) in relation to a literary or dramatic work in poetry, an extract or extracts up to a total of forty lines and in no case exceeding one-fourth of the whole of any poem may be deemed to be fair dealing with such work: Provided that in a review of a newly published work reasonably longer extracts may be deemed fair dealing with such work.”)

⁷⁷ Chile’s copyright law includes:

Artículo 71 J. Libraries and archives that are not for profit may, without requiring authorization from the author or owner, or payment of any remuneration, make copies of fragments of works that are in their collections, at the request of a user of the library or archive exclusively for your personal use.

⁷⁸

Article 71K. Not-For-Profit libraries or archive centers may, without having to seek the authorization of the author or copyright holder or make any form of payment, electronically reproduce works in their collection for free and simultaneous consultation by a reasonable number of users, only in the computer terminal networks of the respective institution and on conditions guaranteeing that the users are unable to make electronic copies of such reproductions.

1. *Reproduction Only*

Some laws provide research exceptions that are open in their application to all works by all users, but restrict the exclusive rights covered to only reproduction. The exception thus does not enable communications of works by researchers, which may pose barriers to collaboration, validation, and communication of results. We color these countries blue in Figure 2, which are listed in Table 3, below.

Table 3. Countries/Regions Coded Blue (4)

Afghanistan	Austria	Bulgaria	Samoa	
-------------	---------	----------	-------	--

Our blue category only includes a few countries because most countries that limit a research exception to reproduction also restrict the exception in other ways, for example that the use be by an individual, e.g. for a “private” or “personal” use.

As described above, EU countries have the freedom to adopt research exceptions that are coded green in our map by taking full advantage of the 2001 Information Society Directive (“InfoSoc”) permitting exceptions or limitations for “use” for scientific research.⁷⁹ Most EU countries do not, however, use this flexibility fully. Austria, for example, permits only “individual copies of a work” for research purposes – a term that does not appear to extend to communications of those copies.⁸⁰

Afghanistan has perhaps the broadest exception in this category. It has a general exception for any “copying and replication of a work ... If the user does not make direct or indirect commercial gains.”⁸¹ This exception is thus open to application to any purpose, work, and user. But the use permitted is only reproduction – earning it the blue categorization in our typology.

2. *Restrictions to Institutional Users*

We color in purple in Figure 2 those countries whose only research exception that permits reproductions of whole works is limited to institutions, not individuals.⁸² Most commonly the beneficiary institutions include non-profit or public libraries or other non-profit organizations.⁸³

⁷⁹ European Parliament and Council Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society, 2001 O.J. (L167).

⁸⁰ Copyright Act, 1965 (Federal Law on Copyrights on Literary and Artistic Works and Related Rights, as last amended by Federal Law Gazette (BGBl) I No. 99/2015). § 40h clarifies that the exception “applies to database works with the proviso that duplication on paper or a similar medium is also permissible..”).

⁸¹ Afghanistan’s Copyright Act provides:

Art 40. (2) Copying and replication of a work for the purpose of using it, is lawful under the following conditions:

1. If the user does not make direct or indirect commercial gains.

⁸² See e.g. Copyright Law, 1990 (Law No. 101/III/90 of December 29; Decree-Law No. 1/2009 of April 27, 2009, revising the Law on Copyright) (Cape Verde), Article 62(1)(b) (permitting reproduction “for exclusively didactic, research or professional training purposes, by libraries, archives and noncommercial documentation centers, scientific institutes or teaching establishments, as long as the copies reproduced do not exceed the needs of the purpose for which they are intended”).

⁸³ See e.g. Law on Copyright and Related Rights, 2016 (Law No. 35/2016 of March 31, 2016) (Alb.)Article 75 (applying to “Public archives, national libraries, educational and scientific institutions, preschool educational institutions and social institutions, that have no direct or indirect economic benefits”); Law on the Protection of Copyright and Related Rights in Burundi, 2005 (Law No. 1/021 of December 30, 2005) (Burundi), Article 26(1)(a) (Article 26(5) (“public libraries, non-commercial documentation centers, scientific institutions and educational establishments”); Copyright Act, 2018 (Act No. 8 of 2018) (Kiribati), , Article 20. Copying by Cultural Institution

Many of these countries also have private reproduction rights for use by individuals that are more restrictive in the kinds of works that may be used than the rights given to institutions.⁸⁴ The countries we reviewed with open research exceptions that only apply to institutional uses appear in table 5, below.

