E

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

SCCR/44/4
ORIGINAL: ENGLISH
DATE: OCTOBER 17, 2023

Standing Committee on Copyright and Related Rights

Forty-Fourth Session
Geneva, November 6 to 8, 2023

SCOPING STUDY ON THE PRACTICES AND CHALLENGES OF RESEARCH INSTITUTIONS AND RESEARCH PURPOSES IN RELATION TO COPYRIGHT

prepared by Professor Raquel Xalabarder

ACKNOWLEDGMENTS

I would like to express my gratitude to Professor Tatsuhiro Ueno (Waseda University) and to Professor Ignasi Labastida (University of Barcelona) for their collaboration and valuable contributions to this Study, notably Annex I on Text and Data Mining and Annex III on Open Science, respectively.

I am deeply grateful to Ms. Christina Langer and Ms. Geidy Lung (CLD) for their continuous support during this project and crucial help in its last stages.

I also thank Ms. Sylvie Forbin (CCIS) and Ms. Michele Woods (CLD) for their advice and orientation.

Finally, I would like to thank Mr. Joseph Fometeu, Mr. Thomas Margoni, Ms. Graciela Melo, and Ms. Marisella Ouma for their helpful reviews and insightful comments.

ABOUT THE AUTHOR

Professor Raquel Xalabarder holds a Chair of Intellectual Property at the Universitat Oberta de Catalunya, where she also served as the Law School Dean, until August 2023. She earned her Master of Laws degree from Columbia University Law School in New York (1993) and achieved her Doctor of Laws, cum laude by unanimity, from the Universitat de Barcelona (1997).

Professor Xalabarder has been a Visiting Scholar at Columbia University Law School in New York (2000-2001) and an Honorarvertrag at the Max Planck Institute for Intellectual Property in Munich (2008 and 2011). She holds notable positions, including President of ALADDA (Spanish group of ALAI), Treasurer of ATRIP, and membership in the European Copyright Society.

Professor Xalabarder is renowned for her extensive teaching and prolific publications on intellectual property, law of the internet, and private international law. Currently, she serves as the director of the European Patent Office (EPO)'s Academy, showcasing her exceptional dedication and expertise in the field.

TABLE OF CONTENTS

ABBREVIATIONS

AAM	Author Accepted Manuscript
AI	Artificial Intelligence
APC	Article processing charge
API	Application Programming Interface
BC	Berne Convention for the Protection of Literary and Artistic Works
BTAP	Beijing Treaty on Audiovisual Performances
CC	Creative Commons
CCC	Copyright Clearance Centre
CDSM	EU Directive on Copyright in the Digital Single Market (2019)
CJEU	Court of Justice of the European Union
CMO	Copyright Management Organization
DOI	Digital Object Identifier
DRM	Digital Rights Management
ECL	Extended Collective License
EU	European Union
FLOSS	Free/Libre and Open Source Software
GPL	General Public License
IPO	Intellectual Property Office
ISD	EC Information Society Directive (2001)
L&E	Limitations and Exceptions
MPA	Motion Picture Association
MPLC	Motion Picture Licensing Corporation
MVT	Marrakesh Treaty to Facilitate Access to Published Works for Persons Who Are Blind, Visually Impaired or Otherwise Print Disabled (2013)
NC	Non Commercial
ND	Non Derivatives
NLP	Natural Language Processing
OA	Open Access
OAPI	African Intellectual Property Organization
PLOS	Public Library of Science
RC	Rome Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organizations (1961)
RRO	Reproduction Rights Organization
SA	Share Alike
SCCR	WIPO Standing Committee on Copyright and Related Rights
SSRN	Social Science Research Network
STM	Science, Technology, Engineering and Mathematics
TDM	Text and Data Mining
TPMs	Technological Protection Measures
TRIPs	Agreement on Trade Related Intellectual Property Aspects
UK	United Kingdom
UNESCO	United Nations Educational, Scientific and Cultural Organization
USA	United States of America
VoR	Version of Record
WCT	WIPO Copyright Treaty
WIPO	World Intellectual Property Organization
WPPT	WIPO Performances and Phonograms Treaty

INTRODUCTION

Researchers[1] are the driving force behind the expansion of human knowledge, creativity and innovation and are at the center of this dynamic relationship. Through research activities, they engage in a variety of tasks, such as data acquisition, experimentation, analysis and dissemination of results. Research activities give rise to new products and services, new jobs, as well as new industries and markets. They bring scientific and technological development, improve our lives and social welfare.

In this rapidly evolving landscape of modern knowledge creation and dissemination, there is one dimension that has emerged as a crucial area of study: the intricate interplay between researchers (and research) and copyright. Researchers often find themselves in a position where they must assume two distinct roles, and this alternation between roles can be quite fluid. On the one hand, they act as authors or creators of original content, meaning that they produce and generate their own intellectual work, such as research articles, reports, books, databases, and other types of material. On the other hand, these same researchers also act as users or consumers of copyrighted materials created by others in their search for information and resources relevant to their research.

This fluid alternation between the roles of creators and users in research generates a unique set of copyright considerations that need to be explored separately from other topics, such as education or libraries, previously studied by the WIPO Standing Committee on Copyright and Related Rights (SCCR). Therefore, the purpose of this exploratory study is to shed light on those specific facets and complexities that are yet to be examined within the context of research activities, thereby avoiding information duplication.

The rich typology of research activities and the changes generated by technological development serve as a key starting point (as described in section 1). Copyright limitations and exceptions (L&E) appear as critical foundations in this setting, balancing the protection of creative expression with the encouragement of scientific activities. These provisions allow researchers to utilize protected content for specific objectives like criticism, study, research, and education. Text and data mining (TDM), which allows academics to investigate huge volumes of data in pursuit of innovation-driving insights across disciplines, is one of these new digital age's provisions enacted in some national laws (as noted in section 2).

As researchers navigate the boundaries of copyright limitations and exceptions, licensing systems play a vital role in granting permissions that transcend legal boundaries. These licensing arrangements, which can range from conventional to alternative approaches, provide researchers with access to a wide diversity of content, including through Open Science strategies (as discussed in section 3).

Among the several challenges that merit special consideration regarding the application of limitations and exceptions, is the potential for overridability: that is, whether licenses could supersede copyright limitations and exceptions statutorily enacted. A further issue is technological measures of protection (TPMs), which are employed in digital rights management systems to regulate access to and use of copyrighted material. Similarly, the interaction of TPMs and copyright exceptions and limitations may pose the challenge of TPMs preventing uses that have been permitted under statutory limitations and exceptions. Finally, in an increasingly interconnected world, research activity easily transcends borders. The advancement of knowledge is dependent on international collaborations and the utilization of resources from various jurisdictions. Jurisdictional and applicable law challenges for the enforcement of rights in cross-border research efforts result from the variety of copyright laws and regulations, as well as licensing solutions available in different Member States (as discussed in section 4).

[1] See https://data.oecd.org/rd/researchers.htm.

In the following scoping study, we will delve briefly into each of these interconnected issues. Our goal is to equip SCCR Members with the necessary tools and information to navigate the dynamic realm of research cultivating an environment that nurtures both creativity and innovation while also facilitating access to knowledge.

1. TYPOLOGY

The typology of research activities consists of the same elements that were previously offered in relation to education and libraries for the SCCR, namely the various categories of activities, along with the primary and secondary rights involved, the scope of works, the purpose, and conditions of use, as well as the elements for ongoing consideration.

Within the framework of library typologies, for instance, a comprehensive analysis of the copies produced by these institutions for the benefit of researchers was conducted.[2] This analysis listed elements of ongoing consideration such as the implementation of digital technologies, the implications of licensing or acquisition agreements, the scope of extended licenses, the delivery of copies to external users, the relationship with interlibrary loans and cross-border transfers, and the liability associated with subsequent user actions. In addition, issues such as multiple requests of copies for the same work, the administration of supplementary materials, and the retention of copies in the library's archives were addressed.

It is not intended to merely restate the material that has already been presented in those typologies; but rather to provide more information for illustrative purposes that provide an understanding of the range of the issues related to research, which are relevant in the field of copyright.

In a natural order of things, research activities require all kinds of content that can be used as inputs by researchers.[3] Depending on the academic discipline, research may be based on both contents that are not protected by copyright (objects, artifacts, methods, data, as well as works in the public domain), and on contents protected by copyright and related rights. Research activities are usually conducted based on previous research outputs, that could range from academic publications and data to press publications, novels, audiovisual recordings, phonograms and sound recordings, works of art and images, among others.

From a copyright perspective, lawful access to protected contents does not amount to an authorization to conduct further acts of exploitation of the content (reproduction, distribution, communication to the public, adaptation, among others), beyond the reading, viewing, or listening by the human senses.

Traditionally, in analogic contexts, access to copyrighted contents does not require any license or authorization from rightholders under copyright: having “physical” access to a copy of the work (usually in a tangible format) already grants “lawful” access to the work. For instance, it is possible for researchers to go to a library and to access printed copies of academic papers and books that are relevant to their research off the shelves. Subsequent actions such as reading or examining the work by a human brain do not qualify as an act of exploitation under copyright. In contrast, in the digital environment, the access landscape changes significantly.

Consumption and exploitation of copyrighted materials are muddled online. The “lawful” access has expanded from mere physical possession to digital rights affecting how researchers and institutions use digital content. Digital copies may be protected from copying, distribution, or alteration, but also TDMs may prevent routine and non-exploitative actions such as reading and viewing.

---

[2] See https://www.wipo.int/edocs/mdocs/copyright/en/sccr_38/sccr_38_4.pdf, page 8.

[3] The kind of contents depends on the academic field of research: arts and humanities (history, religion, literature, visual arts, music and audiovisual works, entertainment, communication studies, marketing, linguistics), social sciences (archeology and geography, cultural studies, psychology and sociology, economics, law and politics, business, journalism, education), natural sciences (physics and chemistry, biology, earth and space sciences), medical sciences (medicine, pharmacy) and technical sciences (mathematics, statistics, computer sciences, logic).

In the context of copyright, researchers perform a number of acts, such as copying, transforming, translating, distributing, and communicating to the public. These uses affect various aspects of the exclusive rights established by copyright law.

Historically, research activities were carried out in academic institutions and universities, mainly involving the creation of hard copies of specific library collections. These actions focused primarily on published works and, to some extent, on the reproduction of artistic works, such as films or songs, within the context of research. To a large extent, these uses were limited to the right of reproduction, especially in analog formats such as paper copies or photocopies. As a result, many national legislations only mentioned reproduction and communication to the public in analog contexts and in relation to research institutions.

In contrast, when research is conducted online and through digital means, it is not limited only to reproduction, but also involves communication to the public and making it available to the public.

Digital technologies have greatly improved the conduct of research. Materials used for research purposes can now be digitized from analog formats, such as books or journals, or recorded from television broadcasts. These materials, as well as those born digital, can be locally stored and compiled on shared storage drives or in the cloud for later access and reuse.

Researchers share copies with each other, either directly from one researcher to another or through shared access drives and internal networks. These actions may extend through the Internet and social networks. However, the categorization of these activities in the realm of distribution and communication to the public may change as the size of the team and the audience reached increases.

The role of non-human (machine) reading, such as artificial intelligence (AI) analysis, is growing in importance within research methodologies. Text and data mining (TDM) has gained prominence thanks to digital technologies. Researchers extract information from a variety of protected works, from academic works to music and press publications. It should be noted that, in many circumstances, the specific and unambiguous categorization of acts involved in TDM lacks a well-defined frame of reference in national laws.

Numerous other significant factors affect the copyright implications in research. Researchers operate in a variety of settings, including laboratories, specialized research institutions, universities, and private businesses. Their employment statuses span a broad spectrum, including full-time, part-time, and student positions. Certain researchers may choose to work independently as independent contractors or consultants. Regardless of the institution in which they conduct their research, researchers can engage in both commercial and non-commercial activities. For the purpose of this study, non-commercial research refers to research conducted without direct or indirect profit motives. Conversely, commercial research is carried out with the explicit aim of creating profit or financial benefit.

Given this brief context, it is obvious that research activities cannot be limited only to the scope of copyright L&E, nor only to licensing considerations. As concluded in the previous “Report on practices and challenges in relation to distance learning and online research activities” (Document SCCR/39/6):[4] both licenses (direct and collective) as well as L&E are necessary to secure that copyrighted contents in digital formats will be available for research purposes. The most plausible assumption is the convergence of these two aspects, encompassing numerous subtleties along the spectrum.

[4] Torres, M./ Xalabarder, R. (2019) “Report on practices and challenges in relation to distance learning and online research activities” (Document SCCR/39/6), https://www.wipo.int/edocs/mdocs/copyright/en/sccr_39/sccr_39_6.pdf.

The industry most directly affected by research activities is, undoubtedly, the academic publishing sector: academic journals and academic publications. Understandably, this is the sector where licensing for research purposes, including Open Access (OA), has advanced further. However, research activities require access and use of contents from other sectors, such as non-academic literary works (novels, press contents), music, audiovisual as well as artistic works, phonograms and audiovisual recordings, software or databases. For these sectors, where licensing options for research purposes are not commonly available, L&E become even more important.

The extent of L&E is inextricably linked to the availability of licensing that exists in each country: the extent of activities allowed under statutory L&E will define the scope of acts of exploitation that require licensing, and vice versa, the availability of licensing in one country will shape the kind and conditions of L&E needed in that country. Yet, the precise balance between them remains uncertain. The following sections will delve deeper into the categories that research efforts can fall under.

2. LIMITATIONS AND EXCEPTIONS

Aware of the importance of research for the development of society and the public interest, national copyright laws provide for L&E that permit specific research uses, without the need to obtain the rightholders' consent. The scope of permitted uses varies widely among different national laws. In order to secure the public interest behind research activities, L&E define a core of uses directly permitted by law (be it for free or subject to compensation). The conditions for these L&E may be certainly different depending on the digital or analogic formats.

The specific balance of L&E and the specific conditions for its application are frequently defined at a national level, considering the specific conditions of each country, such as the structure of research institutions, the academic publishing markets, language and cultural issues, the development of collective management organizations (CMOs), among others.

Research uses are often permitted by several L&E, both in international treaties as well as in national copyright laws:[5]

- Quotations;
- Private use/copying;
- Fair use/dealing provisions (in common law countries);
- Research purposes (often jointly with teaching purposes);
- Library uses;
- Text and Data Mining exceptions.

In some countries, compensation is required for some of them (for instance, for private copying and for some library uses); compensation schemes are usually managed by CMOs.

Having a variety of L&E available does not mean that all research uses and activities are permitted by law. As always, the scope of L&E will very much depend on the specific language used to define them and their interpretation. Digital technologies have expanded the potential of research activities, but not all L&E in national copyright laws are equally responsive to them. Unclear and outdated L&E provisions generate legal uncertainty and ultimately challenge the development of research activities in digital and online means, including across-borders. Interpreting their scope and/or the need to revise them and update them remains the responsibility of national legislators and, ultimately, courts.

[5] For the sake of simplicity, the terms 'exception' and 'limitation' will be used in this Study indistinctively to refer to statutory provisions which authorize specific exploitation acts (or uses), whether the authorized act/use is free or remunerated.

2.1 INTERNATIONAL TREATIES

Research purposes have historically been regarded as a possible basis for L&E in copyright and related rights treaties. Research purposes have been present in the Berne Convention since its very adoption in 1886.[6]

a) Berne Convention (Art.10(1) and (2) BC)

Scientific research purposes have been acknowledged in the Berne Convention for the Protection of Literary and Artistic Works (hereinafter BC) since its very origins. Art. 8 of the Berne text of 1886 stated:

> *"As regards the liberty of extracting portions from literary or artistic works for use in publications destined for educational or **scientific purposes**, or for chrestomathies, the effect of the legislation of the countries of the Union, and of special arrangements existing or to be concluded between them, is not affected by the present Convention."*

This provision subsequently evolved into the current Art.10 BC which has two provisos: quotations and illustration for teaching.

(i) Quotations (Art.10(1) BC).

Quotation exceptions are paramount for research activities and, in fact, are inherently linked to them. According to Art.10(1) BC:

> *"It shall be permissible to make quotations from a work which has already been lawfully available to the public, provided that their making is compatible with fair practice, and their extent does not exceed that justified by the purpose, including quotations from newspaper articles and periodicals in the form of press summaries."*

The current Art.10(1) BC[7] exempts **any acts of exploitation**: reproduction, distribution, communication to the public and making available, as well as translations;[8] it applies to all kinds of works (provided they have been "*lawfully made available to the public*"), without any specific limitation as to the amount that may be quoted. Of course, the term 'quotation' itself already suggests some restriction, but its length will be ultimately determined *in casu,* subject to the conditions of the "*extent justified by the purpose*" and in a manner that is "*compatible with fair practice.*"[9]

As Prof. Ricketson explains,[10] quotations for "*scientific, critical, informatory or educational purposes*" are clearly included within its scope. Similarly, since the quotation exception is

---

[6] See Berne Convention for the Protection of Literary and Artistic Works, of 9 September 1886, as revised at Paris on 24 July 1971 and amended in 1979 [hereinafter, BC].
[7] The Stockholm revision of 1967 provided its current language, substantially enlarging its scope. See Ricketson/Ginsburg, op. cit. supra, §13.39
[8] See WIPO (1976), WIPO Reports op. cit. supra, § 205. See Ricketson, S., WIPO Study, op. cit. supra, p.37-39: "the exclusion of translations from the exceptions provided in these Articles will lead to a manifestly absurd or unreasonable result". Aligned with this conclusion, Sec.7 "Fair use" of the WIPO Tunis Model Law on Copyright of 1976 expressly allows (under all the listed exceptions) the use of works "either in the original language or in translation".
[9] See Ricketson, S., WIPO Study, op. cit. supra, p.12.
[10] See Ricketson, S., WIPO Study, op. cit. supra, p.13.

neither restricted in terms of beneficiaries nor technology, it may exempt quotations for research purposes by any means of exploitation (i.e., digital formats and online contexts).

Art.10(1) BC is drafted as a mandatory provision that Member States must apply in their national laws. Being an imperative restriction to the authors' rights, Member States cannot reduce its extent in favor of Union authors and works. As Ricketson explains:

> *"It will be contrary to the Convention for national legislation to provide protection in a case where this has been specifically prohibited. In the same way that there is a principle of minimum of protection that operates under article 19 in favor of Union authors, so (it can be argued) there is a corresponding principle of maximum protection to be implied in those few cases where the Convention limits or excludes protection"*[11]

Yet, as important as they are for research activities, quotation exceptions will always be restricted to using a part or a fragment of the work (to the extent necessary). They will hardly permit the making of whole copies of works, except in very specific cases.[12]

(ii) Illustration for teaching (Art.10.2 BC).

Its origins can be traced back to Art.8 BC (1886). The Brussels Revision of 1948 modified Art.10 BC, as follows:

> *"The right to include excerpts from literary or artistic works in educational or scientific publications, … in so far as the inclusion is justified by its purpose, shall be a matter for legislation in the countries of the Union, and for special Arrangements existing or to be concluded between them."*

It was at the Stockholm conference of 1967, that the formal reference to "scientific publications" was formally lost in the current text of Art.10(2) BC:

> *"It shall be a matter for legislation in the countries of the Union, and for special agreements existing or to be concluded between them, to permit the utilization, to the extent justified by the purpose, of literary or artistic works by way of illustration in publications, broadcasts or sound or visual recordings for teaching, provided such utilization is compatible with fair practice."*

As stated by Masouyé in his Guide to the Berne Convention, it is clear from Art.10(2) language that "mere scientific research is not within the scope of this paragraph".[13] Little explanation may be found as to why "scientific purposes" were gradually deleted from the conventional text: from use in publications destined for "educational or scientific purposes" (Art.8 Berne Act, 1886), to "educational or scientific publications" (Art.10(2) Brussels Act, 1948), and its final disappearance in Stockholm.[14] In any case, be it under Art.10(1) or Art.9(2) BC (see below on the three-step test), quoting from and making copies of works for research purposes are permitted under the Berne Convention and remain – as always – a matter for national legislation.

---

[11] See Ricketson, S./ Ginsburg, J., op. cit. supra, §13.42. Ficsor prefers another explanation: the compulsory nature of Art.10(1) BC is not so much an exception to the principle of minimum of protection granted under the BC but rather "it follows from a basic human freedom - the freedom of speech and criticism - that is justified and necessary to allow free quotations in appropriate cases." See Ficsor, M. (2002), The Law of Copyright and The Internet, The 1996 WIPO Treaties, their Interpretation and Implementation, Oxford University Press, §5.12.

[12] For instance, when commenting on a work of art, a photograph or a poem.

[13] See Masouyé, C. (1978), Guide to the Berne Convention for the Protection of Literary and Artistic Works (Paris Act, 1971), WIPO, Geneva. #10.10.

[14] See Ricketson, S./ Ginsburg, J., op. cit. supra, §13.34 and §13.45.

b) Berne Convention (Appendix for developing countries)

In 1971, the Appendix to the Berne Convention set up a special framework for supporting research in developing countries by setting up a licensing system regarding translation. Even though this special system has been around for more than 50 years, its use has not been proven. This is a point that has been brought up by many experts in past reports of the SCCR. Furthermore, the introduction and growth of digital technology have had little impact on this terrain.

As stated in Art. II of the Appendix, compulsory licenses may be granted for translation and publication in print of such translations for purposes of teaching, scholarship or research of works, published in printed or analogous form (paragraphs (1) and (5)), provided they have not been translated and published by the owner of rights in a language in general use in the country of protection within three years of the first publication, or, if such publication has taken place, it is out of print (paragraph (2)). The term is reduced to one year if the translation is into a language which is not in general use in one or more developed Union countries, and for other languages the effected countries may agree on shorter terms among them, however not shorter than one year and not for translation into English, French and Spanish (paragraph (3)).

To those terms are added: for the three-year term six months; and for the one-year term nine months, after the completion of various formalities concerning notification of the rightholders, and within those time limits the rights owner can avoid any compulsory license by publishing or reissuing the work (paragraph (4)). A compulsory license may eventually be terminated by the rightholders by publishing or reissuing the work in translation, but copies already made under the license may continue to be distributed until the stock is exhausted. A compulsory license cannot be granted if the author has withdrawn the work (paragraph (8)). As regards works which are composed mainly of illustrations, the compulsory licenses for translation cannot be used on their own, but only if the conditions for a compulsory license for the right of reproduction under Art. III of the Appendix are fulfilled as well (paragraph (7)).

