

Traditional Medicine Databases

Draft: Traditional Medicine Databases

August 19b, 2026, Send comments to james.love@keionline.org

Introduction	4
Challenges in obtaining authorization to use copyrighted works	4
Challenges in obtaining authorization to use information protected by sui generis regimes	5
The EU Directive on the protection of databases	5
Regulatory data relating to medical or agricultural products	5
Sui generis rights in genetic resources and digital sequence information	6
Sui generis rights in traditional knowledge	6
Consent to use traditional knowledge is not always possible	9
Traditional knowledge available without evidence of consent	9
National or regional text and data mining exceptions	9
Multilateral access and benefit sharing regimes I.	9
Using intellectual property rights to regulate access and use of information	10
Human rights norms	10
Intellectual property	12
The WTO TRIPS Agreement	13
Relevant WIPO negotiations at the IGC	14
Intellectual property rights and access to products or services	18
Recommendations for managers of databases	20
Concerns over practicality of enforcing conditions on the uses of information	21
Triggers for benefit sharing obligations	22
WHO PIP Framework	22
WHO PABS	23
Cali Fund	24
BBNJ Treaty	24
Data Standards	24
FAIR Guiding Principles for scientific data management and stewardship	24
The CARE Principles for Indigenous Data Governance	26
IEEE 2890-2025. Recommended Practice for Provenance of Indigenous Peoples' Data	26
Health Data Spaces	27
The European Health Data Space Regulation	28
Persistent Identifiers of Data	31
Missing or bad metadata	32
Legal scope, jurisdiction and enforcement	32
Blockchain	33
Transparency, and mandates for evidence-based policies	35
Consent, opt-in, opt-out, safeguards, and mandatory sharing	37
Orphaned copyrighted works legislation	39
Opt-in, Opt-out	40

Artificial Intelligence	40
Use of copyrighted works to train GenAI services	40
EU Copyright Directives	41
Japan's Article 30-4 exception for use of copyrighted works to train AI	42
Singapore	42
United States	43
Jurisdictions with an explicit lawful-access requirement built into the statute	44
Training AI on copyrighted works of a scientific, legal or technical nature	45
Using traditional medical knowledge to train AI	46
Remuneration approaches	47
Training data attribution (TDA) challenges	48
Tracking data to weights	48
Tracking weights to outputs	49
Collective management of rights	50
Music Streaming and the evolution of collective licensing	53
Merlin Network	54
Pro-rata, artists-centric and user-centric allocations of music streaming royalties	54
Pro rata distributions or royalties	55
Artist-centric models	55
User-centric models	55
Blur-Banff Model	55
Access and benefit sharing regimes and collective management of rights II.	56
Identifying indigenous peoples or local community beneficiaries	58
Local Communities	63
Local communities in South Africa	64
Definitions of local communities	66
Self-identified	70
Possible approach to benefit sharing that anticipates TMK used to train GenAI services	70
EU, China and California rules for transparency of training data	71
Formalities	73
Brazil's Biodiversity Law (Law No. 13,123/2015)	73
Proposal for downstream benefit sharing obligations	74
Access plans	74
Time-limited obligations for benefit sharing	75
The Bellagio compulsory cross-licensing proposal	75
The Swakopmund Protocol	75
Peru: Law No. 27811 (Collective Knowledge Regime)	76
South Africa and Kenya: Fixed Post-Commercialization Terms	76
The Open Source Dividend Proposal	76
Recommended practice for downstream benefit sharing	78

Introduction

There are a number of databases that focus exclusively or primarily on traditional medicine,¹ and others that touch on it incidentally, capturing genetic resources or traditional knowledge as part of a broader scope. These databases are important for evaluating safety and efficacy, quality control, and drug discovery, and for preserving and disseminating traditional knowledge. Databases are used to aggregate real-world evidence, standardize clinical metrics, provide cross-references for thousands of regional, vernacular names with precise botanical taxonomies, and countless other purposes.

Beyond their scientific or public health uses, databases are also important to holders of intellectual property rights. Databases can be used to identify prior art, blocking patenting of inventions, to track the provenance of traditional knowledge, and to implement access and benefit sharing (ABS) systems.

The elements of databases may include articles or other works protected by copyright, as well as elements that are not protected by copyright. There is information concerning traditional knowledge, which may or may not have been collected after consent was obtained from a party that asserts rights. Non-copyrighted data may be subject to various intellectual property (IP) claims in some jurisdictions, and other types of data may be subject to different legal, regulatory and contractual obligations of confidentiality.

The extent to which copyright or other IP rights need to be cleared for database elements varies by jurisdiction, as do the standards for due diligence. The differences in national legal systems present challenges for managers of databases when access to and the dissemination of the data crosses borders.

Challenges in obtaining authorization to use copyrighted works

It is very challenging to obtain permissions from holders of copyrighted works, particularly for inclusion in large databases. One important database associated with the World Health Organization (WHO) has more than 2 million database elements, including copyright-protected articles.² Among the challenges in obtaining permissions to include articles in a database are those

¹ A 2024 review by Xiao-lan and others provides a comprehensive overview and summary of 182 databases that are relevant to traditional medicine research, including 73 databases for chemical component analysis, 70 for pharmacology research, and 39 for clinical application and quality control from published literature (2000–2023). Li XL, Zhang JQ, Shen XJ, Zhang Y, Guo DA. Overview and limitations of database in global traditional medicines: A narrative review. *Acta Pharmacol Sin.* 2025 Feb;46(2):235-263. doi: 10.1038/s41401-024-01353-1. Epub 2024 Aug 2. PMID: 39095509; PMCID: PMC11747326.

² <https://tmgl.org/databases-and-repositories>. The WHO Traditional Medicine Global Library. (Which one?)

associated with identifying the holders of the copyrights, which are often an entity different from the author or authors, such as an employer, a funder, or a publisher. When authors are rights holders, there may be an obligation to obtain authorizations from multiple authors. There may have been, at one time, contracts that set out the various rights, if any, that authors, employers, funders, or publishers claim in a copyrighted work, and these contracts may no longer be findable or even exist. Some authors may be difficult to locate or contact, and some may have died, with heirs (known or unknown) inheriting rights in their copyrights. The older the works are, the more challenging it will be to identify, contact and negotiate with rights holders. Even when rights holders can be identified, it may be costly and certainly time-consuming to obtain consent.

From an economic perspective, the transaction costs of negotiating permissions on an item-by-item basis often dwarf the value of a single article.

Challenges in obtaining authorization to use information protected by *sui generis* regimes

There are different types of *sui generis* regimes that may be relevant for managers of databases.

The EU Directive on the protection of databases

One type of *sui generis* regime is a legal regime that provides rights to data, including facts, extracted from a protected database, including those bound by a European Union directive: Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. This directive makes it illegal to make available to the public all of a substantial part of the contents of a database, even when data is obtained through a subscription. The remedies for infringing this right are harmonized across the European Union by Directive 2004/48/EC on the enforcement of intellectual property rights (IPRED). The EU has sought to export this norm in a variety of trade agreements, with limited success, and some countries, like Mexico or the United Kingdom (UK), have their own versions of the right. In some cases, the owners of a commercial database may assert these rights to prevent a user from republishing the data. While the EU Directive may be important in some cases, the more general issues for many database users are not the EU Directive, but the terms in contracts that are used to condition access.

Regulatory data relating to medical or agricultural products

A number of states limit access to or use of data submitted to regulatory agencies, including those relating to the regulation of medical or agricultural products. One of the more controversial issues involves the ability to rely upon evidence used to register a product, for registration of generic or

similar products. These restrictions may not create issues for the developer and manager of a database, but they may have consequences for how data distributed from a database is used.

Sui generis rights in genetic resources and digital sequence information

The Convention on Biological Diversity (CBD) provides that national governments have sovereignty over “genetic resources, organisms or parts thereof, populations, or any other biotic component of ecosystems with actual or potential use or value for humanity” collected from *in-situ* sources. The CBD states that “access to genetic resources shall be subject to prior informed consent of the Contracting Party.” Not all CBD parties have a statutory framework for obtaining such consent. There can be ambiguity or controversy over the original source of a genetic resource, particularly those resources that have a translational characteristic. There are decisions that have to be made when a genetic resource has been collected, without consent, and shared or described, including but not limited to digital sequence information.

Sui generis rights in traditional knowledge

Article 8(j) of the 1992 CBD refers to national legislation on promoting the wider application of knowledge, innovations and practices of indigenous and local communities embodying traditional lifestyles, subject to the “approval and involvement of the holders of such knowledge.”

The 2010 Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization provides (in Article 7) that parties shall take measures to ensure that access to traditional knowledge associated with genetic resources is conditioned on “the prior and informed consent or approval and involvement of these indigenous and local communities, and that mutually agreed terms have been established.”

Obtaining such consent involves a determination of which, if any, indigenous or local communities are the appropriate holders of rights in traditional knowledge, and what constitutes informed consent. The modalities for obtaining consent vary considerably, both among indigenous or local communities, and under national laws. For example:

- In the Philippines, the legal requirements for obtaining such consent generally require 17 different steps,³ which can take anywhere from 105 days to more than 6 years to complete.

³ Phase I: Field-Based Investigation (FBI)

1. Application filing: The proponent submits a project/research profile to the NCIP Regional Office.
2. First review: NCIP determines if the area overlaps with a documented ancestral domain.
3. Pre-FBI conference: Meeting between the NCIP, the proponent, and local leaders to set the plan and budget.
4. Posting of notices: Notices of the upcoming field investigation are posted in local barangay (village) halls.
5. On-site investigation: The FBI team conducts a physical site inspection and interviews elders.
6. FBI report preparation: The team drafts a report confirming if indigenous communities are affected.
7. Report submission: The FBI report is forwarded to the NCIP Regional Director for review.

Phase II: The FPIC Consultation & Consensus Building

1. Pre-FPIC conference: Preparing the specific community engagement plan.
2. Formal notice of consultation: Officially notifying the community elders and Council of Leaders.

- The Swakopmund Protocol, developed by the African Regional Intellectual Property Organization (ARIPO), provides that the modalities, custody, and internal decision-making regarding consent must be guided by the customary laws and practices of the specific community. However, the Protocol also provides that a National Competent Authority (NCA) designated by the member state can act "on behalf of and in the interests of the relevant communities," and that the central ARIPO Office will play a role in addressing disputes when traditional knowledge is shared across borders. Moreover, Section 12.1 of the Protocol provides that "Where protected traditional knowledge is not being sufficiently exploited by the rights holder, or where the holder of rights in traditional knowledge refuses to grant licences subject to reasonable commercial terms and conditions, a Contracting State may, in the interests of public security or public health, grant a compulsory licence."
- India's requirements for the consent process involve a three-tier state bureaucracy established by the Biological Diversity Act, 2002, and subsequent legislative overhauls, and the state effectively acts as the legal gatekeeper and trustee of the country's biodiversity and traditional knowledge.
- In China, the Traditional Chinese Medicine Varietal Protection System (TCMVPS) makes the state, rather than individual tribes or communities, the holder of rights in ancient medicine knowledge. China's Traditional Chinese Medicine Varietal Protection System (TCMVPS), established in 1993, is a regulatory exclusivity mechanism, separate from patent law, that lets the state grant protected status to specific formulations, providing market exclusivity for a defined term in exchange for quality and safety assurances. Protection is granted to the manufacturer or applicant, not to originating communities, and there is no consent or benefit-sharing requirement running to indigenous or local groups; the state functions as the effective holder of rights in the underlying knowledge.

Separately, China has built a defensive-documentation apparatus similar in purpose to India's Traditional Knowledge Digital Library (TKDL). The China National Intellectual Property Administration (CNIPA, formerly SIPO) created the Traditional Chinese Medicine Patent Database (TCMPD) in 2002, initially for internal use by patent examiners, containing tens of thousands of TCM-related patent records and formulas drawn from historical and contemporary sources. As with TKDL, the core function is defensive: establishing prior art to prevent foreign patent offices from granting patents on formulations already known in China.