Table 5. Countries/Regions Coded Purple (10)

Albania	Bahrain	Burundi	Cabo Verde	Cook Islands
Croatia	Cuba	European Union	Hungary	Kiribati

Many of the provisions in the purple classification specifically permit sharing reproductions with the users of the library – which may be an especially useful provision for researchers. Kiribati, for example, provides:

A cultural institution does not infringe copyright in a work in its collection by copying the work for the purpose of allowing access to that copy by users of the institution whether for personal use or study on the institution's premises (with or without technical equipment) or by way of a loan.⁸⁵

We only classified a country as purple if the exception applies to all kinds of works. Some laws permit relatively broad rights to libraries and other institutions but contain restrictions on applications to important classes of works. Sweden, for example, has a very broad exception for libraries and other public institutions to make “copies of works, *other than computer programs*.”⁸⁶ We were unclear on the extent to which an exclusion only of computer programs from the scope of the exemption may implicate important TDM projects. In this version of the Article we are classifying such exclusion in the Orange category.

We did not count as purple an exception that permits reproductions of full works only for display on computer terminals on the premises of the library. Such exceptions are common in the

(“In this section, “cultural institution” means a library, archive, museum, or gallery that is publicly funded in whole or in part.”).

⁸⁴ Compare Law on Copyright and Related Rights, 2016 (Law No. 35/2016 of March 31, 2016) (Alb.), Article 72. The reproduction of the works for private and personal use (restricting private use of “an entire book,” “graphic editions of the musical works (scores),” “electronic databases,” “cartographic works,” “architectural structures”, and “computer programs”), with Article 75. Restrictions on the Benefit of Special Institutions (non-profit institutions “may reproduce the work from their copy for internal use, by any means”). Accord Law on the Protection of Copyright and Related Rights in Burundi, 2005 (Law No. 1/021 of December 30, 2005) (Burundi), Article 26(1)(a) (closed private reproduction); *id.* Article 26(5) (open research exception for “public libraries, non-commercial documentation centers, scientific institutions and educational establishments”); Copyright Act, 2013 (Cook Islands) Article 14(1) (closed private reproduction right); *id.* Article 14(5) (“A cultural institution does not infringe copyright in a work in its collection by copying the work for the purpose of allowing access to that copy by users of the institution whether for personal use or study on the institution's premises (with or without technical equipment) or by way of loan.”); Copyright and Related Rights Act and Acts on Amendments to the Copyright and Related Rights Act (OG Nos. 167/2003, 79/2007, 80/2011, 141/2013 & 127/2014) (Croat.), Articles 82 (closed private use) and 84 (open research reproduction for “benefit of particular institutions”); Copyright Law, 1977 (Law No. 14 of December 28, 1977, as amended by Decree-Law No. 156 of September 28, 1994) (Cuba) Article 38 (“d) reproduce a work by a photographic or other analogous procedure, when the reproduction is made by a library, a documentation center, a scientific institution or an educational establishment, and provided that it is done on a non-profit basis and that the number of copies is strictly limited to the needs of a specific activity;”)

⁸⁵ Copyright Act, 2018 (Act No. 8 of 2018) (Kiribati), Article 20(4).

⁸⁶ Act on Copyright in Literary and Artistic Works, 1960 (Act (1960:729) on Copyright in Literary and Artistic Works (as amended up to Act (2018:1099)) (Swed.), Article 16 (proving exception for “governmental and municipal archival authorities, the scientific and research libraries operated by the community at large, and the public libraries”).

EU, for example, because of Article 3(n) of the EU InfoSoc Directive, which permits “use by communication or making available, for the purpose of research or private study, to individual members of the public by dedicated terminals on the premises of” libraries and other institutions. We assume that TDM projects need to be able to make reproductions and use materials outside the physical premises of supporting institutions.

3. Restrictions to Individual Users

We label in light blue in Figure 2 those countries with laws that provide a research exception that is restricted to reproductions by individuals, excluding uses by libraries or other institutions. The countries in this category are listed in Table 4, below.

Table 4. Countries/Regions Coded Light Blue (38)

Andean Decision	Bangladesh	Bosnia and Herzegovina	Chad	China
Congo, Rep.	Czech Republic	Democratic People's Republic of Korea (N. Korea)	Djibouti	Egypt
El Salvador	Gabon	Greece	Guinea	Honduras
Iceland	India ⁸⁷	Iran (Islamic Republic of)	Italy	Jordan
Lesotho	Libya	Luxembourg	Malawi	Mexico
Mongolia	Netherlands	North Macedonia	Qatar	Romania
San Marino	Spain	Suriname	Tunisia	Turkey
Uganda	Yemen	Zambia		

The most common of these exceptions extend to research uses as a category of “private” or “personal” use. By virtue of the use of the term “private” or “personal,” we assume that none of these exceptions authorize sharing with other researchers, and also do not extend to commercial or institutional uses.