Additionally, under the aforementioned conditions, the Appendix permits broadcasting organizations in developing countries to obtain compulsory licenses in order to translate works that have been published in printed or analogous forms of reproduction, as long as the translation was made from a lawfully made and acquired copy and that it was only used in broadcasts intended solely for teaching or for the dissemination of the results of specialized technical or scientific research to experts in a particular profession; that the translation is used exclusively for those purposes through broadcasts made lawfully and intended for recipients on the territory of the country of protection, including broadcasts made through the medium of sound or visual recordings lawfully and exclusively made for the purpose of such broadcasts; and that all uses made of the translation are without any commercial purpose. Recordings made specifically for such use may under the same conditions also be used by other broadcasting organizations in the country of protection. Similarly, under the same conditions the provision also permits compulsory license for translations with the purposes of subtitling or dubbing of educational programs (paragraph 9).

c) The three-step test

The formal introduction of a general right of reproduction “in any manner or form”, for all kinds of works, in Art.9(1) BC at the Stockholm Conference (1967) posed the subsequent question regarding what L&E could be envisaged in national laws. The possibility of adopting an exhaustive list of exceptions to the “new” reproduction right was considered and rejected, because it would be too difficult to agree on them and they could never cover all the special cases existing in national legislation.[15] Instead, the three-step test was adopted: a list of three

[15] See Ricketson, S./ Ginsburg, J., op. cit. supra, §13.07.

cumulative conditions that national legislators must meet when implementing L&E to the reproduction right in their national laws.

According to Art.9(2) BC:

> *"It shall be a matter for legislation in the countries of the Union 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."*

Subsequently, Art.13 Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPs) (1995), Art.10 WIPO Copyright Treaty (WCT) and Art.16 WIPO Performances and Phonograms Treaty (WPPT) (1996), as well as Art.13 Beijing Treaty on Audiovisual Performances (BTAP), extended the application of the test (with slightly different language) to all exclusive rights of authors and related rights owners, beyond reproduction, and with a clear intent to apply it also in the digital environment. As explained in the respective Agreed Statements, national laws must comply with the test when adapting and introducing L&E to exclusive rights, including in digital means: "these provisions should be understood to permit Contracting Parties to devise new exceptions and limitations that are appropriate in the digital network environment."[16]

Over the years, the three-step test has been the guide for national legislators to secure a variety of L&E in national laws, such as for private copying / use, research purposes and library uses, and to keep them updated, as necessary, to exempt uses under new technologies, including digital and online means. At the same time, the three-step test confirms that, as a general rule, L&E remain the responsibility of national laws.

According to the three-step test, L&E for research purposes must be carefully defined to avoid exempting uses in a manner that conflicts with the normal exploitation of a work;[17] but at the same time – foreseeing that all exempted uses do so, to some extent – the requirement of remuneration may clear the last step (unreasonable prejudice to the legitimate interests of the author) and help strike a "reasonable" balance between the normative, public interest justifying the L&E (in our case, research purposes) and the possible prejudice caused to rightholders.

d) Other references to research purposes in international Instruments

Beyond the Berne Convention, scientific and research purposes have also been formally acknowledged as L&E by other international instruments.

As far as related rights, Art.15(2) of the Rome Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organizations (RC) permits the application of the national L&E envisaged for copyright (authors' rights) and, specifically, Art.15(1)(d) of the Convention formally refers to the *"use solely for the purposes of teaching or scientific research."*

In 1976, WIPO and UNESCO adopted the Tunis Model Law on Copyright for Developing Countries. Sec.7 in this agreement includes an extensive list of L&E, such as quotations, the making of copies for private and personal use, as well as *"the reproduction, by photographic or*

[16] See, for all, Agreed Statement concerning Article 10 WCT (1996): "It is understood that the provisions of Art.10 permit Contracting Parties to 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. Similarly, these provisions should be understood to permit Contracting Parties to devise new exceptions and limitations that are appropriate in the digital network environment."

[17] For instance, by excluding from the L&E works meant for the teaching market or by setting flexible but clear quantitative or qualitative restrictions so as to avoid depleting the normal markets for the exploitation of the work.

similar process, by public libraries, non-commercial documentation centers, scientific institutions and educational establishments."

Many national laws have mirrored L&E in their national laws to these international provisions.

In turn, research purposes and access to information are also formally mentioned in the 1996 WCT preamble when recognizing "*the need to maintain a balance between the rights of authors and the larger public interest, particularly education, research and access to information, as reflected in the Berne Convention*." This same imperative is underscored in subsequent agreements such as the 1996 WPPT, the 2012 BTAP, and the 2013 Marrakesh Treaty (MVT).

And last, but not least, Art.7 TRIPs reminds States that "*the protection and enforcement of intellectual property rights should contribute to the promotion of technological innovation and to the transfer and dissemination of technology, to the mutual advantage of producers and users of technological knowledge and in a manner conducive to social and economic welfare, and to a balance of rights and obligations."*

At a regional level, the Bangui Agreement, of December 14, 2015,[18] instituting the African Intellectual Property Organization (OAPI), includes in its Annex VII devoted to Copyright and Related Rights an Art.13 setting an exception for reprographic (facsimile)[19] copies by libraries and archives, as well as for preservation purposes and for research or private study within that establishment.

> *"Article 13*
>
> *Reprographic reproduction by libraries and archive services*
> *Notwithstanding the provisions of Article 9, a library or archive service whose activities are not directly or indirectly profit-making may, without the consent of the author or other holder of copyright, make individual copies of a work by means of reprographic reproduction:*
> *(i) where the making of such copies is for the purpose of preserving and, if necessary, in the event of it having been lost, destroyed or made unusable, replacing them, or for replacing a copy that has been lost, destroyed or rendered unusable in the permanent collection of another library or other archive service; and*
> *(ii) where the reproduction of such a work is for the purpose of preservation or intended to preserve the conditions for its consultation for research or private study by individuals on the premises of establishments and on dedicated terminals of libraries that are open to the public, by museums or by archival services, provided that they seek no economic or commercial benefits."*

To conclude, research purposes have always been envisioned by international treaties to justify L&E in national laws. It is up for national legislators to implement these L&E for research purposes in their national laws and to keep them updated to new technological means of exploitation, including digital and online ones, always in compliance with the three-step test.

2.2 NATIONAL LAWS

As noted in Section 1, research uses involve a varied typology of activities that affect all exclusive rights: reproduction (i.e., photocopying, downloading and printing) and making

[18] See https://wipolex.wipo.int/en/text/582620.
[19] According to the definition in Art.1 of the Annex to the Bangui Agreement (xvii), "La « reproduction reprographique » d'une œuvre est la fabrication d'exemplaires en facsimilé d'originaux ou d'exemplaires de l'œuvre par d'autres moyens que la peinture, notamment par tout procédé impliquant une technique photographique ou assimilée, y compris la photocopie, l'impression, la numérisation, le stockage dans des bases de données ou système d'information".

available (i.e., uploading the contents on an intranet), communication to the public or transmission, including the making of multiple transient copies), further subsequent copying and distribution (a second copy made or downloaded by a "recipient" researcher), and even adaptation and the making of derivate works (i.e. when translating a work for research purposes, and perhaps also when building a dataset for machine readable / TDM purposes). Naturally, the moral rights of the authors, namely attribution and integrity, are also implicated and have significant importance for research endeavors, albeit their impact is not extensively influenced by the extent of L&E and/or licensing frameworks applicable to research objectives.

Academic and research needs are addressed by national laws in a fragmented manner, under several L&E. In addition to the general exceptions for quotations and private use/copying that may be found in most countries and exempt an important spectrum of research activities, some countries go further and exempt research purposes (often together with teaching purposes). Common law countries tend to rather rely on fair use/fair dealing provisions and, in many cases, in combination with a non-exhaustive list of L&E to authorize research uses. In parallel, L&E for libraries remain fundamental to exempt some research uses, namely: obtention of copies for research purposes, public lending, and access to works in the library collection through terminals located in the library premises. Only a few national laws provide for specific L&E specifically authorizing TDM uses for purposes of research.

This section does not intend to offer an exhaustive analysis of how national L&E address research activities, but rather to identify a few common and most relevant issues and challenges that research purposes face under national L&E, namely:

a) Technology restrictive language

There are still many L&E in national laws that specifically refer to analogic copies or to photocopying. This may apply to all sorts of L&E: quotations, private copying and personal use, fair dealing, research and library L&E. This is not a problem when the technological distinction (i.e., analogic private copies or digital private copies) is only made to subject them to different conditions (i.e., amount of copying, reference to TPM, compensation or for free). However, issues may arise when this distinction is used to delineate the boundaries of these L&E. To illustrate, consider a scenario where a provision allowing for quotations is exclusively designed for analog formats. In such a case, this limitation could fall short of effectively safeguarding the public interest that underpins the concept of fair use for quotations. Additionally, this approach could also miss out on harnessing the full potential of the diverse range of exempted uses outlined in Art.10(1) BC.

Virtually all national laws include a quotation exception (the only mandatory exception of the BC) that may apply to research uses. Research purposes are often mentioned in quotation exceptions[20], but even when they are not, quotations made for research purposes may be easily exempted. Quotations only allow for the use of a fragment of a work, not the whole work. So, they would hardly allow making copies of the whole quoted work or sending these copies to other researchers (perhaps also located in a different country) or posting them on an intranet.

b) Digital copies and digitization

Scanning a physical printed item fundamentally requires making a digital reproduction. As a result, if a national L&E framework grants such an exemption, the digitization process could

[20] See for instance Chile Copyright Act 1970, Art.71 B: *"It is lawful to include in a work, without remuneration or obtaining authorization from the owner, short fragments of a protected work, which have been lawfully disclosed, and their inclusion is made by way of quotation or for the purposes of criticism, illustration, teaching and research. as long as its source, title and author are mentioned".*

potentially come within the reproduction exemption. However, if the language used in the L&E is specifically limited to analog forms or to the act of photocopying, this exemption may not allow digitization. In other words, if the L&E clearly refers to physical copying or photocopies, the exemption will hardly apply to cover digital uses. Of course, the scope of these provisions may be extended by judicial interpretation, always in compliance with the three-step test; but even then, the legal uncertainty derived from it remains an important challenge for the development of research activities.

It is critical to highlight that any interpretation of the scope of such an exemption must adhere to the three-step test criteria. This test guarantees that any limitation or exception to copyright protection does not interfere with regular utilization of the work, does not unfairly impair the rightholder's legitimate interests, and strikes a fair balance between the interests of rightholders and users.

In essence, while digitization may be permitted under the broader category of reproductions, this interpretation is dependent on the language of the L&E and must follow the three-step test requirements to guarantee a balanced and equitable application of copyright laws.

It is undeniable that the use of digital copies exempted under L&E (for any purposes, also for research and scientific purposes) brings a higher risk of downstream infringing uses than the exemption of analogue copies. In addition, digitization for research purposes may have a higher negative impact on the primary markets of academic works rather than on other markets, such as the music, audiovisual or artistic markets. The economic impact of digitization in these different markets deserves to be carefully considered to find nuanced solutions for academic and research uses exempted under L&E.

The debate regarding whether digital copies and digitization should be permitted under “old” L&E drafted before digital technologies arrived has been going on for decades with different outcomes across the world. For instance, in the EU, digital copies for private use, as well as for research purposes, are formally allowed “on any medium”;[21] the former subject to compensation.[22] This question might take a different trajectory under fair use/ dealing provisions. In many other countries, the exemption of digital private copies and digitization for research purposes is still uncertain, especially so, when specific language refers to analogic or tangible copies.

Ultimately, when norms are silent, it will be a matter for interpretation by courts which will need to look at the purpose of the L&E and decide whether digital copies and scanning may be permitted under the specific L&E, in a manner that is respectful with the three-step test.

The decision regarding the updating or interpretation of L&E in national laws, to facilitate researchers' use of digital and online formats while aligning with the three-step test, rests within the discretion of individual Member States.

c) Adaptations and translations

Very often, L&E applicable to permit (directly or indirectly) research fail to mention whether adaptations and translations of the work are also permitted.

For instance, quotation exceptions usually use neutral and all-encompassing language, such as “use”, “quote”, or “make a quotation”, but a few of them refer to specific acts of exploitation,

[21] See Art.5.2(b) and Art.5.3(a) ISD (2001).
[22] Hence, teaching and research uses may be indirectly compensated through levy systems provided for private copying applicable on equipment (such as photocopiers, printers, and scanners) and/or on operators (schools, colleges, universities, libraries, research institutions, etc.).

such as reproduction, distribution, or communication to the public. Very few of them formally refer to adapting or translating the quoted work.[23] And yet, as concluded when examining the scope of the quotation exception under Art.10(1) BC, excluding translations under a quotation exception (of any kind, and regardless of its formal language) would "lead to a manifestly absurd or unreasonable result".[24] Interpreting a quotation exception overly restrictively would not only be unnecessary for compliance with the three-step test (since modifying or translating a fragment of a work for quotation purposes would not harm the normal exploitation of the work or the author's legitimate interests), but it would also undermine the intended effectiveness and purpose of the exception.

Fair dealing / use provisions may also be conducive to include adaptations or translations, but unless it is formally stated so[25] (or there are specific guidelines established) researchers may be uncertain regarding the scope of the permitted use.

Similarly, L&E for research purposes rarely formally permit the adaptation of the work, despite the making of derivative works (be it a translation, an adaptation, or a summary) of a work used remaining a fundamental element for the development of research activities of any kind.[26]

For instance, within the EU, the right of adaptation has not been generally harmonized in EU *acquis* and L&E for research and teaching purposes in Art.5.3(a) ISD (2001) only formally cover reproduction, distribution and communication to the public. Accordingly, most L&E for teaching and research purposes in EU national laws fail to mention the right of adaptation. Based on the "subsidiarity principle" behind the EU harmonization of copyright, EU national laws may choose to include translations within the permitted scope[27] but this does not happen often.

Likewise, the recently issued Copyright Act of the Federal Republic of Nigeria provides an exception[28] for non-commercial research and private study, which does not specifically mention the right of adaptation.

Accordingly, whether or not a researcher can make a summary or a translation of a work for research purposes under the national L&E will remain a grey area, generating uncertainty and challenges subject to interpretation.

---

[23] See Japan (2010) Art.32(1) and Art.43(ii).
[24] See WIPO (1976), WIPO Reports op. cit. supra, § 205. See Ricketson, S., WIPO Study, op. cit. supra, p.37-39. Aligned with this conclusion, Sec.7 "Fair use" of the WIPO Tunis Model Law on Copyright of 1976 expressly allows (under all the listed exceptions) the use of works "either in the original language or in translation".
[25] Although this may not always be apparent either and, to avoid any doubts, the Australian Copyright Law (2015) Sec.40(1) formally includes "an adaptation" of a work within the fair dealing for purposes of research: "*Sec.40 Fair dealing for purpose of research or study. (1) A fair dealing with a literary, dramatic, musical or artistic work, or with an adaptation of a literary, dramatic or musical work, for the purpose of research or study does not constitute an infringement of the copyright in the work.*"
[26] Exceptionally, China (2010) Art.22.6, allows "translations and reproductions in a small quantity of copies … for use by teachers or scientific researchers. And Cuba (1977) Art.38.d, permits the making of copies by reprographic means (as well as translations) in scientific institutions.
[27] See Poland (2016) Art.27: "Research and educational institutions shall be allowed, for teaching purposes or in order to conduct their own research, to use disseminated works *in original and in translation*, and to make copies of fragments of the disseminated work.
[28] Federal Republic of Nigeria. Copyright Act, 2022. Art. 20. (1) (c): *"The rights conferred in respect of a work under sections 9-13 of this Act, do not include the right to control any of the acts specified in those sections by way of fair dealing for purposes such as — (c) non-commercial research and private study".*

d) The scope of personal / private use

Research uses may also be exempted under L&E for personal use / private copying that exist in most national laws.[29] Copies done by researchers for research purposes may be exempted under the general reference of "personal use" or "private use". The scope of these terms may vary depending on the jurisdiction; as a rule of thumb, private use may result in a more flexible reading than personal use, allowing uses done in family or team-related contexts.

Usually, L&E for private use/copying applies to all kinds of works (computer programs and digital databases may often be excluded). Fair compensation is mostly required within EU countries; elsewhere, compensation for private copying is not the norm.[30,31]

The most common challenge regarding these L&E is defining the scope of "personal or private use". In order to be exempted under the L&E, copies must be restricted to the personal or private use by the person who makes the copy (usually, the researcher). Additionally, personal/private copies must be done for non-commercial (non-lucrative) purposes and cannot be subject to a collective use.

Personal use typically encompasses activities within the immediate family and minor individual uses, devoid of independent economic importance. However, extending this concept to additional uses beyond the family circle might be viewed as overly permissive. For instance, expanding personal use to encompass activities within a research team, so that a private copy is shared among fellow researchers.[32] The situation becomes more complex when considering the scenario where the recipient of the shared copy resides in a different country from where the original private copy was created.

In common law countries, research uses may benefit from fair use / fair dealing provisions.[33] Most of them expressly mention research purposes, together with private use or study. Fair use / dealing provisions are flexible and technology-neutral; hence, they are easily adaptable to new technologies and markets. Fair use / dealing provisions usually rely on the assessment of different factors such as the purpose or character of the use; the amount and substantiality of the part used; the nature of the work; and the effect of the use on the market of that work.[34] These factors will determine, in the circumstances of each specific case, the existence of an infringement or a fair use / dealing.

Fair use / dealing provisions face a very specific challenge: that of unpredictability. In theory, since a fair use / dealing can only be assessed "ex post" in view of the specific circumstances of

---

[29] According to the Study prepared by Prof. Seng, of 192 WIPO Member States studied, virtually all of them (179) provide some L&E permitting private copying or copying for personal use. In addition to the European countries, this includes 45 WIPO countries in the Asia and Pacific region; 50 countries in Africa, and 29 countries in Latin America & the Caribbean. See Seng (2017) *WIPO Study on Copyright Limitations and Exceptions for Educational Activities* (Document SCCR/35/5 REV), available at https://www.wipo.int/edocs/mdocs/copyright/en/sccr_35/sccr_35_5_rev.pdf

[30] See *International Survey on Private Copying Law & Practice 2015,* available at https://www.wipo.int/publications/en/details.jsp?id=4010.

[31] See CISAC (2022) Private Copying Global Study, available at https://www.auvibel.be/wp-content/uploads/2020/11/20201123-private_copying_study_2020.pdf.

[32] Art. 44 of the Colombian Law states that the use of scientific, literary, and artistic works in a private home, non-profit, is also free, available at https://www.wipo.int/wipolex/en/text/506452.

[33] Some 33 out of 192 WIPO Member States provide for some fair use / fair dealing provision for research; this includes most notably the USA, UK, Ireland, Canada and Israel, as well as 13 countries in Asia and the Pacific, 6 countries in Africa, and 9 in the Caribbean. See Seng (2017) *WIPO Study on Copyright Limitations and Exceptions for Educational Activities* (Document SCCR/35/5 REV), available at https://www.wipo.int/edocs/mdocs/copyright/en/sccr_35/sccr_35_5_rev.pdf.

[34] For instance, see Jamaica (1993) Sec. 54. "For the purpose of determining whether an act done in relation to a work constitutes fair dealing, the court determining the question shall take account of all factors which appear to it to be relevant, including— (a) the nature of the work in question; (b) the extent and substantiality of that part of the work affected by the act in relation to the whole of the work; (c) the purpose and character of the use; and (d) the effect of the act upon the potential market for, or the commercial value of, the work."

the case, it is generally agreed that fair use / dealing scenarios are unpredictable and lead to legal uncertainty. However, as caselaw and guidelines develop, the purported unpredictability of fair use / dealing provisions seems to be more apparent than real; at least, not more unpredictable than the interpretation of statutory L&E as applied to specific cases.[35]

e) Sharing copies obtained under other L&E

Following a strict interpretation, when a L&E only refers to "reproduction," researchers might not be allowed to "share" a lawfully obtained copy with other individual researchers in their teams, or libraries would not be allowed to deliver the requested copy to the researcher, because that would involve an act of distribution or of communication to the public (if transmitted online). However, should these acts qualify as acts of distribution to the public and / or communication to the public at all?

Although the concept of public is not defined in international treaties, it is usually accepted that it should not apply to persons within a family circle or a private circle; in other words, the non-public should be "economically insignificant". In Europe, the CJEU has had the opportunity to interpret the concept of public (within the right of communication to the public) and defined it as "a public consisting of an indeterminate and fairly large number of people."[36] The question here would be one of interpretation: should an individual researcher or a team of researchers qualify as a "public"?

A similar problem is posed when dealing with exempted library copies for research purposes. The relevant L&E only formally permits the making of a copy (reproduction), but usually this copy can be "delivered" to the researcher (be it in tangible or intangible format), interpreting that this is not an act of distribution or communication to "the public".

This same question is posed regarding the scope of copies done for research purposes that libraries are allowed to do, under L&E set in their favor as explained below.

f) Acts permitted for research purposes

As explained above, L&E set specifically for research purposes (or scientific research) may also be found in some national laws. Like in Art.8 BC (1886) and Art.15 RC, research purposes are often addressed jointly with teaching purposes in national laws.[37] This is the case in most EU countries where L&E permit acts of reproduction, distribution, and communication to the public for research purposes, also in digital and online formats.[38]

---

[35] See, for instance, USA Conference on Fair Use (CONFU) Guidelines, available at https://libguides.uiwtx.edu/c.php?g=954561&p=6887707.

[36] See CJEU, Judgment of 13 February 2014, *Svensson and Others* (C-466/12, EU:C:2014:76, paragraph 22).

[37] The previous "Report on practices and challenges in relation to distance learning and online research activities" (Document SCCR/39/6) already examined these L&E, as well as the obstacles and challenges nowadays faced by research and teaching activities. For an extensive analysis of L&E for teaching purposes in national laws, see the "*Updated Study and Additional Analysis on Copyright Limitations and Exceptions for Educational Activities"* prepared by Professor Daniel Seng (2017) (Document SCCR/35/5 REV), available at https://www.wipo.int/edocs/mdocs/copyright/en/sccr_35/sccr_35_5_rev.pdf, as well as the several regional studies commissioned by WIPO in 2009: *WIPO Studies on the Limitations and Exceptions to Copyright and Related Rights for the Purposes of Educational and Research Activities*: Monroy Rodríguez, J.C. (Latin America and the Caribbean) SCCR/19/4; Fometeu, J. (Africa) SCCR/19/5; Nabhan, V. (Arab Countries) SCCR/19/6; Seng, D. (Asia and Australia) SCCR/19/7; Xalabarder, R. (North America, Europe, Caucasus, Central Asia and Israel) SCCR/19/8; available at https://www.wipo.int/meetings/en/details.jsp?meeting_id=17462.