-
3. First Community Assembly: Informing the broader community about the project's scope, impacts, and risks in the local language.
 4. Second Community Assembly: The community debates the proposal privately among themselves using customary decision-making.
 5. The Decision Meeting: The community votes or consensus is reached on whether to accept or reject the project.
 6. Drafting the Memorandum of Agreement (MOA): Outlining terms, royalties, and benefit-sharing.
 7. MOA validation and signing: The community, proponent, and NCIP sign the binding document.
 8. Regional review: The NCIP Regional Office evaluates the entire process for compliance.
 9. Commission en Banc review: The national commissioners in Manila review and approve the agreement.
 10. Issuance of Certificate Precondition (CP): The final legal seal is issued, allowing the project to proceed.

An English-language version was later made available through WIPO to extend this defensive effect internationally, and the concept has since been extended into a bilingual World Traditional Medicine Patent Database (WTMPD), which adds patent family, citation, and legal-status information.

- Peru is a member of the Andean Community, which requires that genetic resources and their "intangible components" (a term for traditional knowledge) are matters of national sovereignty and community heritage, and access requires a formal Access Contract with the state and Prior Informed Consent (PIC) from communities. Peru created three distinct registries managed by INDECOPI (the national IP and competition authority). A Public National Register contains traditional knowledge in the public domain (e.g., written about in historical texts or common knowledge), which it uses as "prior art" to block foreign patent offices from claiming the knowledge is a new invention. There is also a Confidential National Register for sacred, secret, or proprietary community knowledge that can only be accessed with explicit permission from the originating community, and local registries maintained directly by indigenous communities for their own internal governance and preservation. Peru requires that companies that use traditional knowledge to develop a commercialized product must deposit 10 percent of their pre-tax gross sales into the Fund for the Development of Indigenous Peoples (Article 8), and this applies even if the knowledge is already in the public domain, but was commercialized within the last 20 years (Article 13).⁴
- Brazil issued Provisional Measure No. 2,186-16 / 2001, which required researchers to apply for prior authorization before undertaking research on a plant or an indigenous community's traditional knowledge. This system created long delays before companies or academic researchers could conduct research; it also created legal uncertainty, criminalized basic science, and failed to provide significant monetary benefits for indigenous populations. In 2015, Brazil adopted a new law (The Biodiversity Law -- Law No. 13,123 / 2015) that made a number of changes. Brazil created SisGen, an online registry system, whereby researchers

⁴ <https://www.wipo.int/wipolex/en/legislation/details/3420>

Law No. 27811 Introducing the Protection Regime for the Collective Knowledge of Indigenous Peoples derived from Biological Resources, Peru.

Article 8.- Percentage accruing to the Fund for the Development of Indigenous Peoples.

A percentage which shall not be less than ten per cent of the value, before tax, of the gross sales resulting from the marketing of goods developed on the basis of collective knowledge shall be set aside for the Fund of the Development of Indigenous Peoples provided for in Articles 37 et seq. The parties may agree on a greater percentage according to the degree of direct use or incorporation of the said knowledge in the resulting end product and the degree to which the said knowledge contributed to the reduction of the cost of research and development work on derived products, among other things.

Article 13.- Collective knowledge in the public domain

For the purposes of this regime, it shall be understood that collective knowledge is in the public domain when it has been made accessible to persons other than the indigenous peoples by mass communication media such as publication or, when the properties, uses or characteristics of a biological resource are concerned, where it has become extensively known outside the confines of the indigenous peoples and communities. In cases where the collective knowledge has passed into the public domain within the previous 20 years, a percentage of the value, before tax, of the gross sales resulting from the marketing of the goods developed on the basis of that knowledge shall be set aside for the Fund for the Development of Indigenous Peoples provided for in Articles 37 et seq.

declare what they are doing. The government reviews legal compliance later, when a commercial product is being readied for the market or a patent is filed. If a traditional remedy is widely known, published in a historical index, or sold openly in city street markets across the country, it is categorized as being of "unidentifiable origin," and no consent is required for its use. If a company successfully has a finished product based on such unidentifiable knowledge, they are required to pay a royalty (from 0.1 to 1 percent of gross revenue) into a centralized, state-managed National Benefit-Sharing Fund (FNRB).

Consent to use traditional knowledge is not always possible

Some academic experts and indigenous people's representatives do not believe consent is ever possible, since traditional knowledge is an intergenerational asset and a single generation lacks the authority to give consent to future generations.

Traditional knowledge available without evidence of consent

Policy decisions are required when traditional knowledge is already widely available, either orally or published, without any evidence of prior informed consent or mutually agreed terms, and this is addressed in some but not all national legal systems.

National or regional text and data mining exceptions

A number of countries have enacted through statutes or adopted through court decisions limited exceptions to copyright, for the purposes of using text and/or data for research purposes, or to train generative AI (GenAI) services. These national (or in the case of the EU, regional) approaches vary but generally provide for more robust exceptions for non-profit activity and may involve mechanisms for rights holders to opt-out of the exception. Generally, these exceptions are non-remunerative, but there is considerable pressure from some rights holders for some type of remuneration, compensation or equitable benefit sharing.

Multilateral access and benefit sharing regimes I.

The Nagoya Protocol and several subsequent negotiations have sought to bypass requirements for consent from national governments to use genetic resources by creating multiple access and benefit sharing systems. Examples of such negotiations include the FAO International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) (2001), the WHO Pandemic Influenza Preparedness (PIP) Framework (2011), the UN BBNJ Treaty (2026), the WHO Pandemic Agreement and the Pathogen Access and Benefit Sharing System (PABS) (ongoing), and the CBD Cali Fund (ongoing).⁵

⁵ See: Multilateral negotiations on access and benefit sharing obligations for access to genetic resources and traditional knowledge, Knowledge Ecology International (KEI), February 25, 2026.

One objective of these efforts has been to eliminate the need for country-level consent to use genetic resources, in order to reduce transaction costs and delays, and avoid disputes arising from the transborder nature of resources. To the extent the ABS regimes have retained obligations for contracts for access to materials or information, the transaction costs have not been eliminated, but there have been efficiencies. Other objectives include greater cross-border enforcement of benefit sharing obligations. There have been notable disappointments, in some cases relating to a lack of enforcement of obligations, and in other cases, a lack of buy-in by manufacturers, or in regimes that provide such broad exceptions to obligations that few monetary benefits have been forthcoming.

A significant issue that has emerged since the Nagoya Protocol has been the focus on digital sequence information (DSI). Earlier negotiations about biologic resources focused on obligations relating to access to physical samples. The rapidly expanding use of digital sequence information has made physical samples relatively less essential for researchers, and this has accelerated as researchers have greater internet bandwidth and new tools to analyze data. Researchers do not use generic resource digital sequence information in isolation. The use of multiple datasets presents its own challenges when reconciling contractual provisions of an ABS regime with the needs of researchers, who want data to be as open as possible.

This has been a seismic shift in the environment for ABS regimes. No longer are ABS systems able to rely on access to physical samples of genetic resources to impose benefit sharing obligations; now they are seeking to control access to and subsequent uses of information that can be costlessly and instantly shared.

In recent years there has been an interest, in some national and multilateral negotiations, to shift the triggers for benefit sharing away from access to both physical samples and digital information, to downstream events. The most downstream trigger would be the sale of a product that used a genetic resource (or associated traditional knowledge), but there has also been an interest in such events as registering products for sale, filing or obtaining patents, or publishing research. This is an issue that will be discussed in more detail below.

Using intellectual property rights to regulate access and use of information

The public or private regulation of access to or use of knowledge is challenging and problematic, raising issues and concerns about freedom, fundamental human rights, integrity of science, innovation and a plethora of other issues.

Human rights norms

The Universal Declaration of Human Rights, adopted by the United Nations in 1948, describes in Article 27 conflicting rights set out in two sentences. On the one hand, “everyone” has a right to

participate in cultural life, and enjoy and share in the benefits of scientific advancements. But “everyone” also has the “right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author.”

Article 27

1. Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits.
2. Everyone has the right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author.

The 1966 International Covenant on Economic, Social and Cultural Rights, a UN treaty with 173 State parties as of 2026, restated these rights in Article 15.

Article 15

1. The States Parties to the present Covenant recognize the right of everyone:
 - (a) To take part in cultural life;
 - (b) To enjoy the benefits of scientific progress and its applications;
 - (c) To benefit from the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author.
2. The steps to be taken by the States Parties to the present Covenant to achieve the full realization of this right shall include those necessary for the conservation, the development and the diffusion of science and culture.
3. The States Parties to the present Covenant undertake to respect the freedom indispensable for scientific research and creative activity.
4. The States Parties to the present Covenant recognize the benefits to be derived from the encouragement and development of international contacts and co-operation in the scientific and cultural fields.

The UN Committee on Economic and Cultural Rights issued General Comment 17 (2005) on *The right of everyone to benefit from the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he or she is the author (article 15, paragraph 1 (c), of the Covenant)*.⁶ The comment drew a distinction between this human right and state-regulated intellectual property rights:

⁶ GENERAL COMMENT No. 17 (2005). The right of everyone to benefit from the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he or she is the author (article 15, paragraph 1 (c), of the Covenant) <https://digitallibrary.un.org/record/566430?v=pdf>.

. . . Human rights are fundamental, inalienable and universal entitlements belonging to individuals and, under certain circumstances, groups of individuals and communities. Human rights are fundamental as they are inherent to the human person as such, whereas intellectual property rights are first and foremost means by which States seek to provide incentives for inventiveness and creativity, encourage the dissemination of creative and innovative productions, as well as the development of cultural identities, and preserve the integrity of scientific, literary and artistic productions for the benefit of society as a whole. . .

In contrast to human rights, intellectual property rights are generally of a temporary nature, and can be revoked, licensed or assigned to someone else. While under most intellectual property systems, intellectual property rights, often with the exception of moral rights, may be allocated, limited in time and scope, traded, amended and even forfeited, human rights are timeless expressions of fundamental entitlements of the human person. . . . the scope of protection of the moral and material interests of the author provided for by article 15, paragraph 1 (c), does not necessarily coincide with what is referred to as intellectual property rights under national legislation or international agreements. . .

It is therefore important not to equate intellectual property rights with the human rights recognized in article 15, paragraph 1 (c). . .

The UN Declaration on the Rights of Indigenous Peoples (UNDRIP), adopted by the UN in 2007, two years later, stated that Indigenous peoples have the right to maintain, control, protect and develop their intellectual property over cultural heritage, traditional knowledge, and traditional cultural expressions, and that “States shall take effective measures to recognize and protect the exercise of these rights.”

Article 31

1. Indigenous peoples have the right to maintain, control, protect and develop their cultural heritage, traditional knowledge and traditional cultural expressions, as well as the manifestations of their sciences, technologies and cultures, including human and genetic resources, seeds, medicines, knowledge of the properties of fauna and flora, oral traditions, literatures, designs, sports and traditional games and visual and performing arts. They also have the right to maintain, control, protect and develop their intellectual property over such cultural heritage, traditional knowledge, and traditional cultural expressions.

2. In conjunction with indigenous peoples, States shall take effective measures to recognize and protect the exercise of these rights.

Intellectual property

Intellectual property rights is a term that has a nuanced history. Its recent popularity has been the subject of a growing academic and professional literature⁷ which focuses on the use of the word “property” by industry lobbies to promote policies that benefit rights holders.

The WTO TRIPS Agreement

The 1995 WTO Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement) is a set of norms for its members on the extent to which certain types of knowledge will be considered intellectual property, and the rights associated with such resources. The TRIPS Agreement includes sections on copyright and related rights (Articles 9-14), trademarks (Article 15-21, geographic indications⁸ (Article 22-24), designs (Articles 25-26), patents (Article 27-34), plant variety *sui generis* systems (Article 27.3.b), topographies of integrated circuits (Article 35-38), and protection of undisclosed information (Article 39). Each of these categories have their own provisions, which feature boundaries, rights and limitations and exceptions to rights, and some measure of international harmonization and national flexibility in their implementation and enforcement.