The interpretation of many of the exceptions in this category is difficult by virtue of the qualification to a “private” use is often separated by a comma from a “research” use. And thus one must determine if “private” is meant to qualify the research use or whether the permitted use of research stands alone. For example, we interpreted China’s research exceptions in Article 22(1) and (6) to both apply only to individuals by virtue of the other examples that appear in their context:

Article 22. In the following cases, a work may be exploited without permission from, and without payment of remuneration to, the copyright owner, provided that the name of the author and the title

⁸⁷ We are aware that India may have a broader interpretation of its exception for fair dealing than we give it here. See M. P. Ram Mohan & Aditya Gupta, Right to Research and Copyright Law: From Photocopying to Shadow Libraries in SYMPOSIUM ON THE RIGHT TO RESEARCH IN INTERNATIONAL COPYRIGHT LAW 35-36 (Apr. 21-22, 2022) (providing that “both the established interpretation of inclusive definitions and the constitutional basis of the right to research requires the Courts to interpret Section 52(1)(a)(i) in its broadest possible enunciation.”). However, as stated before, our analysis considered only the plain text reading of the statute and therefore we did not alter our classification of India based on this information.

of the work shall be mentioned and the other rights enjoyed by the copyright owner by virtue of this Law shall not be prejudiced:

(1) use of a published work for the purposes of the user's own private study, research or self-entertainment;

...

...(6) translation, or reproduction in a small quantity of copies, of a published work for use by teachers or scientific researchers, in classroom teaching or scientific research,

In 2021, China renumbered its exceptions but continues to use phrasing that suggests that the research exception is an individual right:

(1) use of a published work of another for purposes of personal study, research or appreciation⁸⁸

The 2021 amendment adds a general opening clause in Article 24(13), permitting uses in “other circumstances as provided by laws and administrative regulations.”⁸⁹ We have not examined such regulations to determine if China has authorized TDM uses.

4. *Restrictions on Types of Works*

The most common restriction in research exceptions is a limitation on the kinds of works that can be used. Many of these laws exclude application to whole books, databases, computer programs, fine art, architectural works, or musical scores.⁹⁰ A few laws in this category only exclude databases or computer programs.⁹¹ We did not place a country in the orange category

⁸⁸ Order of the President of the People's Republic of China No. 62. Decision of the Standing Committee of the National People's Congress on Amending the Copyright Law of the People's Republic of China (Adopted at the 23rd Meeting of the Standing Committee of the Thirteenth National People's Congress on November 11, 2020) <http://www.npc.gov.cn/englishnpc/c23934/202109/ae0f0804894b4f71949016957eec45a3.shtml>.

⁸⁹ *Id.* (“Article 24 In the following cases, a work may be used without permission of, and without payment of remuneration to the copyright owner, provided that the name or appellation of the author and the title of the work are indicated, the normal use of the work is not affected and the legitimate rights and interests enjoyed by the copyright owner are not unreasonably prejudiced:... (13) other circumstances as provided by laws and administrative regulations.”)

⁹⁰ *See e.g.*, Copyright Law, 2014 (Lei n.º 15/14 de 31 de julho de 2014, dos Direitos de Autor e Conexos) (Angl.): (“Art. 51... 2. The private reproduction referred to in subparagraphs d) of this article does not apply in the following cases: a) The reproduction of architectural works covering the form of buildings or other similar constructions; b) A reproduction reprographic of an entire book or of a musical work in graphical form (scores); c) A reproduction of all or important parts of databases in digital form; d) The reproduction of computer programs, in terms of this Law; e) No other reproduction of a work that affects the normal operation of the work or cause damage unjustified to the author's legitimate interests.”). Copyright and Neighboring Rights Protection Proclamation (Proclamation No. 410/2004) (Eth.): “Article 9... 2/ The provisions of Sub-Article (I) of this Article shall not extend to reproduction; I a) of a work of architecture in the form of a building or other construction; b) of musical work in the form of notation; or of the original or a copy made and signed by the author of a work of fine art. c) of the whole or a substantial part of a database in digital form; ' . d) of a computer program except as provided in Article 14 of this Proclamation; or e) which would conflict with or unreasonable harm the normal exploitation of the work or the legitimate interest of the author.”. Law on Literary and Artistic Property, 1995 (Law No. 94-036 of September 18, 1995), J.O. No. 2333 of 06.11.95, p. 3554 (Madag.) (“Article 42 – ... it shall be permitted, without authorization from the author, to reproduce a work lawfully published exclusively for the private use of the user. Paragraph 1 shall not apply: 1) to the reproduction of works of architecture in the form of buildings or other similar reconstructions; 2) to the reprographic reproduction of limited-edition works of fine art, to the graphic presentation of musical works (scores) or to exercise manuals that are used only once; 3) to the reproduction of the whole or large parts of databases; 4) to the reproduction of computer programs, save in the cases referred to in Article 51.”)

⁹¹ *See* Bangui Agreement Instituting an African Intellectual Property Organization (Act of December 14, 2015) (OAPI TRT/OA002/004) (“Article 10...(1)Notwithstanding the provisions of Article 8, and subject to those of

because of a restriction on research uses to “lawfully published”,⁹² “lawfully disclosed”,⁹³ “lawfully made available”,⁹⁴ or “lawfully acquired” works.⁹⁵ The countries in this category are listed in Table 6.

paragraph 2 of this article, it shall be permitted to reproduce a lawfully published work exclusively for the private use of the user without the consent of the author and without payment of remuneration. (2) Paragraph (1) shall not apply to the following: ... (iii) the reproduction of the whole or of significant parts of databases”).