[38] Art.5.3(a) ISD of 2001 permitted Member States to introduce in their national laws L&E to the rights of reproduction (distribution) and communication to the public (including the making available): "(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;" Most EU

[Footnote continued on next page]

The worldwide scenario is a rather different one. In some countries, L&E for research purposes remain restricted to analogic copies (e.g., photocopying) and face-to-face scenarios (known as "chalk and talk"),[39] making its application to digital and online uses rather difficult, if not impossible.

g) Non-commercial purposes

L&E that refer to research institutions may be more conducive to the needs of research teams than those L&E only benefitting individual researchers, because they may also permit the distribution and communication to the public (within the center) of the copies done for research purposes. The EU, instead, took a different approach: the L&E for research purposes in Art.5.3(a) ISD (2001) only refers to the "non-commercial purpose to be achieved" and recital 42 ISD stresses that this requirement should not rely on the organization structure and the means of the funding; accordingly, L&E for research purposes may also benefit private research institutions. Interpreting the term 'non-commercial' is also another challenge and appears to vary widely across the EU.[40] Apart from the EU and the UK, "non-commercial" is not a widely used term in copyright laws and its scope is uncertain. [41] Instead, other legislations, like the Andean Community[42], have adopted the not-for-profit criteria. When research uses are permitted under fair use / dealing provisions (seen above), the existence of a lucrative intent does not automatically disqualify a finding of fair dealing; this is so, unless a non-for-profit use is explicitly required.[43]

countries implemented Art.5.3(a) ISD (2001) combining "illustration for teaching" and "research purposes" in one single L&E; see Italy (Art.70), Hungary (Art.33.4), Luxembourg (Art.10.2), Poland (Art.27) and Spain (Art.32.3 and 4). Instead, only a few national provisions in the EU countries specifically address "scientific research" purposes under a separate L&E; see §60c German Law; also the UK (a former EU member) implemented Art.5.3(a) ISD in two different provisions: fair dealing for purposes of non-commercial research (Sec.29) and for educational uses (Sec.32 – 36A).

[39] This was the general finding of the comparative law Studies commissioned by WIPO in 2009 and the scenario has not changed much in over a decade. See (2009) WIPO *Studies on the Limitations and Exceptions to Copyright and Related Rights for the Purposes of Educational and Research Activities*: Monroy Rodríguez, J.C. (Latin America and the Caribbean) SCCR/19/4; Fometeu, J. (Africa) SCCR/19/5; Nabhan, V. (Arab Countries) SCCR/19/6; Seng, D. (Asia and Australia) SCCR/19/7; Xalabarder, R. (North America, Europe, Caucasus, Central Asia and Israel) SCCR/19/8, available at https://www.wipo.int/meetings/en/details.jsp?meeting_id=17462.

[40] See LIBER (2020) *Limitations and Exceptions in EU Copyright Law for Libraries, Educational and Research Establishments: A Basic Guide* p.9, available at https://libereurope.eu/wp-content/uploads/2020/11/A-Basic-Guide-to-Limitations-and-Exceptions-in-EU-Copyright-Law-for-Libraries-Educational-and-Research-FINAL-ONLINE.pdf: "How the term 'non-commercial' is interpreted appears to vary widely across the EU. Some countries allow copying by individuals in a commercial company under the private copying exception, whereas in other Member States this is viewed as commercial research. In addition to this, given the large Knowledge Transfer agenda between universities and industry, distinguishing between commercial copying and copying for the purposes of research is frequently impossible. … Exceptions for research copying should be both for commercial and non-commercial research. This would facilitate university and industry research collaborations: the so-called Knowledge Transfer agenda."

[41] See Dusollier, S. (2010) Scoping Study on Copyright and Related Rights and the Public Domain (Document CDIP/4/3/REV./STUDY/INF/1), available at https://www.wipo.int/edocs/mdocs/mdocs/en/cdip_4/cdip_4_3_rev_study_inf_1.pdf.

[42] Decision No. 351 of the Commission of the Andean Community on the Common Provisions on Copyright and Neighboring Rights. *"Article 22.- Without prejudice to the provisions of Chapter V and the previous article, it will be lawful to carry out, without the authorization of the author and without payment of any remuneration, the following acts: (…) b) Reproduce by reprographic means for teaching or for conducting examinations in educational institutions, to the extent justified by the purpose pursued, articles lawfully published in newspapers or periodical collections, or brief extracts from lawfully published works, provided that such use is made in accordance with honest uses and that it is not the subject of a sale or other transaction for consideration, nor does it have direct or indirect profit purposes; c) Reproduce individually, a work by a library or archive whose activities do not have direct or indirect profit purposes".*

[43] See, for instance, Malaysia (2006) Art.13 (2).

h) L&E for libraries and archives

Research is intrinsically linked to libraries. L&E applied to libraries have also an important role to play in permitting research uses.[44] The scope of L&E for libraries has been debated over the years, with legal certainty being a primary concern.

In addition to the general questions previously examined under other L&E, regarding the making of digital copies and of translations, L&E for libraries pose specific questions that may find different solutions (if any) under national laws. Here are two instances of such questions:

(i) Digital copies for research purposes and online delivery

This question was already addressed above regarding researchers, more generally. Here it will be addressed in the specific context of copies for research purposes done by libraries.

When the L&E only permits the making of a copy for research purposes, can this copy be delivered to the researcher who requested it? In paper format or also in digital format? In other words, are digital "document delivery services" included under the L&E set in favor of libraries or do these services require specific licensing from rightholders?

L&E must be interpreted not only according to the criterion of the three-step test but also according to additional principles. For instance, some articulate that its interpretation "must enable the effectiveness of the exception thereby established to be safeguarded and permit observance of the exception's purpose".[45] Accordingly, an L&E that allows the making of copies (also in digital formats) for research purposes, but does not allow delivering the copy to the researcher who requested it would not satisfy the ultimate goal sought by this L&E.

(ii) Making available through designated terminals in the library premises

Libraries often make works in their collections available to the public through designated terminals in their premises. For instance, in the EU, Art.5.3(n) ISD (2001) permits the "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 establishments referred to in paragraph 2(c) of works and other subject-matter not subject to purchase or licensing terms which are contained in their collections". Following that, all EU countries contain a similar provision in their national laws.

Other non-EU countries[46] like Côte d'Ivoire, Kyrgyzstan, Moldova, Montenegro, Panama and São Tomé and Principe, also provide L&E for making available the work on dedicated terminals located in the premises of libraries and research institutions.

Such L&E present two main challenges. First, whether the "dedicated terminals" L&E is only open to contents that has been originally licensed in digital format or it could also apply to digitized copies of works in the library collection? Second, regarding the restrictive interpretation of "library premises".

---

[44] For an in-depth analysis of L&E for libraries, see the Study prepared by Prof. Crews in 2015. See Crews, K. (2015), Study on Copyright Limitations and Exceptions for Libraries and Archives (Document SCCR/30/3), available at https://www.wipo.int/edocs/mdocs/copyright/en/sccr_30/sccr_30_3.pdf.

[45] See CJEU, 3.09.2014, *Deckmyn v. Vandersteen* (C-201/13), ECLI:EU:C:2014:2132, # 23; CJEU, 4.10.2011, *Football Association Premier League y otros* (C-403/08 y C-429/08), EU:C:2011:631, # 163.

[46] See Seng, D. (2016) Study on Copyright Limitations and Exceptions for Educational Activities (Document SCCR/33/6), available at https://www.wipo.int/edocs/mdocs/copyright/en/sccr_33/sccr_33_6.pdf, and Crews, K. (2017) Study on Copyright Limitations and Exceptions for Libraries and Archives: Updated and Revised (Document SCCR/35/6), available at https://www.wipo.int/edocs/mdocs/copyright/en/sccr_35/sccr_35_6.pdf.

The CJEU answered to the first question in the *Darmstadt* case: this L&E may not only apply to contents originally licensed in digital format, but also to digitized copies of works in its collection, as long as the library had lawfully obtained them. In other words, the library may benefit from the L&E for preservation to digitize the one tangible copy of the work available in its catalogue, and subsequently make this digital copy available to researchers under the another L&E, in this case that of "dedicated terminals".[47] Notice that the court is not endorsing a massive digitization of the whole library collection, but allowing digitization of works in its collection, in specific circumstances, so as to permit the enforcement of the L&E of dedicated terminals.

A similar provision is found in Art.13 Bangui Agreement allowing the reproduction by publicly accessible libraries of works in their collections, and to make them available for consultation for research purposes or private study by individuals, through dedicated terminals in their premises, as long as made for non-commercial purposes.

Beyond its scope, the second challenge relates precisely to the fact that this L&E is formally "restricted" to terminals in the library premises, thus discouraging online and cross-border research uses. In order to access specific licensed materials, the researcher must physically travel to the library. This discussion encompasses remote access to digital databases through "designed terminals" within the library premises,[48] but also to access unique copies of historical works in the public domain (i.e., the researcher can access these works in the physical library, but cannot access them remotely via the library network despite these works are already in the public domain!).

L&E that are restricted to physical contexts unnecessarily constrain the way research is done in our current world.[49]

(iii) TPMs and libraries

Through the use of TPMs, libraries (like rightholders) are in a position to control that uses made by researchers will remain within the contours permitted by the L&E. TPMs have proven to be necessary so that rightholders can control the exploitation of their works in digital markets, so should libraries be trusted in using TPMs also to enforce the scope of permitted uses under L&E. Similarly, rightholders should ensure that beneficiaries of statutory L&E (i.e., libraries, research centers and researchers, as applicable) will effectively benefit from them and that the specific permitted uses will not be restricted or prevented by TPMs.

---

[47] See CJEU, 11.11.2014, *Technische Universität Darmstadt v. Ulmer* (C-117/13) ECLI:EU:C:2014:2196 #45-47. In this case, the library scanned (digitized) a history book published by Ulmer, which was in the library catalogue in printed form, and subsequently made it available to the public through dedicated terminals in the library premises. Only one tangible copy of that book was available in the library. The publisher was offering a license to access all the treatises published by Ulmer in e-book format (amongst them, the scanned book); the library refused to obtain this global license because it only "needed" to use one book.

[48] See, for instance, Colombia Law 1915 (2018), Art.16(c) permitting libraries, archives and documentation centers the making available through "dedicated terminals" in their premises, for purposes of research or personal study any kind of works and protected subject matter which have been lawfully acquired and are not subject to licensing terms. https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=87419

[49] Communia offers the experience of "Jonas", a lecturer of comparative literature in Sweden:
"To access material from 1956, we have to go to the National Library Lab in Stockholm. It is a small glass cage with three data terminals. You sit in the lab, annotate... You are not allowed to take data in or out, all labs must be done in the cage." See COMMUNIA, (2023) *Nobody puts research in a cage. Researchers' perspective on working with copyright* p.11; available at https://communia-association.org/wp-content/uploads/2023/03/Researchers-on-Copyright.pdf.

(i) Text and data mining

All along history, technological development has transformed the way research is conducted. Computer programs and databases already transformed research activities a few decades ago. Now, it is the turn of TDM.

This Study defines 'text and data mining' as an automated process (using an algorithm) aimed at analyzing data to generate information and knowledge.[50] TDM is inherently linked to the development of AI projects and of natural language processing (NLP) projects, all of them based on algorithms and machine learning, information extraction, and ultimately TDM techniques.

TDM opens opportunities for the development of new products and services, both by private and public entities. It may be a powerful tool to obtain valuable information for the provision of private and public services, for making complex decisions and creating strategies, and for the development of new applications or products for users and citizens. And as always, opportunities come with risks. Whether or not governments and private businesses should be allowed to develop new services and products based on the mining of cultural contents, copyrighted contents, personal data, the personal attributions such as a person's image and voice, is being currently debated in many jurisdictions. Specifically, several Generative AI projects have recently caught all media attention and deserve separate consideration.

This Study is mainly focusing on TDM technologies for research purposes, which is one upstream part within the larger AI realm: whether TDM for research purposes should be permitted by law or whether it should be left to the decision of stakeholders. Or whether it is a question of determining the modalities and rules of the game, which must be accepted by all the stakeholders concerned. A question that may only be solved after a careful analysis of the benefits it may entail for the advancement of science and research, as well as the effects for the legitimate interests of authors and rightholders in that country.

What remains unquestioned is that TDM is a fundamental tool for the advancement of research activities and will redefine the way research is done, worldwide.[51] Researchers may rely on machines to "read" data (works) and find patterns, trends and correlations that may be valuable for the development of future research. Researchers who do not, may be at a competitive disadvantage.

All these systems rely on all sorts of data: ranging from numbers, facts and statistics, to works and other copyright protected subject matter, such as literary works, academic writings, works of visual art, musical works and recordings, audiovisual works and recordings. All of them will be turned into a *corpus* that will be "normalized" and subsequently "mined" for the building of TDM

[50] For other definitions, *see* Art.2(2) of the CDSM Directive (2019/790) which defines 'text and data mining' as 'any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations'. *See* also Elsevier TDM Glossary: "Text mining is the data analysis of natural language works (articles, books, etc.), using text as a form of data. It is often joined with data mining, the numeric analysis of data works (like filings and reports) and referred to as "text and data mining" or, simply, "TDM." TDM depends on the assembly of a working set of data/content against which an analytic process is run. This process breaks down digital information into raw data and text, analyses it, and comes up with new connections, from unexpected patterns. This can eventually lead to the development of a new drug, to subtle shifts in weather patterns that might predict a downturn in the price of wheat." https://www.elsevier.com/__data/assets/pdf_file/0018/102906/TDM-Glossary.pdf

[51] "Researchers have been working across disciplines more often than before the pandemic, reading preprints to stay connected to new ideas and increasingly using new technology such as AI to help analyze the data sources available to them"; see Elsevier (2022) "Research Futures 2.0: A new look at the drivers and scenarios that will define the decade", p.7, available at https://www.elsevier.com/__data/assets/pdf_file/0017/1250423/Research-Futures-2_0-Full-Report.pdf.

projects, including AI. As a result, data "mined" (machine read) can comprise both non-copyrighted matters[52], and creative works protected under copyright.

The use of raw data and mere information is a very valuable asset for research purposes. Governments are starting to require that data generated from publicly funded research projects should be available for free use on publicly accessible databases. Similarly, research institutions and academics foster publication of their research production and data in open academic journals, as well as in open access platforms. The challenge here is the difficulty to distinguish between copyright protected contents and non-protected data. As long as TDM is based on raw data, no L&E or licensing should be necessary. Yet, to the extent that these data sets are available via databases, copyright may be certainly affected at least within the EU and its database *sui generis* right.[53]

Beyond raw data, works and other objects protected by copyright and related rights may also be the subject of TDM. When it comes to these contents, TDM involves several acts of exploitation (that is, reproduction, often also an adaptation, and communication to the public). Because of the specific requirements involved in TDM activities (namely, reproduction, storage, metadata normalization, as well as sometimes communication to the public among different researchers and research institutions), existing L&E (also for research purposes) may not suffice to permit TDM.

Certainly, temporary acts of reproduction may be covered by L&E.[54] But TDM activities often entail making copies (acts of reproduction) that go beyond temporary or *de minimis,* and making other acts of exploitation of the works and recordings that are being mined. To be mined, corpuses must be "normalized" and afterwards, they may be annotated or even transformed for subsequent mining uses. Failing a L&E to cover these acts, TDM uses of copyrighted contents will require the rightholders' authorization.

A few countries, mostly in developed countries, have recently allowed TDM uses of copyrighted contents: either under fair use doctrines (*i.e.*, USA) or under specific statutory L&E for TDM (*i.e.,* the EU, Japan, Singapore, UK, but also and Ecuador[55] and Ukraine[56]). Some countries permit TDM for any purposes, including for-profit TDM projects; this is the case of Japan (Art.30.4) and the USA (as long as qualifying as a fair use). Others permit TDM for research purposes only; this is the case of Singapore and the EU. For a detailed analysis of these and other national provisions allowing TDM uses for research purposes, see Annex I of this Study.

It is important to note that these TDM provisions are very unique and concentrate in a small number of jurisdictions. Most national laws do not formally exempt TDM uses, neither in general nor for research purposes.

The role of licenses in advancing TDM is evident, with many academic publishers incorporating TDM rights into their subscription database licenses (this is further illustrated in the next section of the Study) and L&E also contributing to this landscape. Ultimately, it is up to the Member States to formulate L&E on TDM for certain objectives, including research or, more specifically,

---

[52] TRIPS Agreement and WCT provide protection to the arrangement or selection of the data, not the data or material itself.

[53] Where Database *sui generis* rights exist (such as in EU countries) the maker of a database may claim an exclusive right to authorize and/or prohibit the extraction and reuse of its contents (be it works or data).

[54] See, for instance, Art.5(1) ISD (2001): *Temporary acts of reproduction […], which are transient or incidental, and an integral and essential part of a technological process and whose sole purpose is to enable: (a) a transmission in a network between third parties by an intermediary, or (b) a lawful use of a work or other subject-matter to be made, and which have no independent economic significance, shall be exempted from the reproduction right […]* In similar terms, see Colombian Law 1915 (2018), Art.16(a) permitting transient or accessory temporary copies; available at https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=87419.

[55] Article 212 of the *Código de Ingenios.* (for libraries and archives), available at https://www.wipo.int/wipolex/en/text/439410.

[56] Article 22(2)(14) of the Ukrainian Copyright Law, available at https://www.wipo.int/wipolex/en/text/587392.

non-for-profit research. This entails establishing parameters (such as specific works or protected subject matter) and terms (such as compensation), while allowing opportunity for licensing schemes to develop naturally in the market.

3. LICENSING

Rightholders have the possibility to grant licenses for uses beyond those specifically permitted by law through L&E, either individually or through collective management organizations. Direct licensing takes place when the righthholder permits the use of content under agreed conditions and compensation, while collective licenses are granted by organizations designated by rightholders to manage their rights on their behalf, through CMOs.

Historically, rightholders handled licenses for primary uses, and CMOs administered licenses for secondary uses of published works. However, with technological developments, including the internet and digital media, it has become easier for rightholders to license directly even for secondary uses, such as TDM.

3.1 DIRECT LICENSES

Rightholders, mostly publishers who own the copyright or act as licensors to author researchers, have the option of negotiating licenses directly with academic institutions or through consortia of universities, known as "consortium licenses". This choice is often based on concerns about the impact on their primary markets and the risk of unauthorized use that could affect their agreements with other organizations. For their part, libraries and universities have joined together in consortia to avoid individual negotiation with numerous publishers, representing a large number of libraries from various regions of the world.

Thus, the market for licensed academic content is nowadays vast and diverse, ranging from e-journals to e-books and databases. This ever-expanding market includes diverse business models, forms of content use and multiple stakeholders. In addition to licensing, new services and solutions for research institutions offered through the platforms of large publishers have also been developing.

This approach has given rise to a "primary market" for scholarly licensing, particularly through databases. Large publishers, among others Elsevier, Wiley and Oxford, as well as major universities, sell their copyrighted content, such as journals and books, directly through their own platforms. Smaller publishers are also joining forces to strengthen their offerings and reach users, often through aggregator and distributor platforms, as well as new digital start-ups.

Alternatively, although most digital resources are offered under paid models (purchase, subscription, rental, lending, pay-per-view or other commercial licensing models), academic institutions also use content following the so-called "open model". This topic is discussed below under Licensing and Open Science.

In terms of direct licensing, each publisher tends to have its own model agreement, but standard models have also been developed by publishing organizations, libraries and consortia. One prominent model is the database subscription license, which is widely used globally and recommended by consortia to their members. However, certain clauses in these licenses may generate conflicts in terms of user access to the content offered. These clauses may define the non-exclusive and non-transferable nature of the license, determine the authorized users, as well as the permitted rights and uses, covering aspects such as reproduction, distribution, downloads or internal use within research institutions.

The wide variety of rights and uses allowed in the license agreements enables a whole range of research activities, including TDM (see Annex II). However, these agreements can be a source of misunderstanding and require adequate training and awareness on the part of users (researchers, universities, research institutes) about permitted and prohibited uses. In this regard, machine-readable licenses, such as Creative Commons and software licenses, could offer a solution to improve clarity and understanding of the terms of use.

Most commercial licenses incorporate digital rights management (DRM) and Technological protection measures (TPM) to ensure compliance with terms and conditions, and it is not uncommon that these may conflict with the scope of uses permitted under national Limitations and Exceptions (L&E). In contrast, open access licenses usually do not allow DRM or TPM, with a view to avoid this problem and foster wider access.

Another challenge regarding content availability under direct licensing is that of pricing. Access may be restricted due to (high) subscription costs for individual journals and articles, but also due to (insufficient) shrinking institutional budgets. For this reason, some libraries, universities, or research centers have been reconsidering the content "package" model, opting for less volume and more quality, prioritizing content based on its effective use and relevance rather than a large number of collections that may not be fully utilized. Prioritization in content selection has become crucial as opposed to the traditional business of acquiring large amounts of bundled scholarly content, especially in developing countries. The idea of re-evaluating title by title before investing large budgets in the purchase of content packages has gained ground to ensure effective use of materials.

The fragmented landscape of tools and license types also presents challenges for users, as different resources have different scopes and legal limitations, from unlimited access to per-user or per-territory access restrictions. In addition, linguistic diversity influences licensing practices and the availability of digital resources, leading to variations from country to country.

Not all publishers can offer their content directly from their platforms. Only a few large Science, Technology, Engineering and Mathematics (STM) publishers are able to do so. As a result, much academic content is not available for direct licensing in the marketplace, especially if it comes from small publishers, in minority languages or targeted at smaller territorial markets. Small publishers find it difficult to reach international markets, and many national budgets are focused on acquiring large databases rather than national repositories.

Direct licensing by rightholders is particularly prevalent in the scholarly publishing business, especially when compared to the audiovisual and music industries.[57] For example, the Motion Picture Licensing Corporation (MPLC) license for audiovisual works[58], provided by the Motion Picture Association (MPA) audiovisual distributors is available in different countries but is often considered (too) costly for academic institutions. Even more problematic, some content providers, such as Netflix and Amazon, often refrain from licensing for educational and research purposes.[59] Faced with this situation, academic institutions are looking for alternatives, turning to other streaming services or video licenses available from public institutions, such as national film archives and national broadcasting corporations in some countries.

[57] For more information, see “Report on practices and challenges in relation to distance learning and online research activities” (Document SCCR/39/6)”, prepared by Torres, M./ Xalabarder, R., available at https://www.wipo.int/edocs/mdocs/copyright/en/sccr_39/sccr_39_6.pdf.
[58] See https://www.mplc.org/.
[59] Netflix has original educational documentaries that may be screened for such purposes, provided it is non-profit and access through the Netflix account is restricted to account holders. Amazon Prime does not offer exemptions for research use of its proprietary material, and the terms of use of these services do not grant rights for institutional use. See ibid https://www.wipo.int/edocs/mdocs/copyright/en/sccr_39/sccr_39_6.pdf.