It is common in the WTO TRIPS Agreement that intellectual property regimes for the benefit of ownership are resolved through willing buyer/willing seller transactions, but not always. For many, but not all IP types, there are exceptions to exclusive rights, including cases where use is legal without the consent of rights holders, sometimes associated with remuneration or compensation set by third parties, and sometimes without.

The WTO TRIPS Agreement provisions on the enforcement of rights provide some obligations to address national and cross border infringement of rights, but also important space for limitations on remedies, including the possibility to allow ongoing infringement to take place, particularly when compensation or remuneration is available.⁹

While the TRIPS Agreement does not address rights in genetic resources, traditional knowledge or traditional cultural expressions¹⁰, as such, there have been persistent efforts to raise these issues in

⁷ Cite Lemley (2005), Christopher May (2000), Susan Sell, Boldrin and Levine, William Fisher, Richard Stallman, etc.

⁸ Geographic indications are generally considered collective rights, managed and enforced by governments or associations. The exceptions set out in the TRIPS refer to cases where geographic indications are not recognized or enforced, for generic terms, grandfathered prior use, or the rights to use personal names.

⁹ Article 44 on Injunctions.

¹⁰ The 1994 Marrakesh Ministerial Decision on Trade and Environment instructed the WTO Committee on Trade and Environment to consider the work program “envisaged in the Decision on Trade and Services and the Environment and the relevant provisions of the Agreement on Trade-Related Aspects of Intellectual Property Rights as an integral part of its work.” In 1995, the WTO secretariat published WT/CTE/W/8, Environment and TRIPS, (<https://docs.wto.org/dol2fe/Pages/SS/directdoc.aspx?filename=Q:/WT/CTE/W8.pdf&Open=True>) for discussion at the WTO Committee on Trade and Environment (CTE), as a response to the Ministerial Decision to assist the work of the Committee. This 56-page Secretariat document described the original WTO discussions on the relationship between the TRIPS Agreement and the Convention on Biological Diversity.

WTO meetings and negotiations,¹¹ and some attention to how traditional knowledge or traditional cultural expressions are protected or managed under convention IP regimes.

Relevant WIPO negotiations at the IGC

WIPO began a series of negotiations on traditional knowledge in 1998, and in 2001 published Intellectual Property Needs and Expectations of Traditional Knowledge Holders (2001). In 2000, the WIPO General Assembly established the Intergovernmental Committee on Intellectual Property and Genetic Resources, Traditional Knowledge and Folklore, referred to as the IGC. The mandate for this committee has been renewed on a regular basis. The 53rd meeting of the IGC will take place in September 2026.

The WIPO IGC is considering different possible instruments that would create multilateral norms for intellectual property rights in genetic resources, traditional knowledge and traditional cultural expressions. In 2024, a diplomatic conference was convened and adopted the WIPO Treaty on Intellectual Property, Genetic Resources and Associated Traditional Knowledge (the WIPO GRATK Treaty). Despite its name, the treaty addressed a narrow issue: the obligation of patent applicants to disclose the country of origin of the genetic resources and/or the Indigenous Peoples or local community providing associated traditional knowledge (TK), if the claimed inventions are 'based on' genetic resources and/or associated TK. The treaty's mandatory disclosure requirement applies only to traditional knowledge associated with genetic resources.

The IGC is currently considering two possible instruments, one on traditional knowledge and the other on traditional cultural expressions, both prepared by the Secretariat, and debating whether these should be combined into a single instrument.

- WIPO/GRTKF/IC/52/4: The Protection of Traditional Knowledge: Draft Articles
- WIPO/GRTKF/IC/52/5: The Protection of Traditional Cultural Expressions: Draft Articles

Despite more than 25 years of negotiations, both instruments are far from reaching consensus. WIPO/GRTKF/IC/52/4 has 518 brackets and 56 alternatives to specific paragraphs, and WIPO/GRTKF/IC/52/5 has 442 brackets and 35 alternatives.

Among the many controversies are the definitions of beneficiaries (how to define Indigenous peoples, local communities and “other beneficiaries”), the nature of “measures” and/or “rights” associated with protections,” and the status of knowledge publicly available outside of their original community.

The proposal on the protection of traditional knowledge has an Article 5bis on databases. The WIPO/GRTKF/IC/52/4 version of Article 5bis has 82 brackets, meaning almost every operative

¹¹ Cite recent papers by Brazil, etc, and Daniel Gervais book, the TRIPS Agreement, discussion of Doha round.

phrase is still contested and not agreed-upon language. The article sets out different norms and objectives for databases, depending upon their intended use and purpose.

5bis.1 concerns publicly accessible databases of already-public traditional knowledge.

5bis.2 is for databases visible only to intellectual property offices, for defensive prior-art purposes.

5bis.3 is for non-public databases that are restricted to the originating communities themselves, for internal codification and conservation, governed by their own customary law.

5bis.4 through 5bis.10 provide detail on how databases should operate, including objectives, norms, safeguards and other operational features.

5bis.6, is heavily bracketed. It states that when protected traditional knowledge is included in a database, it should only be made available to others with “the free, prior and informed consent or approval and involvement of the traditional knowledge holder.”

WIPO/GRTKF/IC/52/4

ORIGINAL: ENGLISH

DATE: NOVEMBER 5, 2025

Intergovernmental Committee on Intellectual Property and Genetic Resources, Traditional Knowledge and Folklore

Fifty-Second Session

Geneva, March 4 to 13, 2026

THE PROTECTION OF TRADITIONAL KNOWLEDGE: DRAFT ARTICLES

Document prepared by the Secretariat

[ARTICLE 5BIS

[DATABASE], [COMPLEMENTARY] [AND] [DEFENSIVE] PROTECTION

Database Protection

[Recognizing the importance of cooperation and consultation]

Member States should endeavor to cooperate and consult with Indigenous [Peoples] and Local Communities [in determining access] [to traditional knowledge], [Member States should endeavor to], subject to and consistent with national [and customary] law, [facilitate and encourage the development of] [the following] [national traditional knowledge databases to which beneficiaries may voluntarily contribute their traditional knowledge] as follows:

5BIS.1 Publicly accessible national [traditional knowledge] databases publicly available traditional knowledge for the purpose of transparency, certainty, conservation, and transboundary

cooperation, and to facilitate and encourage, as appropriate, the creation, exchange and dissemination of, and access to traditional knowledge.

[5BIS.2 National [traditional knowledge databases of publicly available traditional knowledge accessible only by intellectual property offices for the purpose of prevention of the erroneous grant of intellectual property rights. Intellectual property offices should seek to ensure that such information is maintained in confidence, except where the information is cited during the examination of an application for intellectual property protection.]

5BIS.3 Non-public national traditional knowledge databases for the purpose of [codify] [codifying] and conserving traditional knowledge within Indigenous and local communities. Nonpublic national traditional knowledge databases should only be accessible by beneficiaries in accordance with their respective [customary] laws] and established practices that govern the access or use of such traditional knowledge.

[Complementary][Defensive] Protection

5BIS.4 [Member States]/[Contracting Parties] should [endeavour to], subject to and consistent with national and customary law:

- (a) facilitate/encourage the development of [publicly accessible] national traditional knowledge databases for the defensive protection of traditional knowledge, [including through the prevention of the erroneous grant of patents], and/or for transparency, certainty, conservation purposes and/or transboundary cooperation;
- (b) [facilitate/encourage, as appropriate, the creation, exchange and dissemination of, and access to, [publicly accessible] databases of genetic resources and traditional knowledge associated with genetic resources;]
- (c) [provide opposition measures that will allow third parties to dispute the validity of a patent [by submitting prior art];]
- (d) encourage the development and use of voluntary codes of conduct;
- (e) [discourage information lawfully within the beneficiaries' control from being disclosed, acquired by or used by others without the beneficiaries' [consent], in a manner contrary to fair commercial practices, so long as it is [secret], that reasonable steps have been taken to prevent unauthorized disclosure, and has value;]
- (f) [consider the establishment of [publicly accessible] databases of traditional knowledge that are accessible to patent offices to avoid the erroneous grant of patents compile and maintain such databases in accordance with national law;

(i) there should be minimum standards to harmonize the structure and content of such databases;

(ii) the content of the databases should be:

a. languages that can be understood by patent examiners;

b. written and oral information regarding traditional knowledge;

c. relevant written and oral prior art related to traditional knowledge.]

(g) [develop appropriate and adequate guidelines for the purpose of conducting search and examination of patent applications relating to traditional knowledge by patent offices;]

5BIS.5 [In order to document how and where traditional knowledge is practiced, and to preserve and maintain such knowledge, efforts [should]/[shall] be made by national authorities to codify the oral information related to traditional knowledge and to develop [publicly accessible] databases of traditional knowledge.]] in consultation with Indigenous [Peoples] and Local Communities that hold this information.

5BIS.6 [Member States]/[Contracting Parties] [should]/[shall] consider cooperating in the creation of such databases, especially where traditional knowledge is not uniquely held within the boundaries of a [Member States]/[Contracting Parties]. [If [protected] traditional knowledge [pursuant] as defined in to Section 2 is included in a database, the [protected] traditional knowledge should only be made available to others with the free, prior and informed consent or approval and involvement of the traditional knowledge holder.]

5BIS.6 [Member States]/[Contracting Parties] [should]/[shall] consider cooperating in the creation of such databases, especially where traditional knowledge is not uniquely held within the boundaries of a [Member States]/[Contracting Parties]. [If [protected] traditional knowledge [pursuant] as defined in to Section 2 is included in a database, the [protected] traditional knowledge should only be made available to others with the free, prior and informed consent or approval and involvement of the traditional knowledge holder.]

5BIS.7 Efforts [should]/[shall] also be made to facilitate access to such databases by intellectual property offices, so that the appropriate decision can be made. To facilitate such access, [Member States]/[Contracting Parties] [should]/[shall] consider efficiencies that can be gained from international cooperation. The information made available to intellectual property offices [should]/[shall] only include information that can be used to refuse a grant of [cooperation] patent, and thus [should]/[shall] not include secret traditional knowledge.

5BIS.8 Efforts [should]/[shall] be made by national authorities to considering codify the publicly accessible information related to traditional knowledge for the purpose of enhancing the

development of [publicly accessible] databases of traditional knowledge, so as to preserve and maintain such knowledge.

5BIS.9 Efforts [should]/[shall] also be made to facilitate access to publicly accessible information including information made available in [publicly accessible] databases relating to traditional knowledge by intellectual property offices.

5BIS.10 [Intellectual property offices [should]/[shall] ensure that such information is maintained in confidence, except where the information is cited as prior art during the examination of a patent application.]]

Intellectual property rights and access to products or services

In their strongest form, intellectual property rights can be implemented as a monopoly to make, use, sell, import or export a product or service. The public and policy makers have recognized that such monopolies can have negative impacts on society, and have included in many important multilateral negotiations a variety of norms and normative statements calling for balance or exceptions to address such issues.

To address such concerns, the 1995 TRIPS Agreement included, in Articles 7 and 8, language on the Objectives and Principles of agreement.

Article 7, Objectives

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.

Article 8, Principles

1. Members may, in formulating or amending their laws and regulations, adopt measures necessary to protect public health and nutrition, and to promote the public interest in sectors of vital importance to their socio-economic and technological development, provided that such measures are consistent with the provisions of this Agreement.

2. Appropriate measures, provided that they are consistent with the provisions of this Agreement, may be needed to prevent the abuse of intellectual property rights by right holders or the resort to practices which unreasonably restrain trade or adversely affect the international transfer of technology.

In 1999, the WHO adopted Resolution **WHA52.19 Revised drug strategy**. This Resolution, which was the outcome of an intense and prolonged negotiation, urged member states to take “all necessary concrete measures in order to ensure equitable access to essential drugs,” to “ensure that public health interests are paramount in pharmaceutical and health policies, and “to explore

and review their options under relevant international agreements, including trade agreements, to safeguard access to essential drugs.”