⁹² See, e.g., Bangui Agreement (“Article 11. Free Reproduction for Private Purposes (1) Notwithstanding the provisions of Article 9, and subject to those of paragraph (2) of this Article and those of Article 58, it shall be permitted, without the consent of the author and without payment of remuneration, to reproduce a lawfully published work exclusively for the private use of the user.”. Copyright Act, 2017 (consolidated text of February 1, 2017) (Est.) (“Section 19. Free use of works for scientific, educational, informational and judicial purposes. The following is permitted without the authorisation of the author and without payment of remuneration if mention is made of the name of the author of the work, if it appears thereon, the name of the work and the source publication: 2) the use of a lawfully published work for the purpose of illustration for teaching and scientific research to the extent justified by the purpose and on the condition that such use is not carried out for commercial purposes; 3) processing of an object of rights for the purposes of text and data mining and provided that such use does not have a commercial objective;”).

⁹³ See, e.g., Law on the Protection of Copyright, Neighboring rights and Expressions of Folklore, 2003 (Law No. 005/PR/2003 of May 2nd, 2003) (Chad) (“Article 34.... Where a work has been lawfully disclosed, the author may not prohibit 2. copies or reproductions reserved strictly for the private use of the copier and not intended for collective use, with the exception of copies of works of art intended to be used for purposes identical to those for which the original work was created;”). Law on the Promotion and Protection of Intellectual Property Rights, 1993 (Decreto no. 604, modificado por el Decreto Legislativo N° 611, de 15 de febrero de 2017) (El Sal.) (“Article 45.... Regarding works already lawfully disclosed, it is allowed without authorization of the author or remuneration: a) The reproduction of a copy of the work for the personal and exclusive use of the user, made by the interested party with his own means, provided that it does not attempt against the normal exploitation of the work, nor does it cause unjustified damage to the interests legitimate of the author;”).

⁹⁴ See, e.g., Copyright Act, 2016 (Act No. 26 of 2016) (Malawi) (“Art 38. The reproduction, translation, adaptation, arrangement or other transformation of a work exclusively for the user's own personal or private use of a work which has already been lawfully made available to the public shall be permitted: “). Copyright and Related Rights Act, 2000 (Act No. 28 of 2000) (Ir.) (“Article. 50...(4) In this Part, “fair dealing” means the making use of a literary, dramatic, musical or artistic work, film, sound recording, broadcast, cable programme, non-electronic original database or typographical arrangement of a published edition which has already been lawfully made available to the public, for a purpose and to an extent which will not unreasonably prejudice the interests of the owner of the copyright.”)

⁹⁵ See Law on Copyright and Related Rights, 1999 (approved by Decree No. 4-99-E, as amended by Decree No. 16-2006) (Hond.) (“Article 47.... Regarding copies of lawfully acquired works by a person, it is allowed without authorization of the author or remuneration, reproduction of a copy of the work for the personal and exclusive use of that person, made by him, with his own means, provided that happens in special cases, that does not attempt the normal exploitation of the work or cause unjustified damage to the author's legitimate interests.”)

Table 6. Countries/Regions Coded Orange (86)

Algeria	Andorra	Angola	Armenia	Azerbaijan
Bangui Agreement	Belarus	Belgium	Benin	Bhutan
Botswana	Burkina Faso	Cambodia	Cameroon	Central African Republic
Colombia	Comoros	Costa Rica	Côte d'Ivoire	Denmark
Dominica	Dominican Republic	Equatorial Guinea	Eritrea	Ethiopia
EU Directive Art. 4**	Fiji	Finland	France	Gambia
Georgia	Ghana	Greenland	Guyana	Haiti
Ireland	Kazakhstan	Kyrgyzstan	Lebanon	Lithuania
Madagascar	Maldives	Mali	Malta	Mauritania
Montenegro	Morocco	Mozambique	Namibia	New Caledonia
Nicaragua	Niger	Norway	Oman	Palau
Papua New Guinea	Paraguay	Peru	Poland	Portugal
Republic of Moldova	Russian Federation	Rwanda	Sao Tome and Principe	Saudi Arabia
Senegal	Serbia	Seychelles	Sierra Leone	Slovenia
South Sudan	Sudan	Sweden	Switzerland	Syria (Arab Republic)
Tajikistan	Togo	Tonga	Trinidad and Tobago	Turkmenistan
Ukraine	United Arab Emirates	United Republic of Tanzania	Uzbekistan	Venezuela (Bolivarian Republic of)
Vietnam				

Exclusions of types of works from a research exception can be especially detrimental to TDM. Many of the works that are mined in TDM are contained in databases, for example. TDM of broadcasts and audio visual works can be used for media monitoring and language translation.⁹⁶ Bans on uses of whole books can prevent many projects in the digital humanities.