Against this backdrop, direct licensing for research uses offers a mixed picture, depending on sectors and providers, and with uneven availability around the world.[60] Research institutions, especially those supported by public funds, face the dilemma of either devoting time and resources to obtaining a license - with uncertain and limited results - or abandoning use of the work and seeking alternative content - possibly one that is already licensed in some form through a collective license or an alternative license, such as open access.

3.2 COLLECTIVE LICENSES

Collective licensing has proven to be a very efficient system to balance the needs of users for access to content and the protection of copyrights, ensuring adequate compensation to rightholders. Collective licensing is also decisive to secure the enforcement of statutory L&E subject to remuneration or compensation schemes.

However, the availability and licensing practices for research activities are not uniformly developed and vary from country to country, depending on specific legislative, cultural, economic and market aspects. In some countries, the non-existence of CMOs that license research activities, or certain types of works, is an obvious challenge.

Within this category of licensing, three approaches used can be identified: voluntary collective licensing, which is based on the voluntary mandates of members[61]; voluntary licensing supported by legislation, which promotes negotiations through government-approved CMOs; and non-voluntary licensing, which includes statutory and compulsory licenses, requiring grant by law, often with compensation also set by law, and linked to compensation schemes, such as private copying remuneration managed by the CMO.

Collective licensing, mainly by Reproduction Rights Organizations (RROs), focuses on texts and images, covering written works, books, texts, magazines, printed music, among others. These licenses cover offline as well as digital and online activities, expanding in recent decades with the obtaining of digital rights mandates and means of exploitation from rightholders. General, non-exclusive licenses offer permissions such as scanning from paper, digital copies, storage in devices or repositories, enabling sharing in research groups and publishing in secure networks, with quantitative restrictions on the amount of content.[62]

In contrast, specific licensing for TDM is provided mostly by academic publishers, covering both subscribed and non-subscribed content for internal non-commercial research. The practice of collective licensing for TDM is uncommon among CMOs. Some CMOs (such as the Finnish RRO, Kopiosto[63]) include TDM activities in their general licenses for universities and research centers. Another example is the US Copyright Clearance Centre (CCC) license, available worldwide, which allows access and TDM uses of the full text of scholarly articles. In the case of music and audiovisual content for research purposes, collective licensing is even less common.[64]

---

[60] For more information, see “Report on practices and challenges in relation to distance learning and online research activities” (Document SCCR/39/6)”, prepared by Torres, M./ Xalabarder, R., available at https://www.wipo.int/edocs/mdocs/copyright/en/sccr_39/sccr_39_6.pdf.

[61] CMOs must obtain the corresponding mandates from their members before being able to license a new market.

[62] Reproduction Rights Organizations (RROs) have traditionally licensed off-line (analogic) teaching and research activities. Further licensing is directly offered by publishers, through a myriad of individually negotiated as well as consortia agreements.

[63] See https://www.kopiosto.fi/app/uploads/2018/11/11095521/Brochure-The-Kopiosto-copying-licence-Universities_19.pdf.

[64] Yet, the possibilities opened by AI and TDM technologies extend beyond literary and academic works, to all other fields of art, music, audiovisual works, phonograms, as well as recordings of all kinds of sounds. See https://www.wired.com/story/apple-spotify-audiobook-narrators-ai-contract/.

The availability of collective licenses varies widely from country to country due to various factors such as social, economic, cultural ones, copyright awareness, legal framework, existence of organized associations of rightholders, and the very presence of CMOs. In addition, rapid advances in digital technology are transforming research practices, but laws are slow to catch up with these changes, and most CMOs are slow to meet the new needs and standards of research uses, especially in the online world.

In developed countries, licensing for research uses is often aligned with the needs of research institutions. Several factors have contributed to this harmonization, such as a strong legal framework that promotes collective licensing, constructive dialogue between collecting societies and representatives of universities and research institutions, flexibility of publishers and authors in digital licensing, investment in technology by interested parties, and bilateral agreements between collecting societies to grant joint licenses in the form of a consortium, thus strengthening the available repertoire. In contrast, in most developing countries, collective licensing for certain uses has been exceptionally successful, as the presence and operation of CMOs is limited, preventing them from fully meeting the needs of research users. In this case, it is important to highlight that the efficiency of collective licensing (and the development of collective management) can be achieved, for instance, by means of enacting L&E provisions that permit specific uses subject to compensation managed via CMOs, or by means of establishing specific instances of mandatory collective management or extended collective licensing (ECL).

In summary, the availability of collective licenses also does not follow a uniform pattern and faces a number of challenges. From the users' perspective, one common obstacle is the lack of understanding of the need for a license to use a broad repertoire of works in research, which has taken years for some CMOs to overcome. Confusion over the scope of permitted uses under L&E to copyright also creates legal uncertainty, undermining the sale of licenses for research activities. Negotiating the scope of the license appears to be especially confusing for academic and research institutions, as there is often an expectation of being able to copy the entire work in exchange for payment, and licenses do not clearly distinguish uses that are already exempted by a license to use and those that need a specific license.

Furthermore, identifying and locating authors or copyright holders, obtaining timely responses and excessive pricing are obstacles identified when trying to obtain a license from copyright holders, CMOs or copyright agencies. From the CMOs' perspective, collective management faces challenges in developing licenses that respond to new research needs, expanding their availability to new works and territories, raising awareness of the benefits of collective licensing, and working to expand their presence in different regions of the world.

In this regard, a toolkit presenting best practices along with revisions and updates of the copyright law could contribute significantly to the development of collective management licenses to meet research needs. Examples include legal solutions such as "L&E failing licenses"[65] and extended collective licenses (ECL),[66] which extend the scope of voluntary licenses agreed with CMOs and compulsory licenses subject to remuneration for research uses.

3.3 LICENSES AND OPEN SCIENCE

Technological evolution has radically transformed access to scholarly publications. On the one hand, academic publishing markets shifted from analog to digital formats, expanding licensing strategies through massive catalogs of publications and databases under subscription models

[65]A statutory non-remunerated L&E applies if no voluntary agreement is reached among parties.
[66]A license that extends the scope of a voluntary license agreed with a CMO beyond its repertoire and partners, to all works and authors in the same category, as well as compulsory (non-voluntary) licenses.

(known as the "Big Deals"),[67] rather than individual titles and collections. On the other hand, academic institutions soon identified the importance of digital platforms to ensure a wide dissemination of their research results that would allow subsequent uses at lower cost, developing strategies within the so-called “Open Science”, which aims to foster reproducibility, global collaboration and scientific integrity through greater transparency, openness and accessibility of research results and data (see Annex III).

In the area of copyright, there are three elements of Open Science strategies that are particularly relevant to this Study:

- Open access to research publications, including open access publications.
- Open licenses, such as Creative Commons licenses that facilitate the reuse of research results (publications and software).
- Open data, which provides access to structured data generated and/or used for research purposes.

The main challenges with respect to these elements lie in how copyright is managed in these new contexts. It involves clearly defining who retains the copyright on works published under open access through agreements, how these rights are distributed among authors and institutions, and how the appropriate use of works is guaranteed in terms of licenses and rights, as opposed to the principles and policies to guarantee the availability and free use of scientific research.

A first challenge is to find a balance between the common practice of self-archiving, driven by open access mandates, and the conditions imposed by publishers. Although self-archiving has been accepted and permitted by many publishers, there are restrictions, such as the delay in public access (embargo) and the limitation to the "post-print" version rather than the final version. In addition, the choice made by researchers among the several Creative Commons licenses may hinder the reuse of documents in open repositories: the challenge here is ensuring greater open access and reuse, while respecting the licensing terms (and scope) imposed by the licenses.[68]

Another challenge is the transition from the traditional subscription model to the emerging Open Access model, especially through "Read and Publish Agreements" or "Transformative Agreements". These agreements seek to transform the reading and publishing of scholarly research by including provisions that allow researchers to publish in fee-paying journals without incurring individual costs. However, challenges here concern financial sustainability, equity in global access, and equitable participation of diverse institutions, as they tend to be more common in economically developed areas, leaving other regions of the world with limited resources at a disadvantage.

The evolution of open access mandates for researchers, specifically in relation to the results of their funded research, is also another concern. Initially, certain temporal flexibilities were allowed, such as an embargo period before results were made open access. However, this dynamic has changed over time, and researchers are now being urged to publish their results immediately and without access restrictions, granting rights for reuse. In addition, some governments and public funders do not include certain costs – such as those required to cover

[67] See ALLEA, “Statement on Open Access Publication under “Big Deals” and the new Copyright Rules” 17 (2023) JIPITEC 222, para 1, available at https://www.jipitec.eu/issues/jipitec-14-1-2023/5718.

[68] In principle, the CC license that best matches the very definition of Open Access is the Attribution - Creative Commons License, also known as CC BY (https://creativecommons.org/licenses/by/4.0/) which allows any act of exploitation, also for commercial purposes. However, there has been a lot of debate on the use of this license and on whether Open Access goals could also be achieved by means of other CC licenses that are more protective of authors’ interests. In 2017, a group of Latin American Institutions launched the Mexico Declaration advocating for an ecosystem of non-commercial Open Access and supporting, instead, the use of the CC BY-NC-SA license for Open Access journals, which allows complying with OA mandates, including reuse and the making of derivatives works as long as for non-commercial purposes. See: https://redalyc.org/redalyc/documentos/Declaracion-Mexico.pdf.

APC payments – in the budgets of funded projects, which makes it difficult to maintain the hybrid model of publication (subscription and open access).

There are also challenges inherent to the application of solutions at the national or regional level. This is the case of how to define and apply the secondary publication right for authors of scientific publications. Questions arise as to its nature, whether as an inalienable moral right or as a reservation to the exclusive assignment of exploitation rights. Furthermore, it is pointed out that open access implies not only access but also unrestricted use, and the secondary right allows authors to republish the work but says nothing about the recovery of ownership of rights that may have been assigned to the publisher on an exclusive basis. Another challenge is to determine whether this right should be available immediately upon publication or after an embargo period.

Finding effective methods for retaining the researcher's rights as author of the work, especially in earlier versions, is not easy. It is often proposed that researchers transfer only the copyright on the final published version and retain their rights on any earlier versions. This happens in initiatives that seek open access for publicly funded research publications. But at the same time, there is also a need for institutional policies that recognize research institutions as original rightholders and grant authors sufficient rights to proceed with publication.

In the field of open licensing, especially in the free and open-source software arena, one of its key challenges lies at the intersection with artificial intelligence. Advances in AI are transforming the way code is generated, potentially improving efficiency and accuracy for developers. However, there have been legal disputes when authors of open-source software have sued companies that use proprietary software to train AI systems and produce new code.[69]

As for open data, although the data itself is not copyrighted, the structuring and presentation of this data, such as in databases or compilations, may be subject to copyright depending on its originality, or even subject to sui generis protection, as mentioned above. The complexity lies in finding the balance between the openness of data to encourage open science and the need to restrict access not only for copyright reasons, but also in situations where privacy, confidentiality or security require it. In addition, varying national and regional approaches to database protection may lead to differing interpretations of the extraction and reuse of database content.[70]

4. ADDITIONAL CHALLENGES

As we have seen, the better and more precise a definition of L&E in national laws, meticulously balancing the rights of copyright holders, the public access to knowledge, and the indispensable role of research in driving innovation, the more effective these L&E will be.

However, well defined L&E are not enough. Further challenges may derive from the use of licensing terms and TPMs, as well as from the very territorial structure of the international protection of copyright.

Licenses represent versatile tools that have a dual and flexible ability to delineate the terms and conditions governing the use, modification, or distribution of a copyrighted work in research contexts. On the one hand, they can impose additional requirements beyond what would be permitted by copyright exceptions (overriding of exceptions by contract or interplay of TPM and

[69] Furthermore, open licensing is also developing in the field of AI systems. For instance, Open Responsible AI Licenses (Open RAIL) are licenses designed to permit free and open access, re-use, and downstream distribution of derivatives of AI artifacts, as long as the behavioral-use restrictions always apply (including to derivative works). See https://www.licenses.ai/blog/2022/8/26/bigscience-open-rail-m-license.

[70] https://www.ifla.org/wp-content/uploads/2019/05/assets/hq/topics/exceptions-limitations/documents/100contracts.pdf.

L&E) and, on the other hand, they can also relax typical restrictions imposed by copyright law, extending permissions beyond the usual limits (cross-border uses).

First, the overridability of exceptions becomes apparent when, for example, despite the existence of a copyright exception authorizing the use of a work for research purposes without the need for permission from the rightholder, a specific license associated with that work imposes restrictions on such use, effectively negating the intent of the exception. The need to preserve the delicate balance achieved under statutory L&E (usually designed in view of the specific circumstances in each country) highlights not only the critical importance of well-defined licensing agreements, but also the need to formally protect permitted uses under statutory L&E against any overriding licensing terms, as several countries have already done.[71]

Second, the interplay between L&E and TPMs reveals a complex dynamic that influences the use of copyrighted content, especially in research contexts. Technological measures, viewed by some in dire terms and by others as positive catalysts, are central to the modern digital copyright landscape. They provide rightholders with the means to disseminate their works in the digital market and control subsequent copying or exploitation, thus potentially ensuring fair compensation for their creative efforts. Digital markets have enlarged access to the production and dissemination of works, empowering authors to reach a wider public. The exploitation of works in digital markets requires the use of TPMs, but at the same time, such measures may hinder research efforts, as well as other purposes that benefit from specific statutory L&E. Once again, well-defined L&E for research purposes, that harmonize the interests of copyright holders, researchers, and research institutes are not enough, and often specific statutory declarations are necessary to secure that L&E will prevail over the enforcement of TPMs in specific cases.[72,73]

Finally, licenses can easily address cross-border issues by identifying the territorial scope of the licensed rights and content. A common challenge in the field of L&E and licensing of research activities stems from the territorial nature of copyright laws. When developing research activities, researchers and research institutes are often located in different jurisdictions. When assessing whether a specific research use can be exempted under

---

[71] Belgium: Belgian Code de droit économique, Art. XI.193
Kuwait: Kuwait Law No. (75) of 2019, Art. 31
European Union (EU): 1991 Software Directive, Art. 9, and 2009 Software Directive, Art. 8
Ireland: Copyright and Related Rights Act (2000), Sec. 2(10)
Montenegro: Montenegro Copyright Law of 2016, Art. 45, 113, 114
Portugal: Decreto-Lei nº 9/2021, Art. 75(5), 82B
United Kingdom (UK): Copyright, Designs and Patents Act 1988, Sec. 50A, 50B, 50BA, 50D

[72] Article 7 of the Marrakesh Treaty underscored the importance of such mechanisms. It mandates Contracting parties to enact necessary measures ensuring legal protection and effective remedies against the circumvention of technological measures, without obstructing the intended benefits of this treaty for the individuals it aims to assist. In essence, it emphasizes that the pursuit of legal safeguards in the exploitation of copyrighted works should not impede the rightful enjoyment of the limitations and exceptions granted to those in need under the provisions of the Marrakesh Treaty.

[73] In the EU, for computer programs and databases, TPMs are unable to hinder E&L, but under the InfoSoc Directive, particularly Art.6.4(4), TPMs take precedence over L&E, albeit with certain E&L aspects receiving legal protection. Notably, directives do not explicitly address contractual terms. Conversely, the Copyright in the Digital Single Market (CDSM) Directive firmly designates contractual provisions conflicting with mandatory L&E as unenforceable, notably for crucial domains like TDM, education, research, cultural heritage, and out-of-commerce works. However, by referencing Art.6(4) of the InfoSoc Directive, concerns arise regarding potential TPM hindrances in effectively upholding these essential L&E provisions.
In USA, on October 28, 2021, the Librarian of Congress, pursuant to 17 U.S.C. 1201(a)(1)(C) and (D), published the classes of copyrighted works that shall, for a three-year period, be exempted from the prohibition against circumvention of technological measures that effectively control access to copyrighted works, which include multiple scenarios dealing with research.
In Singapore, Section 431 of the Copyright Act 2021 provides an exception for research on encryption technology.
In Australia, Section 116 AN (4) of the Copyright Act 1968 sets forth an exception from TPMs for research on encryption technology.
In Mexico, Article 114 Quarter and Quinquies of the Copyright Act 1996 establishes exceptions from TPMs provisions for research-related activities.

specific L&E, multiple national copyright laws should be consulted and territorially enforced. It is difficult, if not impossible, to consider all of these laws in detail and to comply with them all, more so when they lead to contradictory results (what may be permitted in one country under a L&E may not be so in another country, and so on).

Beyond L&E, "territorial" discrepancies can also be identified in terms of licensing. Often, there is a disparity between the territorial scope of licenses acquired for research purposes (often formally restricted to a single country) and the scope of research activities, which may occur across multiple territories. This is especially the case of collective licenses, obtained from CMOs, which for the most part essentially undertake territorial licensing. However, through bilateral agreements with other CMOs, they may be able to offer solid repertoire licenses, as in the case of consortia that include countries with the same language. Instead, individual and direct licensing may more easily address cross-border issues by identifying the territorial scope of the licensed rights and content covered.[74]

[74] For instance, the University of the West Indies has ongoing negotiations with RROs from the Caribbean Region for facilitating the access to materials from different countries within the regional framework of the university. Indeed, five Caribbean RROs formed a regional body – CARROSA – seeking for a solution for cross-border matters. See further Torres, M./ Xalabarder, R. (2019) "Report on Practices and Challenges in Relation to Online Distance Education and Research Activities" (Document SCCR/39/6) and, most recently, the Presentation on Online Cross-Border Uses of Works for Education and Research from the 43rd session of the SCCR on March 16, 2023, https://webcast.wipo.int/video/SCCR_43_2023-03-16_AM_118726.

5. CONCLUSIONS

Digital technologies are transforming the way research is conducted and there is no single answer to the various issues they raise. Each country must customize its plans to take advantage of the opportunities while effectively addressing the challenges in the copyright arena. This includes not only creating appropriate L&E systems, but also creating licenses that fit the needs of researchers. These approaches are not mutually incompatible and moreover could be complementary depending on the context in which research is carried out and in accordance with the priorities of each Member State to promote the advancement of research.

The scope of L&E is closely related to the availability of licenses in each country. This interconnectedness specifies the operational activities that require licensing and, in turn, has an impact on the development of L&E provisions. Finding the appropriate balance between these factors remains in the hands of Member States.

On the one hand, international treaties have recognized the importance of L&E in national legislation for research purposes. It is up to the Member States to apply the flexibilities arising from L&E according to the new technological means of exploitation. This adaptation must always meet the three-step criteria, ensuring a careful balance between the rights of rightholders and the public interest in research. Unclear and outdated L&E provisions create legal uncertainty and ultimately jeopardize the development of research activities, particularly in digital and online media, including across borders.

On the other hand, rightholders have the possibility to grant licenses for use beyond those specifically permitted by law through L&E, either individually or through collective management organizations. Historically, rightholders managed licenses for primary uses, and CMOs managed licenses for secondary uses of published works. However, with advances in technology, including the Internet and digital media, it has become easier for rightholders to license directly even for secondary uses, such as TDM.

Direct licensing for research uses offers a heterogeneous picture, depending on sectors and providers, and with uneven availability around the world. Licensing practices for research activities are not uniformly developed and vary from country to country, depending on specific legislative, cultural, economic and market aspects. In some countries, the absence of CMOs licensing research activities, or certain types of works, is also an obvious challenge.

Licensing is an enabling tool to foster cross-border research collaborations and addressing the special needs of academic institutions. A key issue is to ensure that research operations can take advantage of the potential offered by digital technologies across jurisdictions.

In addition, the exploitation of works in digital markets requires the use of digital rights management tools but, at the same time, such measures may hinder research efforts as well as other endeavors that benefit from specific statutory L&E. Specific statutory declarations or other mechanisms are needed to ensure that L&E will prevail over the application of TPMs in specific cases. Similarly, related mechanisms are needed to protect uses permitted under statutory L&E from any license conditions that unduly restrict them.

TDM, intrinsically linked to the development of AI projects, has enormous potential to reshape research worldwide by enabling machines to "read" data (copyrighted and non-copyrighted content) and extract important patterns and connections. The question of whether TDM should be allowed for research purposes, and the scope of such permission, has already been addressed at the legislative level by a few Member States. Before any legislative step regarding TDM, it is critical to carefully assess its potential benefits for scientific and research advancement, while taking into account the implications for the legitimate rights of authors and rightholders in each jurisdiction.

Open Science strategies, in particular Open Access programs and Creative Commons licenses, have been very successful in ensuring access and authorizing reuse of works, particularly for non-commercial uses. These programs are essential for disseminating research results worldwide and ensuring subsequent access to research outputs. The main challenges lie in how copyright is managed in these new contexts. It is important to clearly define who retains the copyright in works published under open access through agreements, how these rights are distributed among authors and institutions, and how the appropriate use of the works is guaranteed in terms of licenses and rights, as opposed to the principles and policies to ensure the availability and free use of scientific research.

As technological advances transform research environments, it would be beneficial to map best practices and updated statutory provisions. This approach could be helpful in identifying the various conditions and flexibilities governing the use, modification or distribution of copyrighted works in research contexts. This could further guide future developments to keep pace with new technologies, while taking into account the constantly changing landscape of copyright in research.

ANNEX I: TEXT AND DATA MINING[75]

As seen in chapter 2 of this Study, only a few countries have enacted L&E for text and data mining (TDM) for research purposes. This Annex will further examine a few selected national provisions without aiming at being exhaustive.

As commonly accepted, TDM is quite useful and valuable in the era of the data economy. However, TDM activities may entail machine reading of copyrighted works (i.e., articles, images, movies, music) and other protected subject matter (i.e., performances, phonograms). It is true that a transient machine reading can be excluded from the concept of reproduction covered by the right of reproduction.[76] But TDM activities often entail making (temporary or permanent) copies which are subject to copyright and related rights.[77] If it is not possible to obtain necessary authorizations from rightholders, TDM activities may constitute copyright infringement.

Therefore, a provision on limitations and exceptions to copyright and related rights (hereinafter referred to as 'copyright exception') plays a significant role to facilitate TDM activities. In fact, TDM copyright exception is currently one of the hottest topics in the field of copyright law.

This Annex will show national and regional legal provisions on copyright exceptions for facilitating TDM activities.

1. L&E FOR TEXT AND DATA MINING

a) United Kingdom

The U.K. introduced the provision (Art.29A) for text and data analysis in 2014 in the Copyright, Designs and Patents Act (CDPA, 1988).[78]

> ***Section 29A [Copies for text and data analysis for non-commercial research]***
> *(1) The making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that—*
> *(a) the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose, and*
> *(b) the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise).*
> *(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—*
> *(a) it is to be treated as an infringing copy for the purposes of that dealing, and*

[75] Prepared by Professor Tatsuhiro Ueno.
[76] *See* Art.5(1) of the InfoSoc Directive (2001/29/EC) ('[t]emporary acts of reproduction …, which are transient or incidental [and] an integral and essential part of a technological process and whose sole purpose is to enable: (a) a transmission in a network between third parties by an intermediary, or (b) a lawful use of a work or other subject-matter to be made, and which have no independent economic significance, shall be exempted from the reproduction right provided for in Article 2').
[77] *See* Art.2(1) of the InfoSoc Directive (2001/29/EC) ('[M]ember States shall provide for the exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part,' as the right of reproduction).
[78] *See* Burrow, S., "The Law of Data Scraping: A review of UK law on text and data mining" (2021).