In 2001, faced with growing concern over unequal access to medical inventions, the WTO adopted the Doha Declaration on TRIPS and Public Health, which stated that “the Agreement can and should be interpreted and implemented in a manner supportive of WTO members' right to protect public health and, in particular, to promote access to medicines for all.”¹²

The WHO continued to adopt various resolutions that echoed the mandate to implement intellectual property norms in a manner that would promote access to medicines for all,” such as WHA56.27¹³, WHA60.30¹⁴, WHA61.21¹⁵ and WHA75.14.¹⁶

In 2007, the WIPO General Assembly adopted the WIPO Development Agenda. This included 45 separate recommendations, organized in six clusters.

Cluster A: Technical Assistance and Capacity Building

Cluster B: Norm-setting, flexibilities, public policy and public domain

Cluster C: Technology Transfer, Information and Communication Technologies (ICT) and Access to Knowledge

Cluster D: Assessment, Evaluation and Impact Studies

Cluster E: Institutional Matters including Mandate and Governance

Cluster F: Other Issues

The WIPO development agenda was an effort by developing countries to reset the mandate of WIPO to be consistent with broader UN goals on development and equity, and to make it clear that intellectual property policies should advance the public interest and not only advance the interests of rights holders.

Taken together, developments in the WTO, WHO, WIPO and the broader United Nations system demonstrate a consistent evolution in international norms. While intellectual property rights continue to be recognized as important incentives for innovation, international organizations have increasingly emphasized that such rights are instruments of public policy rather than ends in

¹² WT/MIN(01)/DEC/2, 20 November 2001, Declaration on the TRIPS agreement and public health, Adopted on 14 November 2001 https://www.wto.org/english/thewto_e/minist_e/min01_e/mindecl_trips_e.htm

¹³ WHA56.27, Intellectual property rights, innovation and public health, May 28, 2003.

¹⁴ WHA60.30. Public health, innovation and intellectual property, May 24, 2007. Request the Director General “to provide as appropriate, upon request, in collaboration with other competent international organizations, technical and policy support to countries that intend to make use of the flexibilities contained in the agreement on Trade-Related Aspects of Intellectual Property Rights and other international agreements in order to promote access to pharmaceutical products, and to implement the Doha Ministerial Declaration on the TRIPS Agreement and Public Health and other WTO instruments;”

¹⁵ WHA61.21, Global strategy and plan of action on public health, innovation and intellectual property, “encourage and support the application and management of intellectual property in a manner that maximizes health-related innovation, especially to meet the research and development needs of developing countries, protects public health and promotes access to medicines for all, as well as explore and implement, where appropriate, possible incentive schemes for research and development.”

¹⁶ WHA75.14. Global strategy and plan of action on public health, innovation and intellectual property,”

themselves. Their implementation should promote innovation while also supporting public health, technology transfer, competition, sustainable development and equitable access to knowledge, products and services.

Recommendations for managers of databases

Databases are created for a variety of purposes, and policies that are fit for one may not be appropriate for another. Common (but by no means exhaustive or exclusive) objectives for databases that focus on traditional medicine include:

- Promoting the research and development of commercial products,
- Transferring knowledge and technology,
- Preventing patents from being granted on genetic resources or traditional knowledge,
- Recording and preserving traditional knowledge, for the benefit of indigenous and local communities, and
- Using the contents of the database as an asset that can be used to induce the sharing of benefits.

The contents of databases will include elements that are protected by copyright or a variety of *sui generis* laws, and material that is in the public domain. Databases can be expensive and time consuming to create, update and manage, and access can include a variety of conditions, even when the contents are in the public domain.

In some cases, access to databases is granted without conditions. In other cases, registration and contractual obligations can be imposed.

Intellectual property rights. Among the contractual obligations that can be made to condition access to a database are diverse conditions relating to patents or other intellectual claims. These may include prohibitions or limits on seeking patents, obligations to cross-license patents, or other provisions (see Annex for examples of such clauses).

Restrictions of access and sharing data. There may be restrictions on the use or sharing of the data. For example, for India's TKDL, specified patent examiners can search the database only to evaluate whether a patent application lacks novelty (prior art), and are contractually forbidden from disclosing the contents to third parties or using the data for any other purpose.¹⁷

At present, as per the approval of Cabinet Committee on Economic Affairs, access of TKDL is available to seventeen Patent Offices (European Patent Office, United State Patent & Trademark Office, Japan Patent Office, United Kingdom Patent Office, Canadian Intellectual Property Office, German Patent Office, Intellectual Property Australia, Indian Patent Office, Chile Patent Office, Intellectual Property Corporation of Malaysia, Rospatent- Intellectual Property Office of Russia, Peru Patent Office, Spanish Patent and Trademark Office, Danish Patent, Trademark office, National Industrial Property Institute, France and Eurasian Patent

¹⁷ <https://www.csir.res.in/en/documents/tkdl>

Organization), and Intellectual Property Office of the Philippines under TKDL Access (Non-disclosure) Agreement. As per the terms and conditions of the Access agreement, examiners of patent office can utilize TKDL for search and examination purposes only and cannot reveal the contents of TKDL to any third party unless it is necessary for the purpose of citation. TKDL Access Agreement comprises in-built safeguards on non-disclosure to protect India's interest against any possible misuse.

Linguistic and Conceptual Translation. Some contracts require researchers to translate diagnostic, therapeutic, or pharmacological findings back into the local or indigenous language, or may require that findings be communicated in frameworks that align with local customary concepts, rather than just raw academic prose.

Authorship and Attribution. Contracts require recognition, attribution, consultation, or inclusion of qualifying community researchers on scientific publications, and joint ownership of intellectual property arising from the research. Automatic co-authorship is not simply a negotiable attribution. Under common scientific standards, authorship requires a qualifying intellectual contribution.

Return, deposit, or reciprocal custody of data and specimens. Under physical and digital data stewardship models, researchers are required to return duplicate physical specimens, gene sequences, and raw anthropological records to local archives or national databases (such as community-managed registries).

Fees. The creation, updating and management of databases require resources, and some databases are financed in part through user fees.

Access Guarantees. For traditional medical knowledge that results in a drug or other product, the agreement may require that a community receives priority, subsidized, or free access to the final medical technology or treatment.

Concerns over practicality of enforcing conditions on the uses of information

Conditions governing access are considerably easier to state than to enforce. Once a biological sample leaves its country or community of origin, monitoring subsequent transfers and uses can be difficult. The problem is even greater for digital information, including digital sequence information, because it can be copied, modified, combined with other datasets, and transmitted across borders almost instantly.

Contractual restrictions may be effective against an identifiable contracting party, particularly where the agreement provides for reporting, audits, traceability, remedies, and control over onward transfers. Their effectiveness diminishes when:

- data are placed in an open database;

- recipients pass data to persons who are not parties to the original agreement;
- the same information is available from another source;
- downstream products do not reveal which data contributed to their development;
- users or assets are located in jurisdictions where enforcement is costly or impractical; or
- relevant information is incorporated into aggregated datasets or AI models.

These limitations explain why provenance systems, persistent identifiers, disclosure requirements, user registration, downstream contractual obligations, audit rights, and national or international compliance measures may need to operate together. That said, no single database-access contract can provide complete control once information enters widely distributed digital research networks.

Triggers for benefit sharing obligations

When genetic resources or information are seen as assets that can generate monetary or non-monetary benefits, one approach is to establish ownership in a physical sample of a genetic resource or its digital representation, or knowledge about its use. In the case of physical samples, the 1992 CBD created a legal basis for state sovereignty over its biological resources, while leaving to national governments the responsibility to create legal regimes to regulate access and benefit sharing obligations, if any.

The cross border enforcement of property rights in the physical samples of genetic resources has been challenging. When the genetic resources have a trans-border nature, this is even more complicated and uncertain.

Digital sequence information and TK are generally considered to be unprotected information, unless it is closely held and protected by secrecy.

The 2024 WIPO GRATK Treaty created an obligation for patent applicants to disclose genetic resources and associated traditional knowledge, but, as noted above, does not address what intellectual property rights, if any, would be implicated by such disclosures. The same year, WIPO adopted the Riyadh Design Law Treaty, which allowed but did not mandate the disclosure of traditional cultural expressions in protected designs.

WHO PIP Framework

The WHO PIP Framework requires benefit sharing contracts as a condition for access to influenza samples. Currently, about 85 percent of vaccines for seasonal influenza are sold by companies that have signed benefit sharing contracts with the WHO, while the remaining market is served by companies that have not signed contracts. The companies that obtain samples from the WHO under a Standard Material Transfer Agreement 2 (SMTA2) pay partnership contribution fees and negotiate other benefit sharing obligations, which typically involve an obligation to share 10 percent of production with the WHO, partly as a donation and partly at a concessionary price.

WHO PABS

The negotiations over a WHO Pathogen Access and Benefit Sharing (PABS) system have included, in Article 12, paragraph 6, a requirement that in the event of a pandemic emergency, a manufacturer will make available to WHO 20 percent of real time production.

12.6.(a) each participating manufacturer shall make available to the World Health Organization, pursuant to legally binding contracts signed with the World Health Organization, rapid access targeting 20 percent of their real time production of safe, quality and effective vaccines, therapeutics, and diagnostics for the pathogen causing the pandemic emergency, provided that a minimum threshold of 10 percent of their real time production is made available to the World Health Organization as a donation, and the remaining percentage, with flexibility based on the nature and capacity of each participating manufacturer, is reserved at affordable prices to the World Health Organization; and

For a public health emergency of international concern (PHEIC), or in other cases, negotiations are ongoing, having been extended twice.

The WHO Pandemic Agreement negotiations have largely used the WHO PIP Framework as a model, even though it is recognized that there are considerable differences between pandemics or other health emergencies. The market for seasonable influenza tends to be large, predictable, involve the same manufacturers, and use known technologies. Other pandemics or health emergencies will certainly have different characteristics, and some will have none.

During the negotiations, KEI and some groups representing academic researchers have advocated for the tying of benefit sharing obligations to downstream events, such as registering or selling a commercial product, or filing a patent application, and making the benefit sharing obligation independent of use of or contracts with the PABS. The KEI proposal would require sharing of real time production, for any manufacturer, “regardless of whether that individual manufacturer utilized materials or digital sequence information (DSI) from the PABS System for its development or production.” This obligation would only apply for “products used within the specific field of use and indication required to address the specific pathogen being shared under the Pathogen Access and Benefit-Sharing (PABS) System [and] this obligation shall not extend to other uses or indications of the same products.”¹⁸

Two other multilateral negotiations, the UN Agreement on Marine Biological Diversity of Areas Beyond National Jurisdiction (BBNJ Agreement, adopted in 2023 and in force in 2026) and the CBD Cali Fund, have focused on downstream triggers for benefit sharing, providing more flexibility for sharing data openly with researchers. The WHO PABS, the UN BBNJ treaty and the CBD Cali Fund all propose multilateral mechanisms for the sharing of benefits from the use of both material and digital sequence information.

¹⁸ https://apps.who.int/gb/igwg/pdf_files/IGWG7-written-statements/KEI-07-07-2026.pdf

Cali Fund

The Cali Fund is voluntary and can be described as aspirational at this point, with negotiators discussing the norms for both monetary and non-monetary benefit sharing, and no significant voluntary buy-in from industry at this stage.

BBNJ Treaty

The BBNJ treaty came into force this year, and is a binding instrument. Currently more than 80 countries have ratified the agreement. The BBNJ creates three core trust funds that are funded through mandatory assessed contributions, monetary benefit-sharing from Marine Genetic Resources, and voluntary contributions. The BBNJ treaty designates the Global Environment Facility (GEF) as an institutional trustee and part of its financial mechanism (similar to its role in the Convention on Biological Diversity and UNFCCC). The GEF administers dedicated grants and project funding aligned with BBNJ's objectives.

The triggers for monetary benefit sharing from researchers or companies will include various events. They will likely include such early development events as filing a patent, licensing a synthetic strain, or launching a commercial product, for milestone payments. Monetary contributions are also expected from the actual sale of products. However, before negotiators agree on direct commercial modalities, monetary benefits will come from state contributions.

Data Standards

Standards are important for ensuring that data is accessible and useful, including these three examples.