In this Part we consider a country's TDM exception within the general classification of each country. Thus, for example, Germany is coded green based its TDM exception, although it lacks an open general research exception. In the next section we compare TDM exceptions separately, showing that they parallel many of the restrictions that one finds in general research exceptions.

III. A TYPOLOGY OF TEXT AND DATA MINING EXCEPTIONS

In acknowledgement of the potential benefits of TDM to research and society, several countries have recently adopted specific exceptions for its use. In this part we analyze these exceptions under the same classifications we use above – i.e. on the degree to which each is open to reproductions

⁹⁶ See Amelia Brust, Artificial intelligence helps Voice of America translate broadcasts worldwide, Federal News Network (Feb 18, 2020); Ram Sagar, Netflix Is Using AI For Its Subtitles, Analytics India Magazine (May 6, 2020).

and communications, of all works, by all users. We also include analysis of whether the exceptions are limited to non-commercial uses or to works that are lawfully accessed, which appear in some TDM exceptions. Although we do not change the colors of our coding based on these factors, Figure 3, below, considers these elements in ranking TDM exceptions from most to least open.

A. Open TDM Exceptions

The most open TDM exception that we analyzed is in the law of Japan. Its TDM exception is crafted to introduce a new general standard for what some scholars call a “non-consumptive” or “non-expressive” use.⁹⁷ Japan’s Article 30-4 provides a general exception for “Exploitations not for enjoying the ideas or emotions expressed in a work,” which includes the specific non-exclusive example of “exploitation for using the work in a data analysis.”⁹⁸ There is no restriction to uses for non-commercial purposes or to lawfully published or lawfully accessed works.

Singapore passed the most recent TDM exception, which also is classified as green in our model, but is slightly more restrictive than Japan. Its exception for “Copying or communicating for computational data analysis” includes both reproduction and communication rights and applies to any use, including a commercial use.⁹⁹ It does, however, restrict uses to “lawfully accessed”

⁹⁷ Matthew Sag, Copyright and Copy-Reliant Technology, 103 Northwestern L. Rev. 1607–1682 (2009).

⁹⁸ Copyright Act, 1970 (Act No. 48 of May 6, 1970, as amended up to Act No. 72 of July 13, 2018) (Japan): (“Article 30-4 (ii). Exploitations not for enjoying the ideas or emotions expressed in a work. It is permissible to exploit work, in any way and to the extent considered necessary, in any of the following cases or other cases where such exploitation is not for enjoying or causing another person to enjoy the ideas or emotions expressed in such work; provided, however that this does not apply if the exploitation would unreasonably prejudice the interests of the copyright owner in light of the natures and purposes of such work, as well as the circumstances of such exploitation: ... (ii) exploitation for using the work in a data analysis (meaning the extraction, comparison, classification, or other statistical analysis of language, sound, or image data, or other elements of which a large number of works or a large volume of data is composed;”)

⁹⁹ Copyright Act 2021 (Revised Edition 2020, Act No. 22 of 2021) (Singapore). provides:

244.—(1) If the conditions in subsection (2) are met, it is a permitted use for a person (X) to make a copy of any of the following material:

(a) a work;
(b) a recording of a protected performance.

(2) The conditions are —

(a) the copy is made for the purpose of —

(i) computational data analysis; or
(ii) preparing the work or recording for computational data analysis;
(b) X does not use the copy for any other purpose;

(c) X does not supply (whether by communication or otherwise) the copy to any person other than for the purpose of —

(i) verifying the results of the computational data analysis carried out by X; or (ii) collaborative research or study relating to the purpose of the computational data analysis carried out by X;

(d) X has lawful access to the material (called in this section the first copy) from which the copy is made; and (e) one of the following conditions is met:

(i) the first copy is not an infringing copy;
(ii) the first copy is an infringing copy but —
(A) X does not know this; and

(B) if the first copy is obtained from a flagrantly infringing online location (whether or not the location is subject to an access disabling order under section 325) — X does not know and could not reasonably have known that;

(iii) the first copy is an infringing copy but —

(A) the use of infringing copies is necessary for a prescribed purpose; and (B) X does not use the copy to carry out computational data analysis for any other purpose.

(3) To avoid doubt, a reference in subsection (1) to making a copy includes a reference to storing or retaining the copy.

(4) It is a permitted use for X to communicate a work or a recording of a protected performance to the public if —

(a) the communication is made using a copy made in circumstances to which subsection (1) applies; and (b) X does

works. This makes it slightly less open than Japan.¹⁰⁰

We rate as green Section 19(3) of Estonia’s law which authorizes “processing of an object of rights” for the purposes of non-commercial text and data mining.¹⁰¹ For this law, we interpreted “processing” to be a broader category of activities than mere reproduction, potentially including the communications of works. Estonia’s is the only TDM exception to protecting “processing.” In Figure 3, we rank Estonia as less open than Singapore or Japan based on its non-commercial use restriction.