(b) 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.

(i) Beneficiaries

There is no particular provision limiting the scope of the beneficiaries of the U.K. TDM copyright exception, although it provides for 'a person who has lawful access to the work', which will be touched upon below with regard to the 'lawful access requirement'.

(ii) Purpose

The U.K. TDM copyright exception only applies to the analysis "for the sole purpose of research for a non-commercial purpose".
However, according to the U.K. Intellectual Property Office (IPO), the research output can be used for commercial purposes in so far as the original purpose of carrying out TDM is solely non-commercial.[79]

(iii) Lawful access requirement

The U.K. TDM exception applies to 'a person who has lawful access to the work' (Art.29A(1)(a)). Someone who obtained works by a copyright infringing act cannot benefit from this TDM copyright exception.

(iv) Allowed acts

Under the U.K. TDM copyright exception, '[t]he making of a copy of a work' in 'a computational analysis of anything recorded in the work' is allowed (Art.29A(1)). If the copy made under this TDM copyright exception is transferred to any other person, it constitutes a copyright infringement (Art.29A(2)(a))

(v) TDM not for 'research purposes'

Under the U.K. TDM copyright exception, only the analysis 'for the sole purpose of research for a non-commercial purpose' is allowed. Hence, TDM activities for commercial purposes are not permitted under this TDM copyright exception.

[79] *See* Intellectual Property Office (IPO), Exceptions to copyright: Research, p.10 (2014) ('There are no restrictions on how or where outputs of text and data mining can be published, including journals published for profit by academic publishers and under licences that permit commercial research, such as CCBY. Other commercialisation of the research outputs is not restricted either. But it is important to be scrupulous in assessing whether the original purpose of carrying out the text and data mining analysis is solely non-commercial; if it is not, then researchers are very likely to be infringing copyright.'). Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf.

In 2021, the U.K. Government considered amending the TDM copyright exception to allow TDM activities also for commercial purposes by anyone with lawful access to material protected by copyright.[80] This amendment was not carried on.

> The U.K. Intellectual Property Office, Artificial Intelligence and Intellectual Property: copyright and patents: Government response to consultation (Updated on 28 June 2022) concluded:
>
> *58. The Government has decided to introduce a new copyright and database right exception which allows TDM for any purpose. The Government will identify suitable legislation to make the required changes in due course.*
>
> *59. Introducing an exception which applies to commercial TDM will bring benefits to a wide range of stakeholders in the UK. These include researchers, AI developers, small businesses, cultural heritage institutions, journalists, and engaged citizens. Targeted products and services will benefit businesses and customers. Research outcomes could also benefit the wider public. This could be, for example, by supporting research and innovation in public health. Some in the creative industries also use TDM and AI to understand their market or create new works – they will also see benefits. The benefits will be reducing the time needed to obtain permission from multiple rights holders and no licence fee to pay. This will speed up the TDM process and development of AI.*
>
> *60. These changes make the most of the greater flexibilities following Brexit. They will help make the UK more competitive as a location for firms doing data mining.*
>
> *61. Rights holders will no longer be able to charge for UK licences for TDM and will not be able to contract or opt-out of the exception. The new provision may also affect those who have built partial business models around data licensing. However, rights holders will still have safeguards to protect their content. The main safeguard will be the requirement for lawful access. That is, rights holders can choose the platform where they make their works available, including charging for access via subscription or single charge. They will also be able to take measures to ensure the integrity and security of their systems.*
>
> *62. The Government's ambition is to make the UK a global centre for AI innovation. The new exception will ensure the UK's copyright laws are among the most innovation-friendly in the world. All users of data mining technology will benefit, with rights holders having safeguards to protect their content.*

(vi) Contractual overridability

Art.29A(5) of the CDPA stipulates, '[t]o 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.' Hence, a contractual provision prohibiting TDM activities for the sole purpose of research for a non-commercial purpose is unenforceable.

[80] The U.K. Intellectual Property Office, Artificial Intelligence and Intellectual Property: copyright and patents: Government response to consultation (Updated 28 June 2022). Available at https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-to-consultation.

(vii) Technological overridability

There is no provision regarding technological overridability for the U.K. TDM copyright exception.

(viii) Safeguard

There is no provision regarding safeguards for the U.K. TDM copyright exception.

b) EU Directive

The Directive on Copyright in the Digital Single Market (2019/790) on 17 April 2019 (hereinafter referred to as the CDSM Directive')[81] has copyright exceptions for TDM in Articles 3 and 4 as mandatory copyright exceptions, which were extended from the Proposal for a Directive on Copyright in the Digital Single Market on 14 September 2016 (Art. 3).[82]

> ***Article 3 [Text and data mining for the purposes of scientific research]***
> *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.*
>
> ***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.*

[81] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, OJ L 130, 17.5.2019, p.92–125. Available at https://eur-lex.europa.eu/eli/dir/2019/790/oj.
[82] Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market, COM/2016/0593 final - 2016/0280 (COD). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016PC0593.

(i) Beneficiaries

The beneficiaries under Art.3 of the CDSM Directive are 'research organizations and cultural heritage institutions.'

- Research organizations

Art.2(1) of the DSM Directive defines 'research organization' as 'a university, including its libraries, a research institute or any other entity, the primary goal of which is to conduct scientific research or to carry out educational activities involving also the conduct of scientific research: (a) on a not-for-profit basis or by reinvesting all the profits in its scientific research; or (b) pursuant to a public interest mission recognized by a Member State; in such a way that the access to the results generated by such scientific research cannot be enjoyed on a preferential basis by an undertaking that exercises a decisive influence upon such organization.'

According to Recital 11 of the CDSM Directive, research organizations with public-private partnerships may benefit from this TDM exception[83] and research organizations and cultural heritage institutions may rely on their private partners for carrying out text and data mining.[84]

According to Recital 12 of the CDSM Directive, not only universities and their libraries but also entities such as research institutes and hospitals that carry out research can be beneficiaries under Article 3 of the DSM Directive.[85]

- Cultural heritage institutions

Art.2(3) of the CDSM Directive defines 'cultural heritage institution' as 'a publicly accessible library or museum, an archive or a film or audio heritage institution'.[86]

(ii) Purpose

Under Art.3(1) of the CDSM Directive, TDM activities 'for the purposes of scientific research' are allowed.

The DSM Directive itself does not define 'scientific research'. According to Recital 12 of the DSM Directive, the terms 'scientific research' within the meaning of the CDSM Directive 'should be understood to cover both the natural sciences and the human sciences.'

---

[83] Recital 11 of the DSM Directive states, 'research organisations should also benefit from such an exception when their research activities are carried out in the framework of public-private partnerships.'
[84] Recital 11 of the DSM Directive states, 'While research organisations and cultural heritage institutions should continue to be the beneficiaries of that exception, they should also be able to rely on their private partners for carrying out text and data mining, including by using their technological tools'.
[85] Recital 12 of the DSM Directive states, 'Due to the diversity of such entities, it is important to have a common understanding of research organisations. They should for example cover, in addition to universities or other higher education institutions and their libraries, also entities such as research institutes and hospitals that carry out research. Despite different legal forms and structures, research organisations in the Member States generally have in common that they act either on a not-for-profit basis or in the context of a public-interest mission recognised by the State. Such a public-interest mission could, for example, be reflected through public funding or through provisions in national laws or public contracts. Conversely, organisations upon which commercial undertakings have a decisive influence allowing such undertakings to exercise control because of structural situations, such as through their quality of shareholder or member, which could result in preferential access to the results of the research, should not be considered research organisations for the purposes of this Directive.'
[86] Recital 13 of the DSM Directive states, 'Cultural heritage institutions should be understood as covering publicly accessible libraries and museums regardless of the type of works or other subject matter that they hold in their permanent collections, as well as archives, film or audio heritage institutions. They should also be understood to include, inter alia, national libraries and national archives, and, as far as their archives and publicly accessible libraries are concerned, educational establishments, research organisations and public sector broadcasting organisations.'

According to one theory, “the notion of ‘scientific research’ is to be considered alongside that of TDM in Article 2 No. 2 […] and be intended as encompassing any activity aimed at generating information that allows the uncovering of new knowledge or insights that are based on or characterized by the methods and principles of science”.[87]

(iii) Lawful access requirement

Article 3(1) of the CDSM Directive permits ‘text and data mining of works or other subject matter to which they have lawful access.’

Contents that are freely available online are considered as lawfully accessible works or other subject matter. According to Recital 14 of the CDSM Directive, in the case of subscriptions taken by research organizations or cultural heritage institutions, the persons covered by those subscriptions should be deemed to have lawful access.[88]

(iv) Allowed acts

Under the CDSM Directive, “reproductions and extractions” (Art.3(1) and Art.4(1)) and retaining in certain cases (Art.3(2) and Art.4(2)) are allowed. Other types of exploitation of works or other subject matter such as communication to the public would not be subject to TDM copyright exception in the DSM Directive.

(v) TDM not for ‘research purposes’

Under Art.4 of the CDSM Directive, TDM activities for a commercial purpose or by someone other than ‘research organizations and cultural heritage institutions’ are also permitted, unless ‘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.’ That means that right holders opt out by expressly reserving their rights for TDM activities for commercial purposes or by a business company.

(vi) Contractual overridability

Art.7(1) of the CDSM Directive stipulates, ‘[a]ny contractual provision contrary to the exceptions provided for in Articles 3, 5 and 6 shall be unenforceable.’

Hence, Art.3 of the CDSM Directive is not overridable by contract, while Article 4 is overridable. If someone acquires a work based on a contract and afterwards copies it for commercial TDM in breach of that contract, the act of copying would not only be regarded as a breach of contract but also can constitute copyright infringement in case the right holder made an appropriate reservation.

---

[87] *See* Rosati, E., Copyright in the Digital Single Market: Article-by-Article Commentary to the Provisions of Directive 2019/790, (Oxford Univ. Press, 2021) p.43.

[88] Recital 14 of the DSM Directive states, ‘Lawful access should be understood as covering access to content based on an open access policy or through contractual arrangements between rightholders and research organisations or cultural heritage institutions, such as subscriptions, or through other lawful means. For instance, in the case of subscriptions taken by research organisations or cultural heritage institutions, the persons attached thereto and covered by those subscriptions should be deemed to have lawful access’.

(vii) Technological overridability

Under the second sentence of Art.7(2) of the CDSM Directive, by the application of the first subparagraph of Art.6(4) of the InfoSoc Directive[89], '[n]otwithstanding the legal protection of technological measures, […] Member States shall take appropriate measures to ensure that rightholders make available to the beneficiary of an exception or limitation […] the means of benefiting from that exception or limitation, to the extent necessary to benefit from that exception or limitation and where that beneficiary has legal access to the protected work or subject-matter concerned', for Arts.3 to 6 of the CDSM Directive.

(viii) Safeguard

Under the first sentence of Art.7(2) of the CDSM Directive, by the application of Art.5(5) of the InfoSoc Directive, the limitation and exception 'shall only be applied in certain special cases which do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the rightholder'. That means that TDM copyright exception in the CDSM Directive shall be applied in compliance with the three-step test.

c) National laws of EU Member States

The EU Member States had to implement the CDSM Directive by 7 June 2021. Some of them have already implemented it, while others have not.[90] The situation is evolving in real time.[91]

Since the EU Member States 'shall provide for an exception' (Arts.3 and 4 of the CDSM Directive), the TDM copyright exception under the CDSM Directive is a mandatory provision. Considering also that the list of Limitations and Exceptions contained in Art.5 of the InfoSoc Directive is exhaustive (*see* Recital 32),[92] the EU Member States may not be able to introduce either broader or narrower TDM copyright exceptions. Thus, the scope of the TDM copyright exception is supposed to be basically common in the EU Member States. In practice, the details of each TDM copyright exception may differ slightly among Member States that have already implemented the CDSM Directive, especially in terms of the beneficiaries, the allowed acts and the method of reservation.

To exemplify these differences, this subchapter gathers provisions from a few EU Member States.

---

[89] "4. Notwithstanding the legal protection provided for in paragraph 1, in the absence of voluntary measures taken by rightholders, including agreements between rightholders and other parties concerned, Member States shall take appropriate measures to ensure that rightholders make available to the beneficiary of an exception or limitation provided for in national law in accordance with Article 5(2)(a), (2)(c), (2)(d), (2)(e), (3)(a), (3)(b) or (3)(e) the means of benefiting from that exception or limitation, to the extent necessary to benefit from that exception or limitation and where that beneficiary has legal access to the protected work or subject-matter concerned."

[90] On 19 May 2022, the EU Commission sent reasoned opinions to Belgium, Bulgaria, Cyprus Denmark, Greece, France, Latvia, Poland, Portugal, Slovenia, Slovakia, Finland and Sweden for failure to notify the Commission of transposition measures on copyright and related rights in the DSM Directive (https://ec.europa.eu/commission/presscorner/detail/en/IP_22_2692).

[91] *See* also 'DSM Directive Implementation Tracker' available at https://www.notion.so/DSM-Directive-Implementation-Tracker-361cfae48e814440b353b32692bba879, 'CDSM Implementation Resource Page' available at https://www.create.ac.uk/cdsm-implementation-resource-page/.

[92] Recital 32 of the Info-Soc Directive states, 'This Directive provides for an exhaustive enumeration of Limitations and Exceptions to the reproduction right and the right of communication to the public.'

(i) Germany

The German Copyright Act was amended in 2017 to introduce a TDM copyright exception (§60d [Text und Data Mining]) which only applied to TDM for non-commercial scientific research ("für die wissenschaftliche Forschung"). The 2021 amendment[93] revised §60d [Text und Data Mining für Zwecke der wissenschaftlichen Forschung] and newly introduced §44b [Text und Data Mining] in line with Art.4 of the CDSM Directive, under which TDM for commercial purposes is also allowed unless the right holder has not reserved their right (§44b(3)).

> ***§44b [Text and data mining][94]***
> *(1) 'Text and data mining' means the automated analysis of individual or several digital or digitised works for the purpose of gathering information, in particular regarding patterns, trends and correlations.*
> *(2) It is permitted to reproduce lawfully accessible works in order to carry out text and data mining. Copies are to be deleted when they are no longer needed to carry out text and data mining.*
> *(3) Uses in accordance with subsection (2) sentence 1 are permitted only if they have not been reserved by the rightholder. A reservation of use in the case of works which are available online is effective only if it is made in a machine-readable format.*
>
> ***§60d [Text and data mining for scientific research purposes]***
> *(1) It is permitted to make reproductions to carry out text and data mining (section 44b (1) and (2) sentence 1) for scientific research purposes in accordance with the following provisions.*
>
> *(2) Research organisations are authorised to make reproductions. 'Research organisations' means universities, research institutes and other establishments conducting scientific research if they*
> *1. pursue non-commercial purposes,*
> *2. reinvest all their profits in scientific research or*
> *3. act in the public interest based on a state-approved mandate.*
> *The authorisation under sentence 1 does not extend to research organisations cooperating with a private enterprise which exerts a certain degree of influence on the research organisation and has preferential access to the findings of its scientific research.*
>
> *(3) The following are, further, authorised to make reproductions:*
> *1. libraries and museums, insofar as they are accessible to the public, and archives or institutions in the field of cinematic or audio heritage (cultural heritage institutions),*
> *2. individual researchers, insofar as they pursue non-commercial purposes.*
>
> *(4) Those authorised in accordance with subsections (2) and (3) and pursuing non-commercial purposes may make reproductions made pursuant to subsection (1) available to the following persons:*
> *1. a specifically delimited circle of persons for their joint scientific research and*
> *2. individual third persons for the purpose of monitoring the quality of the scientific research.*
> *The making available to the public must be terminated as soon as the joint scientific research or the monitoring of the quality of the scientific research has been concluded.*
>
> *(5) Those authorised under subsections (2) and (3) no. 1 may retain reproductions made pursuant to subsection (1), thereby taking appropriate security measures to prevent*

[93] Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes vom 31. Mai 2021, Bundesgesetzblatt Teil I, S.1204, available at http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl121s1204.pdf.
[94] Non-official translation is available at https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html#p0328.

unauthorised use, for as long as they are needed for the purposes of the scientific research or the monitoring of the quality of the scientific findings.
(6) Rightholders are authorised to take necessary measures to prevent the security and integrity of their networks and databases being put at risk on account of reproductions made in accordance with subsection (1).

The German TDM copyright exception permits also making the copy made for TDM for non-commercial purposes available to a specifically delimited circle of persons for their joint scientific research and individual third persons for the purpose of monitoring the quality of the scientific research (§60d(4)).

§66d does not apply to works of computer programs (§69d(6)).

Under §60g [Gesetzlich erlaubte Nutzung und vertragliche Nutzungsbefugnis] of the German Copyright Act, in line with Art.7(1) of the CDSM Directive, the right holder may not rely on agreements restricting or prohibiting permitted uses under §60d to the detriment of the right holders. That means, the contractual provision contrary to the TDM exceptions for non-commercial scientific research under Article 60d shall be unenforceable.

Under §95b(1) [Durchsetzung von Schrankenbestimmungen] of the German Copyright Act, in line with Art.7(2) of the CDSM Directive, where a right holder applies technological measures, he or she shall be obliged to make available to the beneficiaries of copyright exception legal access to the works (including §44b and §60d), to the extent that they have lawful access to the work or subject matter, the means necessary to make use of those provisions to the extent required. The beneficiary may require the right holders to provide the means necessary for that (§95b(2)).

(ii) Austria

The Austrian TDM copyright exception (§42h) was introduced by the amendment in 2021.[95]

§42h [Text and data mining] (tentative translation by the Author)
(1) Any person may reproduce a work for a research institution (subsection 3) or for a cultural heritage institution (§42(7)) for the purpose of automated analysis of texts and data in digital form for scientific or artistic research and to obtain information on, inter alia, patterns, trends and correlations, if he or she has lawful access to the work. Individual researchers shall also be entitled to make such reproductions, provided that this is justified for non-commercial purposes.

(2) A reproduction under subsection (1) may be stored and kept subject to reasonable security precautions as long as this is justified by the purpose of the research, including the verification of scientific findings. Reasonable security measures shall be those which have been recognised as best practice by representative associations of right holders on the one hand and research institutions or cultural heritage institutions on the other. Such reproduction may also be made available to a defined group of persons for the purpose of their joint scientific research or to individual third parties for the purpose of verifying the quality of scientific research, provided that this is justified for non-commercial purposes.

(3) A research institution within the meaning of this provision is an institution,
1. whose primary objective is scientific or artistic research or research-led teaching, and

[95] Bundesgesetz, mit dem das Urheberrechtsgesetz, das Verwertungsgesellschaftengesetz 2016 und das KommAustria-Gesetz geändert werden (Urheberrechts-Novelle 2021 – Urh-Nov 2021), BGBl. I Nr. 244/2021, available at https://www.ris.bka.gv.at/eli/bgbl/I/2021/244.

2. which is not profit-oriented in its activities, reinvests all profits in its scientific or artistic research or is profit-oriented and operates within the framework of a state-recognised mission in the public interest and
3. in which an enterprise having a determining influence on the institution does not obtain preferential access to the results of scientific research.

(4) Subsections (1) to (3) shall also apply if the reproduction is carried out within the framework of a public-private partnership in which, in addition to the research institution or cultural heritage institution, a profit-making enterprise or other third party is involved.

(5) The free use of works under subsections 1 to 4 may not be waived by contract. However, this shall not prevent the application of measures intended to ensure the security and integrity of the networks and databases in which the works or other subject-matter are stored, provided that such restrictions do not go beyond what is necessary to achieve this objective. Such restrictions shall be deemed reasonable if they have been recognised as good practice by representative associations of rightholders on the one hand and research or cultural heritage institutions on the other.

(6) Any person may reproduce a work for his or her own use for the purpose of automated analysis of texts and data in digital form to obtain information on, inter alia, patterns, trends and correlations, provided that he or she has lawful access to the work. However, this shall not apply if the reproduction is expressly prohibited and this prohibition is appropriately indicated by a reservation of use, for example in the case of works made publicly available via the Internet by machine-readable means. A reproduction under this paragraph may be retained as long as it is necessary for the purposes of data evaluation and information retrieval.

(iii) France

The French Intellectual Property Code was amended in 2018 to introduce TDM copyright exception (Art.122-5(1)(x)), which was altered by the amendment on 24 November 2021 (Arts.122-5(1)(x) and 122-5-3)) to permit TDM activities (‘fouille de textes et de données‘).[96]

Article L122-5(1)(x) (tentative translation by the Author)
Once the work has been disclosed, the author may not prohibit:
[...]
(x) Digital copies or reproductions of a work for the purpose of text and data mining carried out under the conditions set out in Article L. 122-5-3
[...]

Article L122-5-3
I. Text and data mining, within the meaning of Article L. 122-5, 10°, means the implementation of a technique for the automated analysis of texts and data in digital form in order to extract information, in particular patterns, trends and correlations.

II - Digital copies or reproductions of works that have been lawfully accessed may be made without the authors' authorisation with a view to text and data searches carried out solely for the purposes of scientific research by research bodies, libraries accessible to the public, museums, archive services or institutions that are depositories of the cinematographic, audiovisual or sound heritage, or on their behalf and at their request by other persons, including within the framework of a non-profit partnership with private actors.

[96] Ordonnance n° 2021-1518 du 24 novembre 2021 complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE. Available at https://wipolex.wipo.int/en/text/585553.

The provisions of the previous paragraph are not applicable when a company, shareholder or associate of the body or institution carrying out the excavations, has privileged access to their results.
Digital copies and reproductions made during a text and data excavation shall be stored with an appropriate level of security and may be kept for the sole purpose of scientific research, including for the verification of research results.
Copyright owners may implement proportionate and necessary measures to ensure the security and integrity of networks and databases in which works are hosted.
An agreement concluded between the representative organisations of copyright holders and the bodies and institutions mentioned in the first paragraph of this II may define the good practices relating to the implementation of its provisions.

III - Without prejudice to the provisions of II, digital copies or reproductions of lawfully accessed works may be made for the purpose of text and data searches by any person, regardless of the purpose of the search, unless the author has made appropriate objections, in particular by machine-readable processes for content made available to the public online. Copies and reproductions shall be stored with an appropriate level of security and destroyed after the search of texts and data has been completed.

d) Switzerland

The Swiss Copyright Act was amended on September 27, 2019, to introduce a TDM copyright exception (Art.24d [Verwendung von Werken zum Zweck der wissenschaftlichen Forschung]) which came into effect on April 1, 2020.[97]

Article 24d (Use of works for the purposes of scientific research)[98]
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.
3 This article does not apply to the copying of computer programs.