FAIR Guiding Principles for scientific data management and stewardship

In 2016, a group of scientists published the influential “FAIR Guiding Principles for scientific data management and stewardship.”¹⁹ These are designed to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets, which they summarize as follows:

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

¹⁹ Wilkinson, M.D., Dumontier, M., Jan Aalbersberg, I. et al. Addendum: The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 6, 6 (2019).
<https://doi.org/10.1038/s41597-019-0009-6>

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles

I3. (Meta)data include qualified references to other (meta)data

Reusable

The ultimate goal of FAIR is to optimize the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

The CARE Principles for Indigenous Data Governance

In 2019, the Global Indigenous Data Alliance (GIDA) published the CARE Principles for Indigenous Data Governance (Collective benefit, Authority to control, Responsibility, and Ethics). This effort was led by Dr. Stephanie Russo Carroll (University of Arizona, USA) and Maui Hudson (University of Waikato, Aotearoa New Zealand), with the participation of a global coalition of Indigenous scholars, data activists, and allied researchers.

The CARE principles have been described as a people- and purpose-oriented counterweight or complement to the data-centric, technical FAIR Principles.

C. Collective benefit

Data ecosystems shall be designed and function in ways that enable Indigenous peoples to derive benefit from the data.

A. Authority to control

Indigenous peoples' rights and interests in Indigenous data must be recognised and their authority to control such data be empowered.

R. Responsibility

Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous peoples' self-determination and collective benefit.

E. Ethics

Indigenous peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem.

IEEE 2890-2025. Recommended Practice for Provenance of Indigenous Peoples' Data

The Institute of Electrical and Electronics Engineers (IEEE), the world's largest technical professional organization, has more than 486,000 members in over 160 countries. In September 2025, IEEE approved IEEE 2890-2025, Recommended Practice for Provenance of Indigenous Peoples' Data, which was published in November 2025. The recommended practice was developed through the Standards Committee of the IEEE Society on Social Implications of Technology.

IEEE 2890-2025 helps operationalize the CARE Principles for Indigenous Data Governance -- Collective Benefit, Authority to Control, Responsibility, and Ethics -- by establishing common parameters, descriptors, controlled vocabulary, and metadata fields for recording the provenance of data about or relating to Indigenous Peoples.

The recommended practice addresses several related functions:

Disclosure of relationships. It provides a standardized means of recording and disclosing the relationships between Indigenous Peoples and data concerning their peoples, cultures, knowledge systems, lands, waters, and other relations. Its scope encompasses historical collections as well as contemporary datasets, including environmental and spatial data, genomic and health data, biodiversity information, cultural materials, and data used in artificial-intelligence and machine-learning systems.

Support for Indigenous governance. Provenance information can identify the Indigenous Peoples connected to the data and record information relevant to their governance, decision-making, participation, and potential benefit sharing. This can make Indigenous authority and interests visible to repositories, researchers, companies, and subsequent users of the data.

Interoperability and digital expression of community protocols. The recommended practice establishes common provenance descriptors that can operate across databases and sectors. It also accommodates Indigenous-led mechanisms such as the Local Contexts Traditional Knowledge and Biocultural Labels and Notices, through which communities can communicate culturally specific responsibilities, permissions, restrictions, attribution requirements, and expectations concerning data use.

IEEE 2890-2025 is significant because it converts principles of Indigenous data governance into information that can accompany data through repositories, research systems, and AI or machine-learning pipelines. It does not, by itself, create enforceable Indigenous ownership or control of data, but it can provide infrastructure for recognizing those rights and for monitoring whether institutions respect them.

Health Data Spaces

A "health data space" can be described as a governance and technical architecture for pooling and moving health data across communities, states, or jurisdictions. It is not a physical repository, but rather a set of rules, standards, and infrastructure that let health data flow between different actors (patients, providers, payers, researchers, holders of traditional knowledge, companies, etc.) under common conditions, rather than sitting locked in siloed national or institutional systems.

Federated access is the alternative architecture to centralized pooling; instead of collecting and managing data into a central repository, data can be located in a decentralized system of databases, each with their own systems of policy and control. Queries or algorithms are sent to the data rather than the data being sent to the requester. Local managers are able to implement their own policies regarding security of the data, access and permitted uses.

The federated approach reduces the "single point of failure" problem. When data is centralized, one breach, one policy change, or one bad-faith actor gaining control of the repository can compromise the whole system. A federated approach means no single node holds the whole picture, so trust doesn't have to be extended to one custodian, it's distributed across the origin-holders who retain actual custody, and preserves trusted managers to revoke or condition access at the source.

A health data space approach is appealing for traditional medicine. There is enormous diversity in the policies embraced by national governments and communities that are holders of TK. A health data space can work out issues relating to technical and legal interoperability between federated databases.

The European Health Data Space Regulation

Of interest is the 2025 European Health Data Space Regulation. On March 5, 2025, the European Union published Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847. The European Health Data Space (EHDS) is one of several “Common European Data Spaces” being developed across 14 sectors/domains.

The EHDS regulation is ambitious, setting out obligations for broadly defined data holders and requiring all member states to collaborate on the implementation of standards and systems for expanding access to a wide range of health related-information.

The EHDS regulation is ambitious, setting out obligations for broadly defined data holders and requiring all member states to collaborate on the implementation of standards and systems for expanding access to a wide range of health-related information.

The regulation (1) creates an obligation for entities that hold broadly defined health data to store certain records in standardized electronic formats, (2) gives patients certain rights to control access to those records. The regulation creates (3) technical standards to facilitate the sharing of access to data, and (4) nuanced obligations to share access to data. The regulation (5) defines the permitted and prohibited uses of data, (6) requires EU members to designate entities to implement the regulation, and sets out other mechanisms for the governance and enforcement.

The general approach is to give researchers, businesses and EU governments the ability to use vast amounts of data for both primary (treatment of patients) and secondary uses (research, innovation, policy-making, and regulatory activities), while respecting a number of social objectives. The data spaces approach allows for the data to be held and managed in a decentralized way featuring trusted entities that can anonymize records and/or facilitate queries for users, when necessary to address concerns over privacy or confidentiality. The social objectives include protections of the privacy of patients, improvements in the delivery of health care, and avoiding monopolization.

The EHDS has one set of rules for primary use, and another for secondary uses. These are defined as follows:

‘primary use’ means the processing of electronic health data for the provision of healthcare, in order to assess, maintain or restore the state of health of the natural person to whom

those data relate, including the prescription, dispensation and provision of medicinal products and medical devices, as well as for relevant social, administrative or reimbursement services;

‘secondary use’ means the processing of electronic health data for the purposes set out in Chapter IV of this Regulation, other than the initial purposes for which they were collected or produced;

The Regulation describes the purposes for which secondary uses of data are allowed:

Chapter IV, Article 53, Purposes for which electronic health data can be processed for secondary use

1. Health data access bodies shall only grant access to electronic health data referred to in Article 51 for secondary use to a health data user where the processing of the data by that health data user is necessary for one of the following purposes:

- (a) the public interest in the areas of public or occupational health, such as activities to protect against serious cross-border threats to health, public health surveillance or activities ensuring high levels of quality and safety of healthcare, including patient safety, and of medicinal products or medical devices;
- (b) policymaking and regulatory activities to support public sector bodies or Union institutions, bodies, offices or agencies, including regulatory authorities, in the health or care sector to carry out their tasks defined in their mandates;
- (c) statistics as defined in Article 3, point (1), of Regulation (EC) No 223/2009, such as national, multi-national and Union-level official statistics, related to health or care sectors;
- (d) education or teaching activities in health or care sectors at vocational or higher education level;
- (e) scientific research related to health or care sectors that contributes to public health or health technology assessments, or ensures high levels of quality and safety of healthcare, of medicinal products or of medical devices, with the aim of benefiting end-users, such as patients, health professionals and health administrators, including:
- (f) development and innovation activities for products or services;
- (g) training, testing and evaluation of algorithms, including in medical devices, in vitro diagnostic medical devices, AI systems and digital health applications;
- (h) improvement of the delivery of care, of the optimisation of treatment and of the provision of healthcare, based on the electronic health data of other natural persons.

2. Access to electronic health data for the purposes referred to in paragraph 1, points (a), (b) and (c), shall be reserved for public sector bodies and Union institutions, bodies, offices and agencies exercising the tasks conferred on them by Union or national law, including where

processing of data for carrying out those tasks is done by a third party on behalf of those public sector bodies or of Union institutions, bodies, offices and agencies.

The EHDS Regulations provisions on Intellectual property rights and trade secrets are set out in Article 52. Prohibited secondary uses are set out in Article 54.

Chapter IV, Article 54 Prohibited secondary use

Health data users shall only process electronic health data for secondary use on the basis of and in accordance with the purposes contained in a data permit issued pursuant to Article 68, health data requests approved pursuant to Article 69 or, in situations referred to in Article 67(3), an access approval from the relevant authorised participant in HealthData@EU referred to in Article 75.

In particular, seeking access to and processing electronic health data obtained via a data permit issued pursuant to Article 68 or a health data request approved pursuant to Article 69 for the following uses shall be prohibited:

- (a) taking decisions detrimental to a natural person or a group of natural persons based on their electronic health data; in order to qualify as 'decisions' for the purposes of this point, they have to produce legal, social or economic effects or similarly significantly affect those natural persons;
- (b) taking decisions in relation to a natural person or a group of natural persons in relation to job offers, offering less favourable terms in the provision of goods or services, including exclusion of such persons or groups from the benefit of an insurance or credit contract, the modification of their contributions and insurance premiums or conditions of loans, or taking any other decisions in relation to a natural person or a group of natural persons which result in discriminating against them on the basis of the health data obtained;
- (c) carrying out advertising or marketing activities;
- (d) developing products or services that may harm individuals, public health or society at large, such as illicit drugs, alcoholic beverages, tobacco and nicotine products, weaponry or products or services which are designed or modified in such a way that they create addiction, contravene public order or cause a risk for human health;
- (e) carrying out activities in conflict with ethical provisions laid down in national law.

The preamble of the Regulation describes the intent and purposes of the Regulation in some detail, such as these excerpts from Paragraph 61 of the Regulation.

(61) The secondary use of health data under the EHDS should enable public, private and not-for-profit entities, as well as individual researchers, to have access to health data for research, innovation, policymaking, educational activities, patient safety, regulatory activities or personalised medicine, in line with the purposes as set out in this Regulation. Access to data for

secondary use should contribute to the general interest of society. In particular, the secondary use of health data for research and development purposes should contribute to benefiting society in the form of new medicines, medical devices, and healthcare products and services **at affordable and fair prices for Union citizens**, as well as to enhancing access to and the availability of such products and services in all Member States. . . .

The provision of the data should also support activities related to scientific research. The notion of scientific research purposes should be interpreted in a broad manner, including technological development and demonstration, fundamental research, applied research and privately funded research. Activities related to scientific research include innovation activities such as **training of AI algorithms that could be used in healthcare** or the care of natural persons, as well as the evaluation and further development of existing algorithms and products for such purposes. It is necessary that the EHDS also contribute to fundamental research, and, although its benefits to end-users and patients might be less direct, such fundamental research is crucial for societal benefits in the longer term. In some cases, the information of some natural persons, such as genomic information of natural persons with a certain disease, could contribute to the diagnosis or treatment of other natural persons. . .

While the Regulation states that “the secondary use of health data for research and development purposes *should* contribute to benefiting society in the form of new medicines, medical devices, and healthcare products and services at affordable and fair prices for Union citizens,” the operative articles do not provide further guidance on how the affordable and fair pricing norm will be implemented, if at all.

Persistent Identifiers of Data

When people discuss Persistent Identifiers (PIDs) in the context of Digital Sequence Information (DSI) (such as genomic, transcriptomic, and proteomic data), they are addressing two interconnected challenges: technical infrastructure (how to make vast datasets easily linkable, machine-readable, and permanent) and international policy (how to track access and share benefits arising from genetic resources under agreements like the CBD and the Nagoya Protocol). Examples of PIDs include:

Country of Origin. PID tags can tie sequence accession numbers directly to the physical sample’s country of origin or geographic coordinates.