As we discuss below, we rate the EU’s 2019 directive as setting the minimum EU standard at the level of purple because it allows TDM reproductions of any work only by research and cultural institutions. But Germany took advantage of the flexibility in InfoSoc Directive¹⁰² to implement a TDM exception that extends to reproductions and communications of works, and therefore appears green in our maps.¹⁰³ Like Estonia, another EU country bound by the InfoSoc Directive, Germany restricts its exception to non-commercial uses. The examples of Estonia and Germany indicate that even after implementation of the DSM’s minimum standard for TDM uses, the EU is likely not to become a completely harmonized jurisdiction. Every country will be at least purple in our typology. But some countries may choose to use their flexibility to adopt green standards for all non-commercial research.

B. Reproduction Only

Some TDM exceptions only cover reproduction rights, and therefore do not explicitly exempt the many communications of works that copyright protects and that researchers need at various stages of a TDM project. Switzerland, for example, provides an exception “to reproduce” a work in research using “a technical process,” but does not cover communications of works:

1. For the purposes of scientific research, it is permissible to reproduce a work if the copying is due to the use of a technical process and if the works to be copied can be lawfully accessed.

not supply (whether by communication or otherwise) the copy to any person other than for the purpose of —

(i) verifying the results of the computational data analysis carried out by X; or

(ii) collaborative research or study relating to the purpose of the computational data analysis carried out by X.

(5) For the purposes of this Act, the supply of copies of any material in circumstances to which this section applies

(a) is not to be treated as publishing the material (or any work or recording included in the material); and (b) must be ignored in determining the duration of any copyright in the material (or the included work)

¹⁰⁰ See Michael W. Carroll, Copyright and the Progress of Science: Why Text and Data Mining is Lawful, 53 U.C. Davis L. Rev. (2019) (arguing that the U.S. fair use right would permit TDM on unlawfully accessed works, giving the Scihub database of illegally published works as an example).

¹⁰¹ Copyright Act, 2017 (consolidated text of February 1, 2017) (Est.)

¹⁰² European Parliament and Council Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society, 2001 O.J. (L167) (“3. Member States may provide for exceptions or limitations to the rights provided for in Articles 2 and 3 in the following cases: (a) use for the sole purpose of illustration for teaching or scientific research, as long as the source, including the author’s name, is indicated, unless this turns out to be impossible and to the extent justified by the non-commercial purpose to be achieved;”)

¹⁰³ Act on Copyright and Related Rights, 1965 (Copyright Act, as amended up to Act of September 1, 2017) (Ger.):

Section 60d. Text and data mining

(1) In order to enable the automatic analysis of large numbers of works (source material) for scientific research, it shall be permissible:

1. to reproduce the source material, including automatically and systematically, in order to create, particularly by means of normalisation, structuring and categorisation, a corpus which can be analysed and

2. to make the corpus available to the public for a specifically limited circle of persons for their joint scientific research, as well as to individual third persons for the purpose of monitoring the quality of scientific research.

2. On conclusion of the scientific research, the copies made in accordance with this article may be retained for archiving and backup purposes.¹⁰⁴

The UK's TDM exception is likewise limited to uses for reproduction.¹⁰⁵ We thus classify its TDM exception as blue in figure 3. The UK is classified as green in Part I based on its more open general research exception for "fair dealing."¹⁰⁶

C. Restrictions to Institutional Users

Some TDM exceptions are available for use only by designated institutions, which categorize in purple. This is the minimum standard required in EU member states. Article 3 of the EU DSM Directive requires that EU members permit TDM reproductions of any work by "research organisations and cultural heritage institutions."¹⁰⁷ TDM reproductions by such organizations are not subject to the provision in Article 4(3) – which applies to all users -- "that the use of works ... has not been expressly reserved by their rightholders." Accordingly, if all the countries of the EU harmonize their laws only to the minimum standard of the directive, then all of the EU would be purple in our map.

D. Restrictions on Works

France's law at the time of our study, which had not been adapted to the 2019 DSM Directive, restricts application of its TDM exception to "scientific publications."¹⁰⁸ As we describe above,

¹⁰⁴ Federal Act of October 9, 1992, on Copyright and Related Rights (status as of April 1, 2020) (Switz.) Art 24.d.

¹⁰⁵ Article 29A specifically covers "Copies for text and data analysis for non-commercial research," which applies to "the making of a copy of a work" for "computational analysis". The provision explicitly prohibits a "transfer" of the work or use "for any other purpose", which leads us to classify the exception as blue:

(2) Where a copy of a work has been made under this section, copyright in the work is infringed if— (a) the copy is transferred to any other person, except where the transfer is authorised by the copyright owner, or

(b) the copy is used for any purpose other than that mentioned in subsection (1)(a), except where the use is authorised by the copyright owner.

(3) If a copy made under this section is subsequently dealt with— it is to be treated as an infringing copy for the purposes of that dealing, and if that dealing infringes copyright, it is to be treated as an infringing copy for all subsequent purposes.