(i) Beneficiaries

There is no particular provision limiting the scope of the beneficiaries of the Swiss TDM copyright exception.

(ii) Purpose

The Swiss TDM copyright exception applies to TDM activities for the purposes of scientific research ('Zum Zweck der wissenschaftlichen Forschung') (Art.24d(1)). However, according to one report and the commentary,[99] the Swiss TDM copyright exception can be applied to not

[97] Bundesgesetz über das Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz, URG), Änderung vom 27. September 2019, AS 2020 1003. Available at https://fedlex.data.admin.ch/eli/oc/2020/181.
[98] Non-official translation of the Swiss Copyright Act is available at https://www.fedlex.admin.ch/eli/cc/1993/1798_1798_1798/en.
[99] *See* the National report of ALAI Switzerland group prepared for the ALAI Congress 2021 in Madrid, p.3 (stating "La 'lecture automatique' doit avoir lieu dans le cadre d'une activité de recherche scientifique (il peut s'agir de recherche appliquée, effectuée par une entreprise privée)." Available at https://server5b96310eea735.vservers.es/IMAGENES/CONVENCION/64/COM_PLANTILLA/BOFER/0/5/5/2021_ALAI_SUISSE_FR.PDF; *See* Rehbinder, M./ Haas, L./ Uhlig, K., URG Kommentar, 4. Aufl. (2022) §24d, Rn.7.

only non-commercial scientific research, but also commercial TDM activities. There seems to be no provision under which right holders can opt-out.

(iii) Lawful access requirement

The Swiss TDM copyright exception applies if 'the works to be copied can be lawfully accessed' (Art.24d(1)).

(iv) Allowed acts

Under the Swiss TDM copyright exception, it is permissible to reproduce a work (Art.24d(1)) and 'the copies made in accordance with this article may be retained for archiving and backup purposes' (Art.24d(2)).

However, works of computer programs are not subject to Swiss TDM copyright exception (Art.24d(3)).

(v) TDM not for 'research purposes'

As mentioned before ((ii)), the Swiss TDM copyright exception applies to TDM activities for the purposes of scientific research ('Zum Zweck der wissenschaftlichen Forschung') (Art.24d(1)), however, according to one report, it can be applied to not only non-commercial scientific research, but also commercial TDM activities. There seems to be no provision under which right holders can opt-out.

(vi) Contractual overridability

There is no provision on the enforceability of a contractual provision contrary to copyright exception in the Swiss Copyright Act.

(vii) Technological overridability

There is no provision for the prohibition of a technological override in the Swiss Copyright Act.

(viii) Safeguard

There is no provision regarding safeguards for the Swiss TDM copyright exception.

e) Japan[100]

The Japanese TDM exception was first introduced in 2009 (Art. 47-7) and was extended (Art. 30-4(ii)) by the amendment in 2018 (effective on 1 January 2019) to basically allow TDM activities as they are aimed at neither enjoying nor causing another person to enjoy the work.

[100] *See* Ueno, T., The Flexible Copyright Exception for 'Non-Enjoyment' Purposes: Recent Amendment in Japan and its Implication, 70(2) *GRUR International* 145-152 (2021).

Article 30-4 (Exploitation without the Purpose of Enjoying the Thoughts or Sentiments Expressed in a Work)[101]
It is permissible to exploit a work, in any way and to the extent considered necessary, in any of the following cases, or in any other case in which it is not a person's purpose to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work; provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation:
(i) if it is done for use in testing to develop or put into practical use technology that is connected with the recording of sounds or visuals of a work or other such exploitation;
(ii) if it is done for use in data analysis (meaning the extraction, comparison, classification, or other statistical analysis of the constituent language, sounds, images, or other elemental data from a large number of works or a large volume of other such data; the same applies in Article 47-5, paragraph (1), item (ii));
(iii) if it is exploited in the course of computer data processing or otherwise exploited in a way that does not involve what is expressed in the work being perceived by the human senses (for works of computer programming, such exploitation excludes the execution of the work on a computer), beyond as set forth in the preceding two items.

(i) Beneficiaries

There is no particular provision limiting the scope of the beneficiaries of the Japanese TDM copyright exception. As a result, the Japanese TDM copyright exception applies to TDM activities not only by research organizations but also by business companies.

(ii) Purpose

There is no particular provision regarding the purpose of TDM activities. As a result, the Japanese TDM copyright exception applies to TDM activities not only for non-commercial purposes but also for commercial purposes.

(iii) Lawful access requirement

There is no lawful access requirement in the Japanese TDM copyright exception. As a result, even if obtaining content might not be regarded as legal private copying, the Japanese TDM copyright exception can be applied to those acts in so far as they are conducted solely for the purpose of TDM.

(iv) Allowed acts

Under the Japanese TDM copyright exception, 'It is permissible to exploit a work, in any way and to the extent considered necessary' (Art. 30-4). As a result, not only the copying of a work but also the distribution and communication to the public can be permitted under the Japanese TDM copyright exception. For instance, someone may copy a large number of books for the purpose of machine learning as TDM and afterwards distribute the training dataset of books to other persons conducting machine learning, in so far as it is solely for the other persons' TDM activities.

[101] Non-official translation of the Japanese Copyright Act is available at https://www.japaneselawtranslation.go.jp/ja/laws/view/4001#je_ch2sc3sb5at4.

It should be also noted that the word 'data analysis' in the Japanese TDM copyright exception is defined as 'the extraction, comparison, classification, or other statistical analysis of the constituent language, sounds, images, or other elemental data from a large number of works or a large volume of other such data' (the parentheses of Art.30-4(ii)). The words "by using a computer" that existed before the 2018 amendment (old Art.47-7) were removed in 2018. Therefore, the current Japanese TDM copyright exception can permit not only computational TDM but also non-computational data mining, in so far as it is aimed neither at enjoying a work nor causing another person to enjoy a work.[102]

(v) TDM not for 'research purposes'

There is no particular provision on the purpose of TDM activities in the Japanese TDM copyright exception. Hence, it applies to TDM activities not only for non-commercial purposes but also for commercial purposes. The right holders cannot opt-out of copyright exception. Therefore, even if a copyright holder expressly makes a reservation (opt-out) for the exploitation of a work for commercial TDM, the Japanese TDM copyright exception applies to the exploitation regardless of the reservation.

(vi) Contractual overridability

There is no provision on the enforceability of a contractual provision contrary to copyright exception in the Japanese Copyright Act. If someone acquires a work based on a contract and copies it for conducting TDM in breach of the contract, the act of copying would not constitute copyright infringement based on the Japanese TDM copyright exception, and it can be, however, regarded as a breach of contract.

A recent study on a contractual provision and Japanese TDM copyright exception[103] pointed out the possibility that a contractual prohibition of conducting AI learning using works for the purpose of research and development of a company can be regarded as unenforceable by the Court taking into account the significance of the development of technology and business.[104]

(vii) Technological overridability

There is no provision for the prohibition of a technological override in the Japanese Copyright Act. If a right holder applies technological measures to his/her works, the right holder has no obligation to make available legal access to the works to the beneficiaries of copyright exception under Japanese law. As a result, right holders might be able to prevent users from conducting TDM activities by using TPM (Technological Protection Measures).

(viii) Safeguard

The Japanese TDM copyright exception has a proviso, which stipulates 'that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation' (the proviso of Art. 30-4).

---

[102] Regarding the Japanese flexible copyright exception for 'non-enjoyment' purposes (Art. 30-4) introduced in 2018, *see* Ueno (2021) *supra* at 145.
[103] 『新たな知財制度上の課題に関する研究会報告書（令和４年２月）』[Study Report on the New Issues of IP (February 2022)] (in Japanese), p.45. Available at https://www.meti.go.jp/policy/economy/chizai/chiteki/pdf/reiwa3_itaku_designbrand.pdf.
[104] *See* ibid.

For instance, the Japanese TDM copyright exception is not applicable to works of database that are made for the use by a person who conducts TDM activities (*see* the proviso of the old Art.47-7[105]). There are ongoing consultations lead by the Japanese Government with scholars and stakeholders regarding especially the interpretation of the proviso of Article 30-4.

f) Singapore

Singapore introduced the TDM copyright exception (Arts. 243 and 244) in 2021.[106]

Division 8 — Computational data analysis

Interpretation: what is computational data analysis
243. *In this Division, "computational data analysis", in relation to a work or a recording of a protected performance, includes —*
(a) using a computer program to identify, extract and analyse information or data from the work or recording; and
(b) using the work or recording as an example of a type of information or data to improve the functioning of a computer program in relation to that type of information or data.

Copying or communicating for computational data analysis
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 computation nal 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.

[105] Article 47-7 of the Japanese Copyright Act before the 2018 amendment has the proviso stipulating 'However, an exception is made of database works which are made for the use by a person who makes an information analysis.'
[106] Singaporean Copyright Act is available at https://sso.agc.gov.sg/Acts-Supp/22-2021/.

(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 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).

(i) Beneficiaries

There is no particular provision limiting the scope of the beneficiaries of this TDM copyright exception in the Singaporean Copyright Act.

(ii) Purpose

There is no particular provision regarding the purpose of TDM activities. As a result, the Singaporean TDM copyright exception can be applied to TDM activities not only for non-commercial purposes but also for commercial purposes.

(iii) Lawful access requirement

Under the Singaporean TDM copyright exception, it is required that the beneficiary 'has lawful access to the material (called in this section the first copy) from which the copy is made' (Art. 244(2)(d)). Additionally, it is also required that one of the enumerated conditions (e.g., the first copy is not an infringing copy (i)) is met (Art. 244(2)(e)).

(iv) Allowed acts

Under the Singaporean TDM copyright exception, 'to make a copy' of a work or a recording of a protected performance is permitted (Art. 244(1)). It is not permitted to supply the copy to any person other than for the purpose of verifying the results of the computational data analysis or collaborative research or study relating to the purpose of the computational data analysis (Art. 244(2)(c)). On the other hand, it is permitted for the beneficiary to communicate a work or a recording of a protected performance to the public if the communication is made using a copy made in circumstances to which Article 244(1) applies and the beneficiary does not supply the copy to any person other than for the purpose of verifying the results of the computational data analysis or collaborative research or study relating to the purpose of the computational data analysis (Art. 244(4)).

(v) TDM not for 'research purposes'

There is no particular provision on the purpose of TDM activities in the Singaporean TDM copyright exception. Hence, it applies to TDM activities not only for non-commercial purposes but also for commercial purposes. The right holders cannot opt-out.

(vi) Contractual overridability

Article 187(1) of the Singaporean Copyright Act stipulates that "[a]ny contract term is void to the extent that it purports, directly or indirectly, to exclude or restrict any permitted use under any provision in" item "(b) Division 7 (computer programs)". Hence, a contractual provision prohibiting TDM activities is unenforceable.

(vii) Technological overridability

There is no provision for the prohibition of a technological override in the Singaporean Copyright Act.

(viii) Safeguard

There seems to be no provision regarding safeguards for the Singaporean TDM copyright exception.

g) USA

There is no explicit specific provision on copyright exception for TDM in the U.S. Copyright Act. Nevertheless, certain TDM activities may be permitted as "fair use" under Article 107 of the U.S. Copyright Act, which permits 'the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research' taking into account four factors ((1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work).

Considering some precedents[107] and that the U.S. courts have held the systematic scanning of print books to be fair use for the purpose of indexation, some commentators have said that certain TDM activities can be regarded as fair use in the U.S.[108]

2. COMPARISON OF TDM COPYRIGHT EXCEPTIONS

There are various detailed differences between the TDM copyright exceptions among jurisdictions. Although it would be difficult to make a complete comparison, the following gives an outline.

---

[107] *Sega Enters. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992); *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015); Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014); *Google LLC v. Oracle Am., Inc.,* 141 S. Ct. 1183.
[108] *See* also Association of Research Libraries, Issue Brief: Text and Data Mining and Fair Use in the United States (June 5, 2015). Available at https://www.arl.org/wp-content/uploads/2015/06/TDM-5JUNE2015.pdf.

(i) Beneficiaries

The beneficiaries of the TDM copyright exception under Art.3 of the CDSM Directive are 'research organizations and cultural heritage institutions', while there is no particular provision limiting the scope of the beneficiaries under Art.4 of the CDSM Directive and other national laws introduced in this report.

(ii) Purpose

TDM for the purpose of non-commercial scientific research is permitted under the DSM Directive and the national laws introduced in this Annex.
TDM for commercial purposes is not permitted in the U.K. On the other hand, it can be permitted in Switzerland, Japan and Singapore. Under Art.4 of the CDSM Directive, it can be permitted, unless the right holders opt out by expressly reserving their right.

(iii) Lawful access requirement

The lawful access requirement is provided for in the U.K., the CDSM Directive, U.K., Switzerland and Singapore, while it is not in Japan.

(iv) Allowed acts

The detailed scope of the allowed acts permitted by the TDM copyright exceptions differs among jurisdictions.

The act of reproduction is permitted under the CDSM Directive and the national laws introduced in this Study. Additionally, the act of exploitation other than reproduction such as communication to the public can be permitted under Japanese TDM copyright exception.
The TDM copyright exception in Switzerland and Germany does not apply to works of computer programs.

(v) TDM not for 'research purposes'

TDM for commercial purposes can be permitted in Switzerland, Japan and Singapore. Under Article 4 of the CDSM Directive, it can be permitted unless right holders opt out by expressly reserving their right.

(vi) Contractual overridability

There is a provision on the unenforceability of a contractual provision prohibiting TDM activities in the U.K. and the CDSM Directive, while there is no such provision in Switzerland, Japan and Singapore.

(vii) Technological overridability

There is a provision for technological overridability in the CDSM Directive, while there is no provision for technological overridability in the U.K., Switzerland, Japan and Singapore.

(viii) Safeguard

In the DSM Directive and Japan, there is a provision on safeguards which denies the application of the TDM copyright exception if it conflicts with a normal exploitation of the work or unreasonably prejudices the legitimate interests of the rightholders.

ANNEX II: CONTRACTUAL PRACTICES

This Annex shows contractual practices developed by rightholders and research institutions by introducing publicly accessible information.

The following are examples of the policy showed by the rightholders (e.g., publishers and academic associations) for the use of the (non-commercial or commercial) TDM of works (e.g., articles in journals, books) published in any way (e.g., open, based on a subscription contract). There are differences in conditions among them.

a) Elsevier[109]

Elsevier (an academic publishing company based in the Netherlands) adopted 'a license-based approach which automatically enables researchers at subscribing institutions to text mine for non-commercial research purposes and to gain access to full text content in XML for this purpose.'

Elsevier Provisions for Text and Data Mining (TDM)[110]
Access to subscription content for text mining is provided to subscribers for noncommercial research purposes. Please note that for open access content, TDM permissions and reuse are determined by the author's choice of user license. Upon acceptance of these provisions for TDM, you will be provided with the API documentation and API key to allow you to do the following:

- Secure a unique API key for your own personal use
- To retrieve, via your API key, content your institution has subscribed to in order to use it as a corpus for noncommercial text mining.
- You can automatically extract semantic entities from the corpus to create TDM output, and prepare research papers or other scholarly publications using the TDM output.
- You are able to distribute the findings of your text mining, in line with the following conditions:
- Where snippets (which may include a few lines of query-dependent text of individual full text articles or book chapters up to a maximum length of 200 characters surrounding and excluding the text entity matched) and/or bibliographic metadata are distributed, they should be accompanied by a DOI link that points back to the individual full text article or book chapter;
  - You should also include a proprietary notice in the following form: "Some rights reserved. This work permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited."
  - You are permitted to distribute or create a link to the list of DOIs used to perform TDM;
  - Where images are used you should clear the rights for reuse with the relevant rightholder;

You are not allowed to:

- Use snippets of text from individual full text articles or book chapters of more than 200 characters (excluding text entity matches or bibliographic metadata);
- Abridge, modify, translate or create any derivative work based on the corpus;
- Delete information about authorship or copyright notices from the corpus;
- Substantially or systematically reproduce, retain or redistribute the corpus;
- Extract, develop or use the corpus in any direct or indirect commercial activity;
- Use any robots, spiders or other automated downloading programs, algorithms or devices to search, screenscrape, extract, or index any Elsevier web site or web application, instead of using the APIs; and

[109] https://www.elsevier.com/about/policies/text-and-data-mining
[110] https://www.elsevier.com/__data/assets/pdf_file/0012/102234/TDM-sign-up-short-form.pdf

- Utilize the TDM output to enhance institutional or subject repositories in a way that would compete with the value of the final peer review journal article, or have the potential to substitute and/or replicate any other existing Elsevier products, services and/or solutions.

You are responsible to keep your contact information as registered on https://dev.elsevier.com up to date and you may not modify or attempt to circumvent the key for secure access to the APIs. We shall have the right to deactivate the API Keys provided to you, if (a) you have not started using the APIs within six months following the delivery of the API Keys, or (b) you have not been accessing the APIs for at least a year since its last access, or (c) the term of the ScienceDirect® database subscription to the book and journal content the organization you are affiliated with expires or if (d) you sell, transfer, sublicense, or otherwise disclose the API key to any other party or you use them for purposes not described herein.

Upon termination or expiration of the subscribed content, Elsevier shall disable access to the APIs for your API Key. You must permanently delete all Elsevier content or Elsevier data which you stored pursuant to your use of the APIs except for the TDM output and the Snippets. Notwithstanding the foregoing, you are permitted to retain a private copy of the corpus, or excerpts thereof, for reasons of data archiving requirements and to make this corpus available for internal institutional uses or for peer review, funding or ethics purposes (but not for further external distribution by these agencies or reviewers). You can also maintain the list of DOIs as a data object and provide this externally. Instead of obtaining an API Key for text mining directly from Elsevier through registration on https://dev.elsevier.com, you can also use the CrossRef TDM service (http://tdmsupport.crossref.org/) to obtain a cross-publisher API token for use with our APIs. Obtaining this token is subject to the same conditions as outlined in these provisions.
Last updated: 24-Feb-2017

b) Springer Nature[111]

Springer Nature 'grants researchers text and data mining rights via their institutions, provided the purpose is non-commercial research' and 'offers standard TDM terms as well as the TDM API for a fee' for TDM in the context of commercial research.

Springer Nature TDM Policy

Springer Nature recognizes the importance of new research techniques and aims to support innovation in this regard. As the volume of scientific publications increases and TDM software tools improve, Springer Nature appreciates the need for a more formalized process to enable TDM, and strives to make this as simple as possible for researchers.

A growing part of Springer Nature's journal articles is published open access. TDM is usually allowed without restrictions for these publications since the majority of Springer Nature open access content is licensed under CC-by.

TDM for researchers at subscribing academic institutions

For subscribed journals and books, Springer Nature grants researchers text and data mining rights via their institutions, provided the purpose is non-commercial research.

[111] Springer Nature Group is a German-British academic publishing company created in 2015, by the merger of Springer Science+Business Media and Holtzbrinck Publishing Group's Nature Publishing Group, Palgrave Macmillan, and Macmillan Education. https://www.springernature.com/jp/researchers/text-and-data-mining.

Individual researchers can download subscription (and open access) journal articles and books for TDM purposes directly from Springer Nature's content platforms. They are requested to limit this to 1 request per second. The selection of desired articles can be conducted by using existing search methods and tools, such as PubMed, Web of Science, or Springer Nature's Metadata API, among others. An API key can be requested for researchers who want to use Springer Nature's TDM APIs. Use of the API provides additional querying parameters and a higher bandwidth for content requests (150 requests per minute).

Researchers are required to use reasonable measures to protect the security of downloaded content, store content on a secure internal server without access for third parties and only for the duration of the TDM project.

Researchers are requested to be considerate and limit downloads to a reasonable rate which does not impose an undue burden on Springer Nature's systems and servers.

Implementation by academic and government institutions

Subscribing academic and government institutions may include text and data mining rights in all new and renewed Journal and ebook subscription agreements under Springer Nature's standard TDM terms (Springer Nature's specialist Database products excluded). For such customers the rights to perform TDM is at no additional cost for content that their subscription license provides access to. Existing subscribers may also add TDM rights under these terms before their agreement is up for renewal.

The use of Springer Nature's TDM API incurs additional costs.

TDM for commercial research (Industry)

For TDM in the context of commercial research, Springer Nature offers standard TDM terms as well as the TDM API for a fee. In that case, the restriction to non-commercial research does not apply.

In addition, Copyright Clearance Center offers a text-mining solution that covers publications from 25 STM publishers, including Springer Nature.

c) Taylor & Francis[112]

Text and Data Mining

Text and Data Mining means to perform extensive automated searches of content on our platform, including text, data, and images. It includes but is not limited to the sorting, parsing, addition or removal of linguistic structures, and the selection and inclusion of content into an index or database for purposes of classification or recognition of relations and associations.

Taylor & Francis recognises TDM is an emerging research practice. If you or your institution subscribes to content from Taylor & Francis you can carry out TDM activities on this content, as well as open access content, without any additional charge, provided this is on a non-commercial basis.

[112]Taylor & Francis Group is a division of Informa plc, a United Kingdom–based publisher and conference company. The Taylor & Francis Group includes Taylor & Francis, Routledge, F1000 Research and Dovepress. See https://taylorandfrancis.com/our-policies/textanddatamining/.

Taylor & Francis requires those carrying out TDM activity on content available on our platform to adhere to the terms and conditions outlined in the STM model licence, or, as applicable, a TDM supplement to our content license agreement.

If you intend to carry out TDM on a commercial basis, you should contact us at the address below to discuss your request and obtain a quote.

If you are planning to carry out TDM activity, we recommend that you contact us to ensure we can provide any access and support you may require. Please email support@tandfonline.com, advising on your organisational affiliation (if any) and with a brief overview of your planned TDM activity.

d) Cambridge University Press[113]

Cambridge University Press permits 'text and data mining of Cambridge Core content for any purpose, as long as you have lawful access to the content you wish to mine.'

Terms of use *(excerpt)*

Machine Analysis (Text and Data Mining)[114]

You may download, extract, store and index Content to which you have lawful access to for the purposes of text and data mining ("TDM") for any purpose and may mount, load, integrate and analyse the results of TDM subject to the inclusion of a link to the underlying Content on our Site. Any copies of the Content stored locally by you for the purposes of TDM shall be deleted once such research project ends.

All Content made available on our Site is provided "as is" and Cambridge provides no warranty as to its suitability for machine analysis. Cambridge does not currently provide access to Content via an API. Cambridge utilises a number of mechanisms to monitor the usage and downloading of Content, and reserves the right to place restrictions, determined at our sole discretion, on users accessing and downloading Content, including appropriate technical protection measures.