Internationally Recognized Certificate of Compliance (IRCC). Proposals advocate for embedding Internationally Recognized Certificates of Compliance (IRCCs) or ABS permits directly into the metadata of sequence accessions using Digital Object Identifier (DOI) or another standardized identifier.

Commercial Usage. PIDs can be used to monitor how sequence data moves from public databases (like the International Nucleotide Sequence Database Collaboration,²⁰ or INSDC) into commercial patents or products, allowing tracking and reporting of commercial utilization.

The hope is that embedding IRCC identifiers into digital database accessions will make tracking sample provenance seamless without creating administrative bottlenecks for global scientific research.

While linking IRCCs or other metadata to digital database accessions via PIDs can be an elegant technical solution, its operation at scale, including millions of sequences, introduces significant scientific, technical, legal, and administrative hurdles. Discussed below are some of the things that can and do go wrong.

Missing or bad metadata

Incomplete Historical Data. Public repositories like the INSDC (GenBank, ENA, DDBJ) contain hundreds of millions of sequences deposited over decades, most lacking provenance data or IRCC numbers. Retroactively assigning compliance identifiers to legacy data is impossible.

Non-Compliance. Mandatory IRCC fields are administrative overhead for researchers. If database deposition becomes too cumbersome, scientists may delay sharing data, deposit in non-compliant shadow databases, or upload sequence data without proper IRCC tags, creating gaps in the compliance record.

Taxonomic Ambiguity. Biological samples collected in the field may be misidentified at the point of sequencing or their taxonomy may change over time. If an IRCC specifies a particular species name, but later genetic analysis proves the sample was a different or new species, the legal scope of the permit and the sequence metadata diverge.

Copying, processing and editing errors. In computational biology, sequences are constantly downloaded, reformatted, cropped, alignment-processed, and re-uploaded. Standard file headers can drop complex metadata during routine computational pipelines, severing the link between the sequence and its IRCC.

System failures. The ABS Clearing-House (ABSCH) system run by the CBD and the databases run by the INSDC are entirely separate technical ecosystems. If the ABSCH API goes down, experiences latency, or updates its metadata schema, automated validation during sequence submission breaks.

Legal scope, jurisdiction and enforcement

²⁰ <https://www.insdc.org/>

Commercial vs. Non-Commercial Drift. An IRCC may grant access strictly for non-commercial academic research. However, once sequence data is published in an open database, a commercial entity can download and use it for R&D without ever contacting the original depositor. A simple PID link flags the origin, but it cannot inherently enforce or alter access permissions downstream.

Conflicts in National Legal Norms. National contract, intellectual property and ABS laws vary wildly. Linking an IRCC identifier does not resolve underlying conflicts of law when data generated in Country A is processed on a server in Country B by a scientist in Country C.

Monitoring Capacity. Even if millions of sequences are successfully tagged with IRCC PIDs, national authorities may lack the automated tools and workforce required to parse global sequence databases, patent applications, and commercial products to trace and evaluate compliance with the terms of IRCCs.

These technical and administrative friction points explain why many negotiators and technical experts in forums like the CBD are moving away from rigid, sequence-by-sequence tracking tied to national IRCCs. Rather than burdening scientific databases with complex PID-based tracking, the emerging consensus around multilateral mechanisms (such as the Cali Fund) favors keeping DSI open while capturing value downstream through contributions triggered by commercial events, such as patent applications, product registrations, or revenues.

Blockchain

One technology that has been proposed to facilitate the tracking and governance of digital information relating to genetic resources and traditional knowledge is blockchain, a term that describes a distributed ledger technology in which encrypted records are maintained across multiple computer systems (nodes) and linked together in sequential blocks. Once validated, entries are intended to be resistant to alteration, creating a shared record of transactions.

The advocates of using blockchain technology make the following claims:

Immutable audit trails. A blockchain can provide a decentralized, tamper-resistant ledger. Each time a genetic resource, associated traditional knowledge, or digital sequence information is accessed, transferred, or otherwise used, a transaction could be recorded, creating an auditable history that may assist in documenting provenance, such as the country of origin or provider community.²¹

Automated benefit-sharing through smart contracts. Smart contracts could automate certain compliance and benefit-sharing obligations. For example, if predefined conditions are met, such as the commercialization of a product derived from a particular genetic resource

²¹ Adanur Dedetürk B, Soran A, Bakir-Gungor B. Blockchain for genomics and healthcare: a literature review, current status, classification and open issues. PeerJ. 2021 Sep 30;9:e12130. doi: 10.7717/peerj.12130. PMID: 34703661; PMCID: PMC8487622.

or sequence, the system could automatically trigger payments, royalty distributions, or other agreed-upon benefits to provider countries or Indigenous peoples and local communities.

Deterring unauthorized use. A verifiable record of transactions could make it easier to detect or investigate the use of genetic resources or traditional knowledge without prior informed consent (PIC) or without compliance with mutually agreed terms (MAT), potentially discouraging unauthorized commercial exploitation.

Despite these potential advantages, blockchain has not been widely adopted for the governance of genetic resources or traditional knowledge. Several practical and legal challenges have limited its deployment.

Cost and technical complexity. A new blockchain technology requires sophisticated coding to set up, redundant computing, communications and storage systems to maintain, and other additional costs.

Data errors. Depending upon the objectives for recording and tracking data, the processes for entering data can introduce additional complex tasks for researchers. Blockchain preserves records after they are entered, but it does not verify that the information entered is accurate, complete, or lawful. Data entry may require additional administrative steps and remains vulnerable to both unintentional errors and deliberate misrepresentation.

Difficulties correcting or deleting data. The immutability of blockchain records can conflict with governance models that permit correction, deletion, or withdrawal of data. If information is entered without appropriate consent or legal authority, or if a rights holder subsequently exercises a right to withdraw or remove information, blockchain generally cannot erase or modify the historical record. While later transactions can indicate that data should no longer be used, the original record ordinarily remains part of the ledger.

There is a concern that blockchain technology is a solution looking for a problem, and that while there are conceptual benefits, the practical realities of biomedical research present significant barriers. There are no technical requirements in managing Genetic Resources (GR) or Traditional Knowledge (TK) permissions that require a blockchain.

Some of the resource/cost issues relating to blockchain technology could be managed by an implementation that involves a limited number of nodes, significantly reducing some of the latency and storage costs. By moving from a public, permissionless blockchain, like Bitcoin or Ethereum, to a permissioned or consortium blockchain (like Hyperledger Fabric, Corda, or Enterprise Ethereum), it is possible to eliminate most of the classic resource, latency, and environmental criticisms.

Instead of energy-intensive Proof-of-Work (mining), a consortium can use consensus algorithms like Raft, PBFT (Practical Byzantine Fault Tolerance), or Proof-of-Authority (PoA), and transaction fees become negligible, and thousands of transactions per second are possible. With a limited set of validated nodes, run by a dozen trusted institutions like WIPO, national ABS authorities, indigenous

representative bodies, and research consortia, transactions would finalise almost instantly, and storage costs would be much lower than with a public system.

Once you restrict block validation to a consortium of five to twenty trusted entities, the technology mechanically resembles a federated database with shared cryptographic write access (e.g., PostgreSQL with cryptographic signatures across nodes).

At that point, using a blockchain framework often becomes an expensive way to build a database. Distributed databases like CockroachDB or traditional PKI-signed API networks achieve the same performance, but with lower technical complexity, and easier data updates without the structural overhead of a ledger.

Blockchain in the Nagoya Protocol implementation context has been considered as a tool to solve a lack of trust between parties, but it introduces severe operational friction and costs without solving the core challenges of compliance, enforcement, or verification.

Transparency, and mandates for evidence-based policies

In 2019, the World Health Assembly approved the resolution WHA72.8, titled “Improving the transparency of markets for medicines, vaccines, and other health products.”²² The three-page resolution set out five norms for Member States (“URGES Member States in accordance with their national and regional legal frameworks and contexts”) and seven requests to the WHO Director General. The resolution, which was the subject of intense negotiations, called for interested transparency of the entire value chain, from research to development to commercialization, of health products including medicines, vaccines, medical devices, diagnostics, assistive products, cell- and gene-based therapies, and other health technologies.

The WHO Global Traditional Medicine Strategy, 2025-2034, adopted by the WHA as Decision WHA78/18,²³ “emphasizes the importance of evidence-based traditional medicine, while recognizing the rights of Indigenous Peoples and the importance of the protection of and access to traditional medical knowledge.”²⁴

Evidence is mentioned 42 times in the WHO publication of the strategy, and research is mentioned 52 times.²⁵ The Strategic Objective 1 is “Strengthen the evidence base for TCIM.”

²² https://cdn.who.int/media/docs/default-source/essential-medicines/intellectual-property/gspa/a72_r8-en.pdf. The UK, Germany and Hungary declined to join consensus on the regulation.

²³ Decision WHA78(18). WHO traditional medicine strategy: 2025–2034. In: Seventy-eighth World Health Assembly, Geneva, 27 May 2025 (WHO traditional medicine strategy: 2025–2034).

²⁴ Global traditional medicine strategy 2025–2034. Geneva: World Health Organization; 2025; (Health Systems Performance Assessment (HSPA) toolkit for universal health coverage). ISBN 978-92-4-011317-6 (electronic version)

²⁵ *Ibid.*

Under Guiding Principles of the WHO Global Traditional Medicine Strategy, evidence-based is described as follows:

Guiding principles

4.1 Evidence-based

Scientific or research evidence refers to knowledge that is explicit, systematic and replicable, and can be judged by its methodological standards. Decisions for the use of TCIM should be based on the best available evidence of safety and effectiveness from research and practices. Correspondingly, no health care practices or treatments, be they in biomedicine, public health or TCIM, will be supported or recommended by WHO unless they are evidence-based.

As with all health sciences, traditional medicine should incorporate the highest scientific standards by employing rigorous and robust research methods, integrating continuous methodological and technological advancements, adhering to ethical principles, and using a multidisciplinary approach to select the most stringent and relevant research methods for specific topics and applications to ensure validity, generalizability, reliability and positive impact for people's health and well-being.

Robust scientific validation includes ensuring the quality of the design of research and the quality of the evidence in the way that research is interpreted to ensure safe and effective interventions, and taking into account research ethics, conflicts of interest and issues such as scalability and sustainability for global health, including through health technology assessments. These principles are agnostic to the discipline from which the interventions come and focus on benefits to people's health and well-being based on the most stringent standards of science.

The Strategy seeks to use technological advancements and innovative approaches in research, including the use of new artificial intelligence services.

Direction 1.2

Explore relevant research approaches and optimize the utilization of technological advancements.

Actions for Member States

- Explore innovative, scientifically valid approaches for research appropriate to TCIM.
- Enable the development and application of digital technologies in TCIM research.
- Facilitate digitization and the use of electronic health records inclusive of TCIM related information to enable comprehensive health care in a responsible and ethical manner.
- Develop mobile health solutions and telehealth services and utilize advanced technologies such as those based on artificial intelligence for relevant TCIM.
- Explore relevant research approaches for non-codified traditional medicine.

Actions for partners and stakeholders

- Contribute to developing research methods for the ethical and robust scientific validation of individualized TCIM approaches and knowledge in ways that are culturally appropriate, socially relevant and inclusive.
- Develop digital health applications together with TCIM end-user communities and beneficiaries in support of delivering people-centred care.
- Contribute to developing/implementing electronic patient record systems accessible by TCIM practitioners and promote interoperability.

Actions for the WHO Secretariat

- Support the development of research methodologies appropriate to complex, holistic and individualized approaches of TCIM.
- Strengthen capacity-building on TCIM research methodologies and evidence collection strategies.
- Encourage the development and use of TCIM-specific responsible artificial intelligence tools (17).
- Contribute to the bridging of digital and technological innovations across the TCIM continuum of care, translate collected information into actionable knowledge tailored to Member States, and propose interventions maximizing TCIM contributions.

17. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021 (<https://iris.who.int/handle/10665/341996>). License: CC BY-NC-SA 3.0 IGO.