(4) In subsection (3) "dealt with" means sold or let for hire, or offered or exposed for sale or hire.

(5) To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable.

¹⁰⁶ See Copyright, Designs and Patents Act 1988, c. 48 (incorporating amendments up to the Digital Economy Act 2017) (UK.) Article 29 ("Fair dealing with a work for the purposes of research for a non-commercial purpose.").

¹⁰⁷ European Parliament and Council Directive 2019/790 on copyright and related rights in the Digital Single Market, 2019, art. 3, O.J. (L130). Article 3 provides:

1. Member States shall provide for an exception to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, and Article 15(1) of this Directive for reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access.

2. Copies of works or other subject matter made in compliance with paragraph 1 shall be stored with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results.

3. Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.

4. Member States shall encourage rightholders, research organisations and cultural heritage institutions to define commonly agreed best practices concerning the application of the obligation and of the measures referred to in paragraphs 2 and 3 respectively."

¹⁰⁸ Intellectual Property Code, 1992 (amended by Act No. 2016-925 of July 7, 2016) (Fr.). France's TDM exception states:

Article L122-5....

When the work has been disclosed, the author cannot prohibit: 10°. Copies and digital reproductions made from a

many private research exceptions exclude application to various works. France’s TDM exception may be the most restrictive among these depending on narrowly the category of “scientific” works is drawn. Many digital humanities and other TDM projects analyze popular literature, social media, movies, phonograms, broadcasts, and other works that are not easily described as “scientific.” It is noteworthy that although France’s restriction in the types of works that can be used is quite narrow, it does extend the uses permitted beyond mere reproduction. The statute authorizes a decree to set conditions for “the methods of conservation and communication of the files produced at the end of the research activities for which they were produced.”

Finally, Article 4 of the EU DSM Directive contains a novel restriction on the kinds of works that can be used by any user. Whereas Article 3 permits use of any work by specified organizations, Article 4 applies to any user but is subject to the “condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.”¹⁰⁹ Thus, we code Article 4 as orange.

E. *Safe harbor-only*

We reviewed one TDM provision – in Ecuador - that we classify as red because it does not appear to permit TDM at all. Ecuador’s law can be difficult to decipher because of some odd drafting. Elkin-Korin and Netaniel include Ecuador in their list of fair use countries based on the newly enacted Article 211 for “fair use” which is interpreted by some local experts to as a general exception applicable to a use for any purpose.¹¹⁰ Our methodology is different because we do not

lawful source for the purposes of mining text and data included in or associated with scientific publications, for public research purposes, excluding all commercial purposes. A decree fixes the conditions under which the exploration of texts and data is implemented, as well as the methods of conservation and communication of the files produced at the end of the research activities for which they were produced; these files constitute research data;

¹⁰⁹ European Parliament and Council Directive 2019/790 on copyright and related rights in the Digital Single Market, 2019, art. 4, O.J. (L130). Article 4 of the DSM provides:

Article 4. Exception or limitation for text and data mining

1. Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining.

2. Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining.

3. The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.

4. This Article shall not affect the application of Article 3 of this Directive.

¹¹⁰ Niva Elkin-Koren and Neil Weinstock Netanel, *Transplanting Fair Use Across the Globe: A Case Study Testing the Credibility of U.S. Opposition*, 72 HASTINGS L.J. 1121 (2021). Intellectual Property Law (Consolidated version of February 10, 2014) (Ecuador). Ecuador’s law provides:

Art. 211. Fair Use.

It will not constitute a violation of the patrimonial rights the use or exploitation of a protected work or benefit, in the cases established in the following article, as long as they do not attempt against the normal exploitation of the work or protected benefit and do not cause unjustified damage to legitimate interests of the holder or holders of the rights. To determine if the use of the work or service is in accordance with the provisions of this article, the provisions of this Code will be taken into account and the International Treaties to which Ecuador is a party. In addition, at least the following should be considered factors:

The objectives and nature of the use;

The nature of the work;

The quantity and importance of the part used in relationship with the work as a whole, if applicable;

The effect of use on current market value and potential of the work; and,

The enjoyment and effective exercise of other rights fundamental

engage local interpretive sources and rather base our classifications on the words of the statute alone. The wording of Article 211 applies a four factor fair use test “to the cases in the following article” – referring to the list of specific exceptions in Article 212. We thus interpret this as an additional restriction on the exceptions in Article 212 rather than an opening clause.

Article 212 also contains significant interpretive issues. The Article is titled “Acts that do not require authorization for their use.” Its opening clause states that “the cases determined in this article shall not constitute a violation of the patrimonial rights of the rights holder” and that “the following acts do not require the authorization of the rights holder nor are they subject to any remuneration.”¹¹¹ It thus appears to define a set of exceptions. Article 212(9) defines rights of libraries, including as acts that do not require authorization:

Individual reproduction of a work by a library, archive or museum, when the copy respective is in the library collection, archive or museum, and such reproduction is made with the following purposes.