If you would like to carry out programmatic or large-scale downloading which may be affected by technical protection measures, or if you need delivery of content in specific formats, please contact openresearch@cambridge.org.

You may use the results of your TDM in your research and make the results of your TDM available on externally facing websites provided no Content, or part of any Content, is made available other than as expressly permitted by applicable law.

For any queries about text and data mining, please contact openresearch@cambridge.org.

e) John Wiley & Sons (Wiley Online Library)[115]

Text and Data Mining Agreement

John Wiley & Sons, Inc. and its affiliates ('Wiley') grants subscribers and other lawful users ('Authorized Users') the right to text and data mine subscribed online content for non-

[113] https://www.cambridge.org/core/services/open-research/text-and-data-mining
[114] https://www.cambridge.org/core/legal-notices/terms
[115] https://onlinelibrary.wiley.com/library-info/resources/text-and-datamining

commercial purposes. In accepting this Text and Data Mining Agreement (the “Agreement”), Authorized Users consent to the terms and conditions set out below.

1. Definitions
‘Text and Data Mining’ (‘TDM’) as used in this Agreement means any automated computational technique for accessing, extracting, copying, or analytical processing of content subscribed to by Authorized Users or otherwise made available to Authorized Users on Wiley Online Library. and TDM shall include but not be limited to the identification of entities, structures and relationships within the Wiley Content.

‘Text and Data Mining Output’ (‘TDM Output’) means the result of any TDM activity carried out by Authorized Users.

‘Wiley Content’ – the electronic content made available (under agreement or otherwise) by Wiley to Authorized Users on Wiley Online Library (or other platforms owned or operated by Wiley).

2. Text and Data Mining
During the term of this Agreement, Wiley grants Authorized Users the non-exclusive, non-transferrable right to text and data mine Wiley Content for the purposes of non-commercial, scholarly research related to specific projects. TDM and TDM Output will not be used for direct or indirect commercial purposes without prior consent in writing from Wiley. Except as permitted by this Agreement or by statutory rights under applicable legislation, Wiley reserves all rights to make reproductions and extractions for TDM. By entering this Agreement, Authorized User acknowledges and agrees to Wiley’s reservation of TDM rights.

Authorized User acknowledges that Wiley Content is protected by copyright and that all right, title and interest in and to Wiley Content remains with Wiley and its licensors and that unauthorized use or redistribution of Wiley Content or the TDM Output would materially harm Wiley and its licensors.

In the event that certain Wiley Content is made available under a more permissive article-level license, such as a Creative Commons CC BY license, then the terms of the article-level license will apply, and users may use that content in accordance with the article-level permissions. In addition, this Agreement is not intended to restrict any statutory TDM rights held by Authorized Users under applicable legislation.

3. Security, Grant of Access Rights, Formats and Delivery Mechanisms
Authorized Users must access Wiley Content for TDM using a Wiley-approved API service and must abide by any rate-limiting conveyed in machine readable form from time to time, and may not bypass the API or disrupt the working of Wiley Online Library. Except in accordance with Wiley’s API requirements (such as load rates and download rates conveyed in machine-readable form from time to time) and the terms and conditions determined by any approved API service, Authorized Users may not use any robots, spiders, crawlers or other automated downloading programs, algorithms or devices to continuously and automatically search, scrape, extract, deep link, or index Wiley Content.

Authorized Users shall implement and maintain adequate and effective data security systems and measures to safeguard Wiley Content downloaded for TDM, in line with international industry standards and best practice.

In the event that TDM results in the processing of personal data, Authorized Users shall be responsible for complying with relevant data protection and privacy laws.

4. TDM Output
Authorized Users may load and technically format Wiley Content on Authorized Users' servers in order to enable access to and use of such Wiley Content for TDM purposes by Authorized Users for specific research projects by using automated programs or devices to continuously and automatically:

- extract and index information such as semantic entities from Wiley Content;
- mount, load and integrate the results for access and use by User for secure TDM;
- communicate TDM Output to third parties as part of original non-commercial research carried out by Authorized Users, including in articles that describe, analyse and interpret research. Publications or analyses resulting from TDM of Wiley Content may include brief quotations from the original text as permitted under Section 107 or 108 of the 1976 United States Copyright Act in the United States, or as permitted by other applicable national copyright laws internationally. Any such extracts, as well as bibliographic metadata, must cite the original Wiley Content in the form of a DOI link. Permission to reproduce images shall be required in accordance with clause 5.

5. Restrictions
Except as expressly stated in this Agreement or otherwise permitted in writing by Wiley, Authorized Users may not:

- perform systematic or substantive extracting for the purposes of creating a product or service for use by third parties, or that has the potential to substitute and/or replicate any other existing Wiley product, service and/or solution;
- create any form of central repository containing Wiley Content except as described above for the purpose of specific TDM projects;
- make the results of any TDM Output available on an externally facing server or website, except as permitted by clause 4;
- allow a third party to harvest any Wiley Content or TDM Output to an internal server; extract, develop or use Wiley Content in any direct or indirect commercial activity; abridge, modify, translate or create any derivative work based on Wiley Content, except to the extent necessary to make it perceptible on a computer screen to the User for research purposes; remove, obscure or modify in any way any copyright notices, other notices or disclaimers as they appear in Wiley Content;
- use any robots, spiders, crawlers or other automated downloading programs, algorithms or devices to continuously and automatically search, scrape, extract, deep link, index or disrupt the working of Wiley Online Library, except as permitted by clause 3; or
- reproduce any illustrations, including photographs, figures and line drawings, in the TDM Output without the consent of the rights holder (unless permitted under the article level license).

6. Fee
Authorized Users shall not be charged an additional fee for TDM, provided the scope remains purely non-commercial. Authorized Users requiring access to non-subscribed Wiley Content should contact their Wiley Account Manager for further information.

7. Term and Termination
The term of this Agreement shall commence on the date on which Authorized User accepts the Agreement and shall continue until terminated by either party. The Agreement may be terminated by either party for breach or for insolvency. Authorized Users will be able to access subscription products for TDM if Authorized Users their institution maintains a subscription to such products. Upon completion of any specific TDM project or upon termination of this Agreement for any reason, Authorized User will delete all Wiley Content downloaded for purposes of TDM.

8. General
Neither party's delay or failure to perform any provision of this Agreement as a result of circumstances beyond its control shall be deemed a breach of this Agreement.

The relationship between the parties does not constitute a partnership, joint venture, or agency relationship, and neither shall have any authority to bind the other in any way.

Authorized User shall not transfer, license or assign any of Authorized User's rights or obligations under this Agreement without the prior written consent of Wiley.

Wiley reserves the right to make changes to the Agreement and its TDM service from time to time and such changes will be clearly posted by Wiley and deemed accepted by Authorized Users.

Choice of law and venue for this Agreement will be the same as that set forth in the agreement under which Wiley Content is subscribed to or made available to Authorized Users

f) Copyright Clearance Center (CCC)

Copyright Clearance Center offers a text-mining solution that covers publications from scientific, technical, and medical (so-called 'STM') publishers including Springer Science+Business Media, Wiley, BMJ, the Royal Society of Chemistry, Taylor & Francis, SAGE, Cambridge University Press, American Diabetes Association, American Society for Nutrition, Future Medicine, with the software 'RightFind XML' that provides a single point of access to full-text journal article content in normalized XML format.[116]

g) PLOS

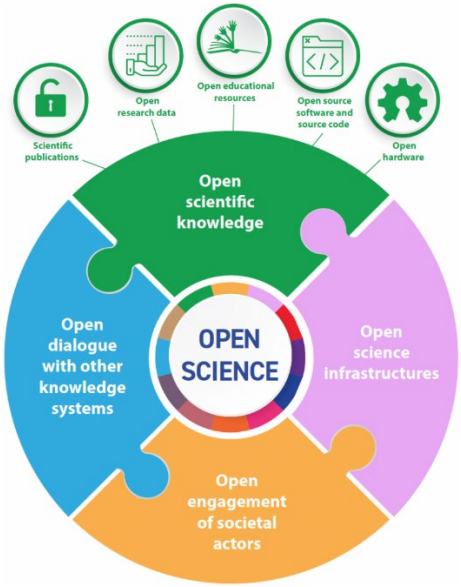
PLOS (The Public Library of Science, a nonprofit and open access publisher) tells, '[o]ur approach to TDM is simple: PLOS articles may be mined, reused, and shared by anyone, anywhere, for any purpose.'[117]

[116] https://www.copyright.com/solutions-rightfind-xml/
[117] https://plos.org/text-and-data-mining/

ANNEX III: OPEN SCIENCE[118]

Open Science can be defined following different approaches. For instance, the UNESCO Recommendation on Open Source[119] defines it very broadly as an "inclusive construct that combines various movements and practices aiming to make multilingual scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community."

Open Science aims at opening up not only research outputs, but also all the practices involved in the research process. The following image provided by UNESCO shows the diverse components of Open Science.[120]



In some instances, Open Science has been defined with reference to the openness of research outputs (Open Access), often protected by copyright and potentially other intellectual property rights. However, the broader concept of Open Science embraces many other aspects, not necessarily related to IP, such as research integrity, and public engagement.

Three key elements of Open Science are particularly relevant for this Study:

- Open Access to research publications, including Open Access journals;
- Open Licensing, such as CC licenses that facilitate the reuse of research outputs (publications and software);
- Open Data offering access to structured data generated and/or used for research purposes.

1. OPEN ACCESS

Open Access (OA) is a set of principles and practices aiming at the online dissemination of research outputs, free of access charges and of other conditions (such as TPMs). Accordingly,

[118] Prepared by Professor Ignasi Labastida.
[119] https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en
[120] See UNESCO Recommendation on Open Science (2021), available at https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en.

OA includes not only free availability but also the right to reuse these contents, broadly with almost no restrictions.

At the end of 2001, a group of researchers, librarians and scientific publishers met in Budapest to discuss about the future of scholarly communication. The Budapest Open Access Initiative of 2002 defined the foundations of the Open Access movement to scientific publications:[121]

> *By "open access" to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.*

Other declarations followed over the years.[122] At the same time, countries have been adopting and updating their Open Access initiatives and policies. The European Union[123] and the Latin American region[124] have been very active in this respect, particularly regarding academic publications. Governments and research institutions established "Open Access" mandates attached to their funding:[125] any output coming from research activities that have been funded with public money will be published in an Open Access format. Similarly, data used in these research projects should be made available on Open Access, to facilitate its checking and use.

The aforementioned Budapest Declaration of 2001 recommended two complementary strategies to achieve Open Access to scholarly journal literature:

- Scholars need the tools and assistance to deposit their refereed journal articles in open electronic archives, a practice commonly called, self-archiving.
- Scholars need the means to launch a new generation of journals committed to open access, and to help existing journals that elect to make the transition to open access.

---

[121] Budapest Open Access Initiative (2002) https://www.budapestopenaccessinitiative.org/read/

[122] See Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2002), available at https://openaccess.mpg.de/Berlin-Declaration; Bethesda Statement on Open Access Publishing (2003), available at https://www.ugr.es/~afporcel/bethesda.pdf; Declaración de Santo Domingo "Ciencia para el siglo XXI: Una nueva version y un marco para la acción" (1999), available at https://rieoei.org/historico/documentos/rie20a12.htm; Salvador Declaration on Open Access (2005), available at https://blog.scielo.org/en/2015/10/23/the-salvador-declaration-on-open-access-completes-10-years/; Declaración de Panamá sobre Ciencia Abierta (2018), available at https://archive.org/details/declaracion_panama_ciencia_abierta; Sorbonne Declaration on Research Data Rights (Jan. 2020), available at https://sorbonnedatadeclaration.eu/.

[123] See, for instance, France: "Second French Plan for Open Access (2021-2024)" https://cache.media.enseignementsup-recherche.gouv.fr/file/science_ouverte/96/6/Second_French_Plan-for-Open-Science_web_1420966.pdf; Germany (2008) "Digitale Information" https://www.allianzinitiative.de/about-us/?lang=en; and (2018) "Shaping digital transformation in science. "Digital Information" Initiative by the Alliance of Science Organizations in Germany. Mission statement 2018 – 2022" https://gfzpublic.gfz-potsdam.de/rest/items/item_2829904_5/component/file_2829905/content; FINLAND: (2014) "Open Science and Research Roadmap" https://julkaisut.valtioneuvosto.fi/handle/10024/75210; USA (2016): "Public Access Plan: Increasing Access to the Results of Federally Funded Scientific Research" https://www.usaid.gov/open/public-access-plan.

[124] See, for instance, Colombia: (2018) "Lineamientos para una política de Ciencia Abierta en Colombia" https://minciencias.gov.co/sites/default/files/ckeditor_files/Lineamientos%20ciencia%20abierta%2017-dic-2018-doc.pdf (2022) "Política Nacional de Ciencia Abierta 2022-2031" https://minciencias.gov.co/sites/default/files/politica_nacional_de_ciencia_abierta_-2022_-_version_aprobada.pdf

[125] See OECD Principles and Guidelines for Access to Research Data from Public Funding (2007), available at https://www.oecd.org/sti/inno/38500813.pdf; see UNESCO Policy guidelines for the development and promotion of open access (2012), available at https://unesdoc.unesco.org/ark:/48223/pf0000215863; see European Recommendation on access and preservation of scientific information, 17 July 2012, COM(2012)4890final, available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012H0417&rid=1.

a) Self-archiving

Self-archiving has become a common practice among researchers, especially because of OA mandates set by research funders. Research institutions and research communities have built digital archives known as repositories to facilitate self-archiving, based on a set of standards to enable interconnection and to improve searches.[126]

The self-archiving strategy has also been accepted by publishers and integrated in some OA models (i.e., Green OA model). Following the transfer or assignment of exclusive rights in favor of the publisher, most publishers allow authors to self-archive a digital copy of their contribution under the following conditions:[127]

- Publishers accept that researchers upload (self-archive) not the final version but the accepted version of their papers, known as author's accepted manuscript or "post-print". This is the reviewed text without the final editorial layout of the journal.
- Public access through self-archiving must be delayed for a period after the publication of the paper; this is known as an "embargo".[128] After this time, public access to the "post-print" version is made available (this the only version allowed by the new copyright holder).
- Repositories must include a clear reference to the publication source and its copyright holder (i.e., the publisher).

Several versions of an academic publication

When disseminating research results through academic publications at least, three different versions of the same work may be identified. The first version is the *original text* that is submitted to a publisher in order to be published in an academic journal or a book. This first version is usually called *preprint or submitted version,* and its relevance has increased enormously in recent years. The number of dedicated servers hosting preprints has grown exponentially because this is the fastest way to disseminate research outputs, even before the work has been reviewed. Moreover, the dissemination of a paper before its publication was a practice not accepted in the past by publishers, but now it is generally accepted and sometimes encouraged by them (i.e., Green OA).

The second version is the *accepted manuscript*, sometimes also called *postprint or authors' manuscript*. This version is the final version of a text once it has gone through all the review process, which is usually conducted by peers and coordinated by a scientific editor. The differences between these two versions (*prepint* and *postprint*) may be sometimes significant (if many modifications were introduced during the review process), or rather insignificant.

And finally, the *published version* is the third one: the version that is published in the journal or the book. This version contains substantially the same text as in the accepted version and therefore it is the same work. However, this third version, also known as *version of record*, also includes the final layout of the publication.[129]

[126] Bashir, A./ Mir, A.A./ Sofi, Z.A., "Global Landscape of Open Access Repositories" (2019). *Library Philosophy and Practice* (e-journal). 2445, available at https://digitalcommons.unl.edu/libphilprac/2445.
[127] Laakso, M. "Green open access policies of scholarly journal publishers: a study of what, when, and where self-archiving is allowed", *Scientometrics* 99, 475–494 (2014). https://doi.org/10.1007/s11192-013-1205-3.
[128] Some publishers establish a default embargo period while others have different embargo times for each journal. Governments or funders play an important role in this regard.
[129] The "Version of Record" (VoR) holds significant value for publishers. It represents the typeset, copyedited, and officially published version of an article. Recognizing that version control is essential for traceability, identifiability, clarity, reduced duplication, and minimized errors, it serves as the definitive version among potentially numerous pre-record versions, ensuring a clear and sequential progression.

Originally, self-archiving might not have been seen as Open Access because it only secured public access but failed to permit reuse of the documents in the open repositories,[130] published under an "all rights reserved" model. In the last years, many institutions have changed their self-archiving policies so that researchers can license their works under the most restrictive *Creative Commons* licenses, such as the CC BY-NC-ND which permits neither the making of neither derivative works nor commercial uses.[131]

b) Open Access journals

The second strategy recommended in Budapest was establishing a new generation of journals that would *"use copyright and other tools to ensure permanent open access to all the articles they publish".* T his new generation of journals is known as Open Access journals.[132] To a large extent, this new generation of Open Access journals has been possible because major STM publishers joined in.[133][134]

Open Access journals do not charge for accessing and reading; for this reason, they must find alternatives to be sustainable. This issue was already pointed out in the Budapest declaration:

> *"Because price is a barrier to access, these new journals will not charge subscription or access fees and will turn to other methods for covering their expenses. There are many alternative sources of funds for this purpose, including the foundations and governments that fund research, the universities and laboratories that employ researchers, endowments set up by discipline or institution, friends of the cause of open access, profits from the sale of add-ons to the basic texts, funds freed up by the demise or cancellation of journals charging traditional subscription or access fees, or even contributions from the researchers themselves. There is no need to favor one of these solutions over the others for all disciplines or nations, and no need to stop looking for other, creative alternatives."*[135]

Among the several existing business models, the most well-known is the one that requires a fee to publish, also known as the "article processing charge" or APC.[136] APC are typically paid by the author, through institutional or grant fundings. It should be mentioned that the "Author pays" business model was not established with the creation of Open Access journals: it existed in the past, and it is still present in some non-Open Access journals (that use the subscription model).

In the Gold OA model, the publisher makes all articles and related content available for free *immediately* on the journal's website.

The Green OA model permits self-archiving by authors. Independently from publication, the author is entitled to post his or her work on an institutional repository (usually, that of the research institution that funded or hosted the work), or to an independent central open

[130] Publishing a paper (i.e., in self-archiving), without any license, does not permit further reuse.
[131] Any further authorization (i.e., authorizing a commercial exploitation) can only be granted by the copyright holder, and this may not always be the author.
[132] See SPARC (March 2019). "Landscape Analysis: The Changing Academic Publishing Industry – Implications for Academic Institutions", p.21-25, available at https://infrastructure.sparcopen.org/landscape-analysis.
[133] See STM Open Access Position at https://www.stm-assoc.org/stm-oa-position/ and STM Open Access Dashboard at https://www.stm-assoc.org/oa-dashboard/
[134] https://www.stm-assoc.org/oa-dashboard/uptake-of-open-access/
[135] https://www.budapestopenaccessinitiative.org/read/
[136] This is the reason why Gold Open Access journals are often identified as the journals "where authors are required to pay for publication."

repository, where people can download the work without paying. Usually, embargoes are set for the Green OA model.[137]

Some Open Access journals do not charge for publishing because they have other income sources: these have recently become known as Diamond or Platinum OA model journals.[138] Diamond or Platinum OA journals often require funding from external sources such as academic institutions, government grants or even advertising. Diamond OA journals are usually small (<25 articles per year) and more likely to be multilingual (38%).[139] The Diamond model has been especially successful in Latin America-based journals (95% of OA journals) following the emergence of large publicly supported platforms, such as SciELO and Redalyc. In 2022, new national and international policies, such as the abovementioned UNESCO recommendation on Open Science, and the Action Plan for Diamond Open Access[140] aim at further supporting the development of non-for profit, community-driven forms of Open Access publishing.[141] Diamond Open Access journals have received a lot of attention in recent years due to the increase of publications fees in the majority of open access journals. [142] Some funders and research institutions are supporting an Action Plan published in 2022 to develop and expand a sustainable, community-driven Diamond OA scholarly communication ecosystem.[143]

Some subscription journals are also offering authors the choice to open their individual contributions through Open Access models. This option is known as the Hybrid OA model: academic journals which offer a mix of OA articles (often under APCs) and closed access articles (under a subscription model). The Hybrid OA model has been used by most of the paywalled scientific journals.[144] This entitles the publisher to be partially funded by subscriptions, and provide open access only for individual articles whose authors (or research sponsor) pay a publication fee (APC).[145,146] Usually, publishers offering this hybrid model provide a choice of open licenses, on top of the transfer of rights to the publisher.

---

[137] The variation of "Green OA/Zero Embargo" refers to the practice of making the accepted manuscript version of a manuscript available in an open access repository immediately upon acceptance for publication, without any embargo period. Its long-term sustainability is yet to be fully determined.

[138] Fuchs, C./ Sandoval, M. (2013-09-09). "The Diamond Model of Open Access Publishing: Why Policy Makers, Scholars, Universities, Libraries, Labour Unions and the Publishing World Need to Take Non-Commercial, Non-Profit Open Access Serious". *TripleC: Communication, Capitalism & Critique*. 11 (2): 428–443. doi:10.31269/triplec.v11i2.502.]

[139] See https://en.wikipedia.org/wiki/Open_access.

[140] https://www.coalition-s.org/action-plan-for-diamond-open-access/

[141] The OA Diamond Study (Fuchs, op. cit. supra) gives an estimation of over 29,000 Diamond Open Access journals in 2021 which represent a significant share of the total number of scholarly journals. Diamond journals make up for 73% of the Open Access journals registered on the Directory of Open Access Journals with 10,194 entries out of 14,020 in September 2020. See https://en.wikipedia.org/wiki/Diamond_open_access.

[142] Morrison, H., Borges, L., Zhao, X., Kakou, T. L., & Shanbhoug, A. N. (2022). "Change and growth in open access journal publishing and charging trends 2011–2021". *Journal of the Association for Information Science and Technology*, 73(12), 1793-1805.

[143] Ancion, Z./ Borrell-Damián, L./ Mounier, P./ Rooryck, J./ Saenen, B. (2022). Action Plan for Diamond Open Access. *Zenodo.* https://doi.org/10.5281/zenodo.6282403]

[144] Hybrid models allow serving both researchers from OA mandates countries (i.e. European countries) as well as researchers from countries without OA mandates (i.e., Brazil). However, the hybrid model poses the challenge that journals may not always have any incentive to flip to a full Open Access model anymore. See Björk B. 2017. *Growth of hybrid open access*, 2009–2016. https://doi.org/10.7717/peerj.3878

[145] A controversial practice in Hybrid OA journals is "double dipping", where both authors and subscribers are charged for the same publication, see for instance https://www.cambridge.org/core/services/open-access-policies/open-access-journals/double-dipping-policy.