The WHO Strategy places heavy emphasis on making traditional medicine "evidence-based," defining research evidence as knowledge that is explicit, systematic, replicable, and methodologically stringent. The document asks Member States and stakeholders to enable digitization, electronic health records, AI models, and mobile health tools for Traditional, Complementary, and Integrative Medicine (TCIM).

The promotion of digitizing, sharing, and standardizing data to build a global scientific evidence base, when doing so with traditional medical knowledge, presents tensions with norms regarding the unauthorized commercialization or exploitation of protected traditional knowledge.

Consent, opt-in, opt-out, safeguards, and mandatory sharing

A national legal regime may require consent for uses of protected copyrighted works, genetic resources, or traditional knowledge, and this may be further specified for traditional knowledge as prior informed consent (PIC), or free and prior informed consent (FPIC).

As a practical matter, for large databases, clearing documents for consent, whether it is under copyright, genetic resources or traditional knowledge, will be challenging. It may require identifying, locating and establishing consent from holders of copyright (often not the authors). Likewise, when it is unclear who holders of traditional knowledge are, particularly if such knowledge is widely

diffused and shared, or the genetic resources have a trans border nature, proof of consent can be impractical or impossible.

Copyright & Ownership Chains: Finding the original author or authors is rarely the end of the search. Copyrights are frequently assigned, transferred, or licensed to corporate entities, publishers, or estates. Locating the actual current rights holder to secure explicit licensing across large numbers of dataset entries becomes an unrealistic administrative burden.

Diffuse & Transboundary Traditional Knowledge: Traditional knowledge rarely adheres to modern nation-state borders or singular legal entities. If a particular botanical remedy or cultural motif is shared among multiple communities across different borders, establishing who holds the authority to grant PIC or FPIC can itself be a challenge, and when that knowledge is already widely diffused, the rationale for excluding the information from a database is weak.

Genetic Resources & Transboundary Species: Under frameworks like the Nagoya Protocol, accessing genetic resources requires agreement with the providing country and source communities. However, biological organisms and their DSI do not respect geopolitical borders, creating multi-jurisdictional compliance challenges.

Databases can be managed in a variety of ways. In the EHDS Regulation, patient records, when anonymized for secondary use, are normally allowed to be used, subject to a patient exercising an opt-out right. However, for some records that are considered more sensitive, EU member states can require that records not be shared unless patients exercise a formal opt-in, and for some information, no opt-out is possible.

The opt-out approach is that the default is inclusion, and data is included automatically, and holders of rights must actively request that data is withdrawn. A strict consent approach, opt-in, would have data excluded by default, without evidence of explicit prior authorization.

In the EHDS Regulation, individual EU citizens have an opt-out right for secondary use of their personal patient records but copyright holders and corporate and institutional health data holders do not have an automatic opt-out right, simply because they contain copyrighted elements, trade secrets, or proprietary data, and data holders may be required to make data available for secondary use, subject to some safeguards. As described by Baker McKenzie in a briefing note:²⁶

The EHDS requires health data holders to make in scope electronic health data available, through Health Data Access Bodies, to authorised health data users for a closed list of permitted secondary uses, including scientific research, innovation and the training, testing and validation of algorithms.

From a trade secrets perspective, this is unprecedented.

26

<https://www.bakermckenzie.com/en/insight/publications/2026/06/european-union-european-health-data-space-and-trade-secrets>

Unlike voluntary data sharing frameworks, EHDS access: (i) may be requested by third parties with competing commercial interests; (ii) is channelled through health data access bodies rather than negotiated bilaterally; and (iii) limits the data holder's ability to refuse access outright.

Although the Regulation recognises the need to protect intellectual property and trade secrets, it does so primarily through protective conditions, not through a general right to withhold data. This shifts the legal battleground from "whether data must be shared" to "under what safeguards, and at what level of abstraction".

Orphaned copyrighted works legislation

In copyright law, the term "orphan work" is used to describe a copyrighted work, when the owner of the copyright cannot be identified or located. When an owner cannot be reached to grant permission, prospective users face a risk of copyright infringement liability. This can and frequently does lead to cultural, historical, and academic material being effectively under utilized.

A work generally qualifies as an orphan work if it meets three criteria:

- The work is not in the public domain; its copyright term has not expired.
- The identity of the rightsholder is unknown. Even if an author or original publisher is known, the terms under which rights have been transferred through a contract may be unknown, and/or their location or legal successor cannot be traced (common with deceased creators, defunct publishers, or transferred corporate assets). Many living authors have no records of agreement signed with publishers regarding who holds the copyright in the works they produced, and it may also be impossible to obtain the contracts from publishers, if they still exist.
- A prospective user has conducted a good-faith, documented investigation across relevant registries, databases, and archives to find the owner, but failed to do so.

The European Union, in Directive 2012/28/EU, Canada in Section 77 of its Copyright Act, and UK Intellectual Property Office (UKIPO), in the Enterprise and Regulatory Reform Act 2013, all define orphan works and set out rules regarding their use.

The United States has struggled to enact legislation on orphaned copyright works, in part over concerns expressed by photographers.²⁷ A proposal that nearly succeeded in the U.S. Congress would have limited the liabilities for using orphaned work, and permitted both non-commercial and

²⁷ Unlike book publishers, photographers often have uncredited works circulating online or in print ephemera without embedded metadata.

commercial uses, and in some cases, even after ownership rights had been established, subject to remuneration.²⁸

The limitations and exceptions to copyright in various legislative approaches were either based upon a three-step test requiring the exceptions be limited to “special cases which do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the right holder” (the TRIPS Standard in Article 13), or, in the case of the U.S. proposal, a limitation on remedies to infringement, relying on the last sentence of Article 44.2 of TRIPS. The second sentence of Article 44.2 of the TRIPS allows a WTO member to deny injunctions and limit financial recovery to adequate compensation, preserving formal liability while eliminating destructive exposure to damages that would otherwise be at risk.

Opt-in, Opt-out

For databases involving traditional medicine, where copyright, genetic resources and traditional knowledge are all relevant, a recommended approach is to identify the types of records that should have an opt-out, opt-in or mandatory inclusion status.

For traditional knowledge, information that is widely diffused should generally have an opt-out or mandatory inclusion status, and information that is sacred or closely held should have the more strict opt-in status.

For copyright, the national laws will differ on whether or not consent or remuneration is required. A challenge for the managers of a database is to conform to national laws where the database is located and operated, but in some cases, to also address legal regimes where data is accessed.

Artificial Intelligence

The use of data to train artificial intelligence (AI) is identified by the WHO as important in its Global Strategy on Traditional Medicine, is a permitted secondary use of data by the European Health Data Spaces (EHDS) regulation, and is the subject of considerable and evolving state legal norms.

Use of copyrighted works to train GenAI services

The most significant legislative activity has involved copyright, and whether or not consent is required before works protected by copyright can be used to train generative AI (GenAI) services, and what terms, such as transparency, tracking or remuneration, are required.

²⁸ Shawn Bentley Orphan Works Act of 2008.

Extracting facts, statistical information, or patterns from copyrighted works generally does not implicate copyright because copyright protects expression, not facts or ideas.

Making and storing copies of works to build a training corpus does implicate the reproduction right and therefore may require authorization or a copyright exception.

Several countries have enacted statutes to address the uses of text and data mining for research purposes, and there are multiple court cases over copyright infringement claims.

Among the countries that have enacted special provisions on the uses of copyrighted works for research purposes (often referred to as text and data mining, or TDM exceptions), some have provided more robust exceptions for noncommercial entities or uses than what are available for profit or commercial uses. These distinctions are being reconsidered in some countries, in order to provide a more welcoming legal environment for the development of commercial GenAI services. On the other hand, in some countries, like the UK, lobbied by copyright holders and creative communities has caused governments to narrow exceptions.

The balance between enabling and supporting commercial AI investment and protecting copyright holders is intensely contested, and state practice is diverse.

EU Copyright Directives

One important legal norm in Europe is Directive 2019/790 on copyright and related rights in the Digital Single Market.²⁹ Article 3, titled “Text and data mining for the purposes of scientific research,” provides an unwaivable TDM exception for “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.”

Article 4 of EU Directive 2019/790 provides for a more general mandatory exception or limitation to rights for TDM, available to commercial entities. But the exception is narrower, and does not apply if the copyright holder has expressly opted out of the exception, by placing a machine-readable tag on a copyrighted work.

In 2024, the European Union adopted Regulation (EU) 2024/1689, to harmonise rules on artificial intelligence.³⁰ This regulation required general-purpose AI (GPAI) developers to identify and respect the reservations expressed under Article 4(3) of the Copyright Directive, and publish a detailed summary of the training content used to build the model so creators can verify if their opt-outs were honored.³¹

²⁹ <https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng>

³⁰ <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

³¹ The EU AI Act and copyrights compliance, April 30, 2025. International Association of Privacy Professionals. <https://iapp.org/news/a/the-eu-ai-act-and-copyrights-compliance>; and AI and copyright: The training of general purpose AI, European Parliament, April 23, 2025. https://www.europarl.europa.eu/thinktank/en/academic/EPRS_ATA%282025%29769585

Japan's Article 30-4 exception for use of copyrighted works to train AI

Japan's copyright exception for AI model training is primarily codified under Article 30-4 of the Japanese Copyright Act (enacted in 2018 and effective in 2019). The exception allows the use of copyrighted works, without consent from copyright holders, to train GenAI, when the use is not for a purpose that would unreasonably prejudice the interests of the copyright holder.³² The exception addresses the training of GenAI services, but not to outputs that are distributed in a way that substitutes for the original work's market.³³ According to Chie Kasahara, of the law firm Atsumi & Sakai:³⁴

"The article 30-4 copyright carve-out allows companies to ingest copyrighted text, images, audio and video into training pipelines without obtaining individual licences, provided that the purpose remains analytical rather than expressive. In practice, this covers the following activities:

- **Corpus construction.** Assembling large-scale datasets of web-scraped text, licensed image libraries or broadcast recordings for the sole purpose of training model parameters.
- **Feature extraction and benchmarking.** Running copyrighted materials through neural networks to derive statistical weights, embeddings or evaluation metrics without reproducing or distributing the source works.
- **Internal research and development.** Using copyrighted datasets to fine-tune proprietary models that will generate original outputs, where the training data itself is never surfaced to end users.

The WIPO analysis on AI and copyright in Japan confirms that Japan's TDM exception is among the broadest in any major jurisdiction, because Article 30-4 does not impose a sector limitation (it applies equally to commercial and non-commercial actors) and does not require that the work be "lawfully accessed," although separate legal theories, such as terms-of-service breach, may still constrain scraping activities."

Singapore

³² Expressed In the Japan statute, "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." <https://www.cric.or.jp/english/clj/cl2.html>

³³ General Understanding on AI and Copyright in Japan" -Overview - published by the Legal Subcommittee under the Copyright Subdivision of the Cultural Council, Agency for Cultural Affairs, Government of Japan. https://www.bunka.go.jp/english/policy/copyright/pdf/94055801_01.pdf

³⁴ Generative AI and Copyright in Japan 2026: Practical Licensing & Compliance Guide. Global Law Experts. <https://globallawexperts.com/generative-ai-copyright-japan/>

In 2021, Singapore enacted a dedicated exception for "Computational Data Analysis" (CDA), codified under Sections 243 and 244 of the Singapore Copyright Act 2021. The new law is designed to attract AI developers and data-driven industries.³⁵ The exception covers both commercial and non-commercial machine learning. The exception is limited to cases where the user had lawful access to a work Section 244(2)(d) and if one of the conditions in Section 244(2)(e) is met: if the user does not and could not reasonably know that the use was from infringing copies or was for a prescribed purpose.

United States

There is no specific exception in the U.S. Copyright law for text and data mining, and with no consensus in Congress, policy is being made by a series of court cases over what constitutes fair use under the four factor test set out in 17 U.S.C. 107.

In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include

(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.