After defining three such purposes, the provision states “A library or archive may also carry out the following acts:

viii. Text mining. Libraries and archives and their officials shall be exempt from responsibility for the acts carried out by their users as long as they act in good faith and have reasonable grounds to believe that the work protected by copyright or provision protected by neighboring rights has been used within the framework allowed by the limitations and exceptions provided in this Paragraph or in a way that is not restricted by rights over the work or performance, or that said work or performance is in the domain public or under a license that allows its use;

This provision is perhaps the most difficult to interpret of all the provisions we reviewed. Are we to interpret that “A library or archive may also carry out ... Text mining”? Or is the provision only the liability safe harbor that follows that title: protecting libraries from “responsibility for the acts carried out by their users”? The safe harbor is provided for acts believed to be “within the framework allowed by the limitations and exceptions provided in this Paragraph.” But the paragraph does not define any “framework” permitting the actual exercise of TDM. Based on the language of the provision, with no consultation to interpretive guidance, we judge that it does not provide an exception for the actual reproduction and communication of works in TDM, and therefore is classified red.

Figure 3 presents a ranking of the openness of each TDM exception according to the six-color scale we use in part II and also considering the two additional criteria that appear frequently in recent TDM exceptions – lawful access or publication requirements for the works used, restrictions on commercial uses. We also indicate (but do not consider in our ranking) whether each exception protects the user right from contracts that attempt to waive or override copyright exceptions. As in the typology above, we do not change the color of the classification based on these factors. The TDM exceptions are listed from the most open to the least open. Thus, one can see that of the green

¹¹¹ Intellectual Property Law (Consolidated version of February 10, 2014) (Ecuador). The full introductory paragraphs states:

Without prejudice to the provisions of the previous article, in accordance with the nature of the work, the international instruments to which Ecuador is a part and the principles of this Code, the cases determined in this article shall not constitute a violation of the patrimonial rights of the rights holder, provided that they do not attempt against the normal exploitation of the works and do not cause unjustified damage to the legitimate interests of the owner or holders of the rights. In this sense, the following acts do not require the authorization of the rights holder nor are they subject to any remuneration:

TDM exceptions, Japan’s is the most open and Estonia’s is the most restrictive, even though they are all classified green.

We do not classify countries here based on whether the laws in question specifically authorize storage of materials for TDM, as several recent exceptions do.¹¹² We assume that granting reproduction rights needed for TDM includes the right to store those materials given that “storage” is not generally a separate protected right in copyright laws.

Figure 3.
TDM Exceptions in Comparative Copyright

Country	Commercial	Lawful Access	Protection from contracts	Uses	Users	Works	Typology
Japan	Y	N		Use	Open	Open	Green
Singapore	Y	Y	Y	Reproduction, communication	Open	Open	Green
Germany	N			Reproduction, communication, storage	Open	Open	Green
Estonia	N	N		“processing”	Open	Open	Green
UK	N	Y	Y	Reproduction	Open	Open	Blue
Switzerland				Reproduction	Open	Open	Blue
EU DSM Art 3	*	N	Y	Reproduction, storage	Cultural institutions	Open	Purple
France	N	Y		Reproduction, communication (decree)	Open	Scientific writings	Orange
Ecuador	*			TDM “acts carried out by their users”	Libraries, Archives safe harbor	Open	Red

¹¹² See Federal Act of October 9, 1992, on Copyright and Related Rights (status as of April 1, 2020) (Switz.) (“Art 24.d- 1. For the purposes of scientific research, it is permissible to reproduce a work if the copying is due to the use of a technical process and if the works to be copied can be lawfully accessed. 2. On conclusion of the scientific research, the copies made in accordance with this article may be retained for archiving and backup purposes...”); Intellectual Property Code, 1992 (amended by Act No. 2016-925 of July 7, 2016) (Fr.) Article L122-5 (indicating that decree will define conditions for “methods of conservation and communication of the files produced at the end of the research activities for which they were produced”).

CONCLUSION

The intent of this report is to fill a gap in the existing literature on copyright exceptions by classifying countries on the degree of openness in their research exceptions generally, and in any exception specifically for text and data mining research. We show that although every copyright law in the world has at least one exception that promotes research purposes, there is a wide degree of variation between countries. We find that one can classify research exceptions into six broad categories along a spectrum from the most open to research uses implicating any exclusive right of all works by any user to those that are the most restrictive – permitting only the quotation or use of excerpts. This same pattern is replicated in TDM exceptions, with only a small number (4) being fully open to TDM uses.

This Article may be useful in helping countries find models for domestic copyright reform as well for consideration of guidelines or norms for harmonization between countries. With regard to the latter, our report points to a serious problem in cross border uses of works for TDM. Only a small portion of our study permits communications of works for research, and none explicitly considers cross border uses.