[146] A publisher explanation about how the accusation is ill-informed: Double dipping and other bad manners. See https://www.elsevier.com/connect/elsevier-chats-double-dipping-and-other-bad-manners.

Exclusive transfer of rights to the publisher

Initially, Open Access journals did not require any copyright transfer or any exclusive publishing license to publish the work. Authors remained full copyright holders and granted sufficient rights to publishers to publish their articles under the agreed license. However, when browsing the Directory of Open Access Journals (DOAJ),[147] one can find many journals where authors do not retain copyright and are requested to transfer their rights to the publisher or scientific society in charge of the journal. For instance, this is an example of the transfer of rights in a publishing agreement for OA journals, from Elsevier:

"As the author of the Article, I understand that I shall have: (i) the same rights to reuse the Article as those allowed to third party users of the Article under the CC-BY-NC-ND License, as well as (ii) the right to use the Article in a subsequent compilation of my works or to extend the Article to book length form, to include the Article in a thesis or dissertation, or otherwise to use or re-use portions or excerpts in other works, for both commercial and non-commercial purposes. Except for such uses, I understand that the license of publishing rights I have granted to the Journal gives the Journal the exclusive right to make or sub-license commercial use."[148]

This is a relevant issue because many of these journals use restrictive licenses like CC BY-NC-ND and once authors have transferred their exclusive rights to the publisher, authors will not be in a position to authorize further acts of exploitation of their works, for instance, acts of exploitation for commercial purposes not covered by the CC license.

Last, but not least, the subscription model is still very much used by research institutions and academic publishers. In some countries, these agreements go beyond the mere access (that is, reading) and include provisions that will enable researchers to publish openly in paywalled journals, as in the hybrid model. These agreements do not require an individual payment by authors (as in the APC models), but the cost is already included in the overall agreement (with the academic or research institution) to obtain access to these journals. This new generation of agreements are known as "Read and Publish Agreements" or "Transformative Agreements" because they are supposed to transform subscription models to Open Access ones, as mentioned in the Budapest Open Access Initiative. These agreements are more widespread in areas that are typically thought of as having more developed economies, but they are less common in areas with limited economic resources. As we have seen, OA initiatives often require institutions with strong economic resources or heavily funded by governments.

c) OA mandates, sustainability, and copyright

Under OA mandates, researchers are required to publish (when they do so) their research outputs resulting of their funded activities on an Open Access basis. Initially, these OA mandates offered some time flexibilities (i.e., *within 6 months after publication*) which allowed compatibility with embargoes set by OA journal models.[149] These initial requirements have been evolving over time and, currently, researchers are obliged to open up their outputs *immediately after publication* without any restriction (i.e., embargo) and granting the needed rights to reuse them.[150]

[147] https://doaj.org/
[148] See: https://www.elsevier.com/__data/assets/pdf_file/0009/1243188/CC-BY-NC-ND-JPLA_updated_March-2022.pdf.
[149] Sometimes, the delays allowed by OA mandates were shorter than the embargo periods established by OA publishers; this is one of the reasons that hybrid OA models were offered by publishers.
[150] For instance, Wellcome Trust, a UK based charitable foundation, required in 2006 that all research papers funded in whole or in part by the foundation be made available via the UK PubMed Central repository as soon as possible,

[Footnote continued on next page]

Open Access mandates in the EU

Governments and Public funders have established similar requirements when granting funding to research projects. This was done, for instance, by the European Commission which in 2014 included a mandatory Open Access requirement in the Horizon 2020 Research Programme. Currently, the beneficiaries of a project funded under the current European Research Framework, known as Horizon Europe:

must ensure open access to peer-reviewed scientific publications relating to their results. In particular, they must ensure that:
- at the latest at the time of publication, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications
- immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights; for monographs and other long-text formats, the licence may exclude commercial uses and derivative works (e.g. CC BY-NC, CC BY-ND) and
- information is given via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication.[151]

In Latin America, for instance, see the Open Access mandate set in Colombian Law 2294 (2022), Art.171 "Democratization of science through access to results derived from publicly funded research".[152]

Additionally, the APC required in some OA models could be traditionally included as a cost within the project budget. Currently, many governments and public funders have decided not to include the costs of the hybrid publication model in the funded project budgets, and they *require immediate Open Access* for publications resulting from the projects funded by them. In its turn, these new policies clash with the copyright transfers of rights or exclusive licenses that researchers are required to sign with the publishers. Publishers are still imposing embargo periods and continue offering hybrid models as a solution to fulfill the funders' requirements. To overcome this challenge, several solutions have been proposed. Among them, the most relevant ones involve legal amendments introducing a "secondary publishing" right or a "rights retention" strategy.

d) A secondary publishing right

Some national legislators in Europe have introduced in their copyright laws (or in a specific law) a right for authors of a scientific publication allowing to republish the work once it has been published in a journal or a book, under some conditions. This is the case in Belgium, France,

---

and in any event within six months of the date of publication. See Mayor, S., BMJ 2008; 336:688 doi: https://doi.org/10.1136/bmj.39526.467951.DB Instead, the current Open Access policy from Wellcome Trust requires that: *All original, peer-reviewed research articles that are supported in whole, or in part, by Wellcome must be: made freely available through PubMed Central (PMC) and Europe PMC by the official final publication date, and published under a Creative Commons attribution licence (CC BY), unless we have agreed, as an exception, to allow publication under a CC BY-ND licence. See* https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy.

[151] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/agr-contr/unit-mga_he_en.pdf

[152] See Colombian Law 2294 (2023) Art. 171, entitled "Democratization of science through access to results derived from publicly funded research"; available at https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=209510.

Germany, and the Netherlands. The idea of introducing this secondary publication right has been recently endorsed by the Council of the EU.[153]

In Germany, a right to re-publish and communicate to the public in digital form any works resulting from public funding was introduced in Sec.38 already in 2013.[154]

In the Netherlands, a new article was introduced in the copyright law to secure that authors of scientific works funded with public funds will be entitled to disseminate them publicly:

Article 25 fa) – NETHERLANDS
The maker of a short scientific work, the research for which has been paid for in whole or in part by Dutch public funds, shall be entitled to make that work available to the public for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work. [155]

In 2019, an association of Dutch universities launched a campaign called "You share, we take care"[156] to implement the new statutory provision. The campaign's goal was to reach a 100% of Open Access by 2020. The goal was not achieved but the number of publications in repositories has certainly increased.[157] Moreover, a study committed by the Dutch government to evaluate copyright contract law indicated that universities could mandate the use of the Art.25fa).[158]

Challenges faced by this strategy are of a different nature. A first question would be how to define this secondary publishing right. As an inalienable moral right[159] or as an exception to the exclusive assignment of exploitation rights. Second, one should not forget that OA is not only about access, but also about use without restrictions and the secondary right allows authors to republish the work, but remains silent about recovering ownership of rights that may have been transferred to the publisher on an exclusive basis. And lastly, some scholars who used to advocate in favor of expanding this secondary publication right to other jurisdictions are now advocating in favor of eliminating this *embargo* and granting the secondary publishing right immediately after publication.[160]

---

[153] See EU Council, Conclusions on High-Quality, Transparent, Open, Trustworthy and Equitable Scholarly publishing, of 23 May 2023, available at https://data.consilium.europa.eu/doc/document/ST-9616-2023-INIT/en/pdf.
[154] See Caso, Roberto, & Dore, Giulia. (2021). "Academic Copyright, Open Access and the "Moral" Second Publication Right", *European Intellectual Property Review* (2022) n.6, pp.334 https://doi.org/10.5281/zenodo.5764841
[155] This provision is known as the Taverne amendment, after the member of the Dutch parliament who introduced it. See https://zoek.officielebekendmakingen.nl/stb-2015-257.html Some uncertainties may be already identified in the Dutch provision: for instance, what does it mean "reasonable period of time"? What is a "short scientific work"? Or will this provision be applied to international journals or only to Dutch ones? See Visser, D., "The Open Access provision in Dutch copyright contract law", *Journal of Intellectual Property Law & Practice*, Volume 10, Issue 11, November 2015, pp.872–878, https://doi.org/10.1093/jiplp/jpv161
[156] https://www.openaccess.nl/en/in-the-netherlands/you-share-we-take-care
[157] Sondervan, J./ Schalken, A./ de Boer, J./ Woutersen-Windhouwer, S. (2021). "Sharing published short academic works in institutional repositories after six months: The implementation of the article 25fa (Taverne Amendment) in the Dutch Copyright Act" *LIBER Quarterly: The Journal of the Association of European Research Libraries*, 31(1), 1–17. https://doi.org/10.53377/lq.10915
[158] van Gompel, S.J./ Hugenholz, P.B./ Poort, J.P./ Schumacher, L.D./ Visser, D.J.G. (2020). Evaluatie wet auteurscontractenrecht: Eindrapport. University of Amsterdam.
[159] The introduction of the secondary publishing right as a moral right is defended by some legal scholars and it has also been proposed in the Italian copyright law. See Caso, R. and Dore, G., "Academic Copyright, Open Access and the "Moral" Second Publication Right" (December 7, 2021). Available at SSRN: https://ssrn.com/abstract=3981756 or http://dx.doi.org/10.2139/ssrn.3981756
[160] https://libereurope.eu/zeroembargo/

e) Rights retention strategy

This strategy is aimed at retaining rights over the work until a transfer to a publisher is necessary to publish the final version of the work (the published version). The proposal is to only transfer copyright on the final published work, while copyright on any previous versions remains with the authors.[161]

Some public funders are already encouraging this initiative. This is the case of the European Commission and the members of the cOAlition S[162], which in a similar way are advocating for this solution as an alternative when authors are not able to publish directly in an Open Access venue. Their plan (Plan S) seeks to make publicly funded research publications freely accessible to the public, and its supporters are working to promote this open and transparent knowledge dissemination.[163]

Over the last years, a similar approach has been developed by research institutions. The rationale here is based on the policy that researchers grant their institutions a right to archive any scientific publication. This kind of policy was first adopted in 2008 by the Harvard Faculty of Arts and Sciences[164] and now it is being implemented in other countries.

In the United Kingdom, several institutions have approved and are approving similar policies following the initiative of the University of Edinburgh[165] and the pilot at the University of Cambridge[166]. Specifically, the text of the policy approved at Edinburgh[167] works like this:

First, the university acknowledges the common practice of ownership of copyright to scholarly works:
"The University of Edinburgh confirms the current practice that members of staff own the copyright to their scholarly works."

Secondly, the policy states how the university is granted with a right to make manuscripts publicly available:
"Upon acceptance of publication each staff member with a responsibility for research agrees to grant the University of Edinburgh a non-exclusive, irrevocable, worldwide licence to make manuscripts of their scholarly articles publicly available under the terms of a Creative Commons Attribution (CC BY) licence, or a more permissive licence."

And to fulfil with the policy, it is explained how to proceed:
"After granting the licence each staff member with a responsibility for research will provide an electronic copy of the accepted manuscript (AM) of each article at no charge to the appropriate representative of the University of Edinburgh in an appropriate electronic format (such as PDF)."

[161] The first challenge here is to identify all the different versions of a work as separate works and, grant rights differently on each of them.
[162] https://www.coalition-s.org/rights-retention-strategy/ Coalition S is an initiative, launched in 2018, by a group of national research funding organizations with the support of the European Commission and the European Research Council (ERC), to make full and immediate Open Access to research publications a reality.
[163] Coalition S Organizations will facilitate this by changing their grant conditions to require that a Creative Commons Attribution licence (CC BY) is applied to all Author Accepted Manuscripts (AAMs) or Versions of Record (VoR) reporting original research, supported in whole or in part by their funding.
[164] Tate, D., & Aspaas, P. P. (2022). The Rights Retention Policy of Edinburgh University. *Open Science Talk*, (43). https://doi.org/10.7557/19.6859
[165] https://www.ed.ac.uk/information-services/about/policies-and-regulations/research-publications
[166] https://www.openaccess.cam.ac.uk/funder-open-access-policies/rights-retention/rights-retention-pilot
[167] Text policy available at:
https://www.ed.ac.uk/sites/default/files/atoms/files/research_publications_copyright_policy_2021_approved.pdf

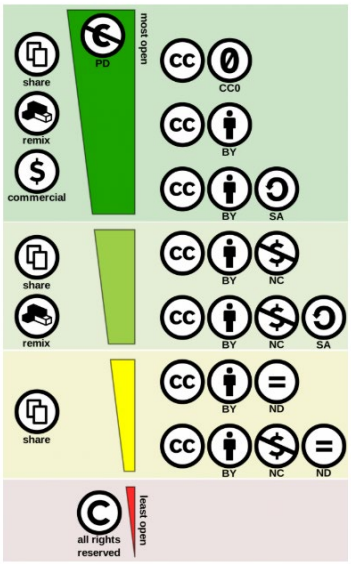
Another approach for institutional policies is to acknowledge research institutions as original rightholders of any scientific scholarly work and then granting enough rights to authors to proceed with the publication.[168] This grant of a right by default should include a provision to keep the right to make publicly available any accepted manuscript at the same time of the actual publication by a journal.

2. OPEN LICENSING

As for Open Access, authors have several options to authorize the public (any readers) the reuse of their publications, subject to the two conditions identified in the Open Access policies: acknowledgment of authorship and integrity of the work. Probably, the best well-known tool to achieve this goal is the set of licenses developed and curated by *Creative Commons.*[169]

Creative Commons licenses

Among all the licenses provided by Creative Commons there is a set of standard licenses, offering some common features and some differences.



"Creative Commons License Spectrum" by Shaddim (CC BY)

Leaving aside the first one, CC0, which appears to be a waiver, rather than a license,[170] the other six licenses allow reproduction, distribution, and public communication of the licensed work, at least, without a commercial purpose. All six require attribution acknowledging not only the author and the copyright holder but any other part required by the licensor. Four licenses, the ones without the NonCommercial (NC) element, allow the aforementioned exploitation acts for commercial purposes. Four licenses, the ones without the NonDerivatives (ND) element, allow making and sharing derivative works. However, two of them, the ones with the ShareAlike (SA) element, have a copyleft requirement: any derivative work must be licensed under the same license (or an equivalent license) as the original work.

[168] European Commission, Directorate-General for Research and Innovation, Angelopoulos, C., Study on EU copyright and related rights and access to and reuse of scientific publications, including open access : exceptions and limitations, rights retention strategies and the secondary publication right, Publications Office of the European Union, 2022, available at https://data.europa.eu/doi/10.2777/891665.
[169] https://creativecommons.org/licenses
[170] Yet, in many countries, the validity of CC0 depends on it being understood as a license (not as a waiver, since copyright cannot be waived by authors): similar to CC-BY but without requiring attribution.

In principle, looking at the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities calling for "*a free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship (community standards*",[171] the CC license that best matches the very definition of Open Access is the Attribution - Creative Commons License, also known as CC BY,[172] which allows any act of exploitation, also for commercial purposes. However, there has been a lot of debate on the use of this license because it is the most permissive. In 2017, a group of Latin American Institutions launched the Mexico Declaration advocating for an ecosystem of non-commercial Open Access and supporting, instead, the use of the CC BY-NC-SA license for Open Access journals,[173] which allows complying with OA mandates, including reuse and the making of derivatives works as long as for non-commercial purposes.

Scientific publications are currently also using proxies, as has been done for many centuries. The most common example of proxy publication would be a journal or a collection managed by a publisher or a scientific association which has established a specific choice of a license to publish with them. Authors who agree to publish in this journal or collection will be implicitly licensing their works with the chosen license.

Free/Libre and Open Source Software (FLOSS) has received a lot of attention, and, in fact, all the current open licenses (such as Creative Commons licenses) were basically created based upon the existing open source software licenses (such as the General Public License or GPL).[174] Software open licensing has been used for over 30 years and it has been tested and validated by courts.[175] The artificial intelligence (AI) is proving to be a challenge for open source. Recently, several lawsuits were brought by Open Source software authors against companies that use open source (open licensed) software to train AI systems and produce (write) new code.

Nevertheless, if the goal is to share software broadly and without restrictions, Open Science goals may be clearly achieved by using free and open source licenses (such as GPL or the EUPL).

FLOSS licenses

FLOSS licenses may generally speaking be divided in two: copyleft and non-copyleft ones.

When the author of a computer program licenses it under an open source license, he or she is authorizing the making of any acts of exploitation that Copyright laws has granted him or her; in order to do so, authors must also make their source code open for use by others.

Under a copyleft license, authors impose on subsequent users/authors a contractual obligation to subject their new (derivative) works to the same copyleft license (this is the "copyleft" component). Non-copyleft licenses do not impose this obligation: the non-copyleft open source license guarantees the freedom to use, modify, and redistribute, while also permitting that derivative works may be subject to copyright by their authors. The original and most well-known copyleft license is the GPL, but a variety of other licenses are commonly used.

[171] https://openaccess.mpg.de/Berlin-Declaration
[172] https://creativecommons.org/licenses/by/4.0/
[173] https://redalyc.org/redalyc/documentos/Declaracion-Mexico.pdf
[174] https://www.gnu.org/licenses/gpl-3.0.html
[175] Reddy, H. R., Jacobsen v. Katzer: "The Federal Circuit Weighs in on the Enforceability of Free and Open Source Software Licenses," *Berkeley Technology Law Journal*, Vol. 24, No. 1, ANNUAL REVIEW OF LAW AND TECHNOLOGY (2009), pp. 299-320.

GNU General Public License (GPL):[176] Richard Stallman originated this license (and the "copyleft" concept) with the goal to use "copyright tools" (that is, a copyright license) to secure that software would not be subject to exclusive licensing terms as permitted by copyright law. GPL is a copyleft license: any software that is written based on any GPL licensed component must be released under the same GPL license and open its full source code.

Other open source licenses include: Apache License, Microsoft Public Licenses (Ms-PL), Berkeley Software Distribution (BSD), Common Development and Distribution License (CDDL), Eclipse Public License (EPL), MIT License.[177] In the EU, the EU Commission developed the EUPL: European Free/Open Source Software License, translated to 22 European languages, and to be used by public administrations within the EU countries (including the EU bodies) to disseminate their software under open source models.[178]

3. OPEN DATA

Beyond scientific publications, Open Science aims at sharing any kind of research outputs, including data.

Research data is another relevant research output.[179] Disseminating results only as an academic publication has proven to be insufficient in order to allow the replication or reproduction of the research process in order to try to achieve the same outcomes. This is the underlying of the scientific methods, and data must be openly accessible to achieve that. That is the reason behind the demand to access data that underlies research publications. Open Data must comply with the FAIR standards: data which meets principles of findability, accessibility, interoperability, and reusability (FAIR).[180]

In general terms, research data cannot qualify as a work protected under copyright[181]. However, since data is generally structured as a compilation or a database, copyright may still have a role to play in its reuse. The arrangement or the presentation of data in a database or compilation may be protected by copyright laws in most countries, if the specific selection or arrangement reaches a minimum level of originality to qualify as a work, specifically, as a compilation of facts or data. Moreover, in some countries, the maker of a database also enjoys a *sui generis* right to control (authorize or prohibit) the extraction and reuse of its content: this is the case of the EU Database *sui generis* right granted in all EU countries.[182]

For Open Science purposes, data should be as open as possible and as closed as necessary. The goal is not to open all data, because often data cannot be made publicly available for reasons such as privacy (and personal data), confidentiality, public security, or commercial secrets. In these cases, closing data is justified.

Several licensing options are available to secure Open Science goals.

---

[176] See GPL https://www.gnu.org/licenses/licenses.html.en.
[177] See OSI – Open Source Initiative, https://opensource.org/licenses/.
[178] https://commission.europa.eu/content/european-union-public-licence_en
[179] European Commission, Directorate-General for Research and Innovation, Senftleben, M., *Study on EU copyright and related rights and access to and reuse of data*, Publications Office of the European Union, 2022, available at https://data.europa.eu/doi/10.2777/78973; See also European Commission, Directorate-General for Research and Innovation, Eechoud, M., *Study on the Open Data Directive, Data Governance and Data Act and their possible impact on research,* Publications Office of the European Union, 2022, available at https://data.europa.eu/doi/10.2777/71619.
[180] https://www.openaire.eu/how-to-make-your-data-fair
[181] See article 5 of the WCT at https://www.wipo.int/wipolex/en/text/295166 and Article 10 of the TRIPS Agreement https://www.wto.org/english/docs_e/legal_e/27-trips_04_e.htm.
[182] Directive 96/9/EC of the European Parliament and of the Council, of 11 March 1996, on the legal protection of databases, available at https://eur-lex.europa.eu/eli/dir/1996/9/oj/eng.

Open licensing of Data

On the one hand, the Open Data Commons initiative created a specific license for data.[183] On the other hand, general public licenses, like the Creative Commons set, have been adapted to be applicable to license the reuse of data and databases by including specific licensing provisions and adjusting the license wording to the specificities of some national legislations.[184] Among them,[185] the CC0 license.[186] This license was designed as a tool for licensing scientific databases, but has ultimately become a tool for dedicating works to the public domain.[187]

The lack of international harmonization for the protection of databases and their contents may lead to different interpretations when extracting and reusing the contents of a database. The use of specific public licenses adapted to these cases could certainly overcome this problem but, at the same time, it may create further challenges when the licensed rights are not protected in certain jurisdictions. Probably this is the reason why some Open Science advocates prefer not to use any public licenses and choose to follow the community (academic) norms when reusing research data, advocating against the protection of databases under copyright and/or *sui generis* rights, and in favor of simple academic norms to favor Open Data initiatives.[188]

[End of document]

[183] Miller, P./ Styles, R./ Heath, T. (2008) Open data commons, a license for open data. LDOW, 369; see https://opendatacommons.org/.
[184] https://wiki.creativecommons.org/wiki/4.0/Sui_generis_database_rights
[185] Hatcher, J., "Implementing Open Data: The Open Data Commons Project." Open Source Business Resource February 2008 (2008). See https://www.timreview.ca/article/122.
[186] https://creativecommons.org/publicdomain/zero/1.0/deed.es_ES
[187] The use of the CC0 license has been severely questioned in some countries, when used to “dedicate” works into the public domain. This is the case in civil law jurisdictions where the protection of copyright is automatically granted from creation and does not depend on a voluntary act such as registration; after all, it is only the term of protection (not the author’s will) that determines when a work will enter the public domain. However, some studies show its feasibility also in these jurisdictions. See Kreutzer, T. (2011) "Validity of the Creative Commons Zero 1.0 Universal Public Do-main Dedication and its usability for bibliographic metadata from the perspective of German Copyright Law."; available at https://www.rd-alliance.org/sites/default/files/cc0-analysis-kreuzer.pdf.
[188] Wilbanks, J., "Public domain, copyright licenses and the freedom to integrate science." *Journal of Science Communication* 7.2 (2008): C04.