Among the recent court decisions:

Bartz v. Anthropic (Alsup. No. 3:24-CV-05417, N.D. Cal). This class action, involving nearly half a million books, held that training large language models on lawfully acquired books was an "exceedingly transformative" fair use.³⁶ However, Judge Alsup held that Anthropic's acquisition and retention of millions of pirated books in a "central library" was not fair use

³⁵ Kin Wah Chow, Singapore: Copyright Infringement Defence for AI Machine Learning, Rouse.com February 26, 2024. "This is the first in the South East Asian region and no doubt with the aim to attract AI investments into Singapore."

<https://rouse.com/insights/news/2024/artificial-intelligence-in-singapore-copyright-infringement-defence-for-artificial-intelligence-machine-learning>

³⁶ In *Bartz v. Anthropic PBC* (No. 3:24-CV-05417, N.D. Cal.), Senior U.S. District Judge William Alsup issued his pivotal summary judgment order on June 23, 2025. Settlement in Principle (August 26, 2025): Following the partial summary judgment order, the parties reached a stipulation to settle the remaining piracy claims out of court. Preliminary Approval (September 25, 2025): The court preliminarily approved a \$1.5 billion class-action settlement covering rightsholders whose books were downloaded via pirate datasets. Final Approval (July 20, 2026): U.S. District Judge Araceli Martínez-Olguín granted final approval to the \$1.5 billion settlement.

and constituted copyright infringement. The remaining infringement claims were later resolved through a \$1.5 billion class settlement, avoiding a trial on damages. The settlement provided compensation averaging roughly \$3,000 per covered title before deductions for fees and expenses.

Kadrey v. Meta (Chhabria, No. 3:23-CV-03417, N.D. Cal.). Issued the same week³⁷, the court similarly concluded that training Meta's LLMs on copyrighted books was transformative fair use on the record presented. The plaintiffs alleged that Meta obtained many of the books from pirated shadow-library sources using BitTorrent, but Judge Chhabria stressed that his ruling was limited to the evidence before him and did not foreclose future cases based on stronger evidence of market harm.

Thomson Reuters v. Ross Intelligence (No. 1:20-CV-00613-SB, D. Del.). A court held that the use of Westlaw notes describing court cases were used to create a service that was directly competitive and undercut Westlaw's market.³⁸ This case is on appeal, and one of the issues in the appeal is whether or not the Westlaw headnotes are themselves protected by copyright.

US case law appears to be converging on findings that lawfully sourced training data is likely to be fair use, while piracy and direct market-substitution outputs are not.

Jurisdictions with an explicit lawful-access requirement built into the statute

Internationally, provenance (licensed vs. scraped vs. pirated) has become an important factor in litigation, and is also an issue in some statutory exceptions.

EU. Both Art. 3 and Art. 4 of the Digital Single Market (DSM) Directive conditions the exception on the user having lawful access to the work.

UK. Section 29A of the Copyright, Design and Patents Act requires lawful access as a precondition to the exception for copies for text and data analysis for non-commercial research.

Singapore. Section 244 requires that the user has lawful access to the first copy, and that the first copy is not itself an infringing copy — the statute gives the specific example that paywall circumvention defeats "lawful access."

³⁷ U.S. District Judge Vince Chhabria granted summary judgment on AI Training & Fair Use on June 25, 2025.

³⁸ U.S. Circuit Judge Stephanos Bibas (sitting by designation in the U.S. District Court for the District of Delaware) issued a landmark summary judgment order on February 11, 2025.

Germany. §§44b, 60d of the German Copyright Act allows TDM reproduction of lawfully accessible works, including commercially, absent a rights reservation.

Japan is an exception. Article 30-4 does not make any explicit distinction between legally accessed and non-legally accessed materials; unlike the TDM provisions in the EU, UK, and Singapore, it doesn't statutorily prohibit training on pirated content. The Agency for Cultural Affairs' "General Understanding" issued guidance that if a developer knowingly collects training data from a site known to contain pirated content, that creates a high likelihood of liability, but this guidance is not consistent with the statute, which omits mention of a lawful-access condition.

Training AI on copyrighted works of a scientific, legal or technical nature

KEI has argued that, with regard to the use of copyrighted works, distinctions should be made between cultural works, such as recorded music or novels, and works of a scientific, legal or technical nature. This is particularly important when it comes to medical information. These passages are from James Love's 2023 article in *Scientific American*:

Clearing rights to copyrighted work can be difficult even for simple projects, and for very large projects or platforms, the challenges of even knowing who owns the rights is nearly impossible, given the practical requirements of locating metadata and evaluating contracts between authors or performers and publishers. In science, requirements for getting consent to use copyrighted work could give publishers for scientific articles considerable leverage over which companies could use the data, even though most authors are not paid.

Differences between who owns what matter. It's one thing to have the copyright holder of a popular music recording opt out of a database; it's another if an important scientific paper is left out over licensing disputes. When AI is used in hospitals and in gene therapy, do you really want to exclude relevant information from the training database?³⁹

Similar points have been made in several government consultations, including by KEI at the 2025 Special 301 hearing.⁴⁰

Any country that provides robust exceptions for using copyrighted material for AI will have a significant advantage in terms of training such services. But a lack of harmonization may create a situation where services developed in one country, such as in the United States, will not be legal in another, because of non-authorized use of copyrighted works to train the service. This makes it a significant trade issue.

³⁹ James Love, We Need Smart Intellectual Property Laws for Artificial Intelligence: "One-size-fits-all" regulation will sideline medical and research benefits promised by the advent of artificial intelligence, *Scientific American*, August 7, 2023.

<https://www.scientificamerican.com/article/we-need-smart-intellectual-property-laws-for-artificial-intelligence/>

⁴⁰ KEI Submissions to 2025 Special 301 Review, February 24, 2025. <https://www.keionline.org/40497>

The most emotive and politically important voices to make it an infringement of copyright to train AI services are cultural industries and journalism. KEI has been concerned that restrictive policies on the uses of copyrighted works to train AI will be extended to much broader classes of works. A general exception should be a broad text and data mining exception for AI, but we'd be open minded about whether there should be carve-outs for certain cultural industries and journalism. What we don't like is a one-size-fits all approach that would treat areas like cultural industries and biomedical research the same.

We believe that AI will be increasingly used in law, scientific and biomedical R&D, and in those areas, society is best off if AI services have access to everything possible in terms of data, and where omissions may cost someone their health or freedom.

In regard to scientific and medical information, we note that unlike novelists or screenwriters, research journal authors are rarely paid for their works, and a handful of companies control a large number of journals, many of them foreign-owned. It would be a very bad outcome if private equity firms were able to significantly limit which companies can use the leading medical journals to train AI services, or for that matter, if any publisher can opt-out of the science being used to train programs that are used for drug discovery or to treat and protect patients. We're not saying that screenwriters could or could not opt-out, but that an opt-out right should be an exception to a more general exception, and not the rule.

Using traditional medical knowledge to train AI

There are a number of differences between copyrighted works and traditional medical knowledge (TMK) to consider.

1. As noted above, extracting facts, statistical information, or patterns from copyrighted works generally does not implicate copyright because copyright protects expression, not facts or ideas. Traditional medical knowledge is mostly about facts and ideas.
2. In many jurisdictions, copyright only applies when there is a fixation of work. Traditional knowledge in some communities is not written and is disseminated orally.
3. When an AI is trained on TMK, it absorbs the core value, the functional knowledge, which is unprotectable under copyright.
4. There is a dynamic and growing production of copyrighted works, while traditional medical knowledge is, in some sense, if not fixed, more limited in terms of its supply. Copyright is an incentive for a continuous flow of new creative output. TMK represents a cumulative body of knowledge passed through generations, whose supply is bounded by cultural heritage and community practice, and at risk of being lost.

5. Copyright and neighboring rights are initially held by authors, performers, or corporate entities (publishers, record labels, film studios) and for a limited term before entering the public domain. TMK may be held collectively by a community, tribe, or lineage, for which it may have a perpetual relationship.
6. Copyright and neighboring rights for performers or producers are initially assigned to authors, performers or corporate producers of audio or audiovisual works. Traditional medical knowledge may be held in a legal sense among a community, or more broadly diffused with no ownership assigned.

A TDM exception for TMK, drafted using the approach for copyright, where expressions are protected and facts and ideas are not, doesn't produce a proportionate result. A copyright TDM exception takes an existing bundle of rights (reproduction, adaptation, distribution) and carves out one specific permitted use: making copies for computational analysis. Before the exception, the rightsholder controls all uses of the expression. After the exception, the rightsholder still controls everything except the training use itself. The market for publishing a book for reading or making an audio version, or publishing derivatives, remains intact and enforceable. A copyright TDM exception removes a part of the exclusivity bundle and leaves the rest standing.

What the text and data miner extracts (e.g., facts, ideas, statistical co-occurrence, syntactic patterns, and semantic relationships) was never protected by copyright. The copyright exception's effect on the copyright holders' protected interest is limited. TMK, on the other hand, is primarily about facts and ideas, and once a GenAI program uses it for training, it captures the entirety of the knowledge asset.

This distinction can be overstated. Some TMK involves expressions, which in theory and to some extent in practice, can be protected on the output side, if a *sui generis* regime protecting the expressions exists. And if you mine a scientific journal for facts, ideas and relationships, you may extract everything a downstream user actually wanted (the data and the findings), leaving behind only the sentence-level prose expression of little economic value to readers.

For many scientific, legal and technical articles, the principal value is informational, and if an AI service is trained on the underlying facts, discoveries, methods, or relationships, the market for the expressive presentation may be much smaller than before TDM, leaving a publisher with a legally protected asset whose commercial value has been significantly displaced.

Remuneration approaches

Creative communities and publishers have pressed governments to narrow copyright exceptions for training AI, and failing that, to provide remuneration to copyright holders.

Some of the policies focus on the importance of transparency of the works used in training a GenAI service. The negotiations and policy discussions in the European Union, the UK, India and in the Cali Fund of the Convention on Biological Diversity are examples that have illustrated the technical challenges of mapping the use of a work into a GenAI weight or its use for a particular query or output.

Training data attribution (TDA) challenges

Mapping how a specific work influences an AI's model weights and its query outputs is complex, especially for large models. During training, a work's information is fragmented and compressed across billions of static parameters (weights). When a user submits a query, the model breaks the text into tokens and routes them through dozens of layers and thousands of specialized 'attention heads.' Because these internal activation pathways change dynamically with every query, tracing an output back to a single original source is very difficult. It is difficult to track the training data to the model weights, and then to track weights to outputs. For example, tracking data to weights and tracking weights to outputs.

Tracking data to weights

Connecting training data inputs to specific parameter values across billions or trillions of model weights faces severe computational and structural hurdles.

- **High Dimensionality and Superposition.** Neural networks compress vast amounts of information into shared vector spaces. Concepts are stored in superposition (where many features share overlapping dimensions) and polysemantic neurons (individual neurons that activate for entirely unrelated concepts). Information isn't saved in discrete "files" or specific weights.
- **Non-Linear, Non-Deterministic Optimization.** Large models are trained via stochastic gradient descent (SGD) across billions of tokens. Factors like batching order, learning rate schedules, random seeds, and float precision mean that two training runs on identical data yield completely different weight matrices, making exact causal attribution non-deterministic.
- **The Computational Cost of Counterfactuals.** A recognized method of proving influence is counterfactual retraining (i.e., "What would the model parameters look like if sample X were removed?"). Doing this for large models is computationally impossible at scale.
- **Approximation Limitations (Influence Functions).** Techniques like Influence Functions and EK-FAC (Eigenvalue-corrected Kronecker-factored Approximate Curvature) attempt to mathematically estimate a data point's gradient effect without retraining. However, math only works for short and simple learning processes. Standard influence functions assume the AI's learning journey was like rolling a marble down a simple, smooth bowl with one clear

bottom, but AI models often are more like complex surfaces with millions of peaks, valleys, and dead ends, and the math used to predict influences can give wrong answers.

- **Non-linear multi-epoch training.** During training, an AI reads through the same dataset multiple times (each pass is an "epoch"). Every time it sees a piece of information again, it interprets it slightly differently based on everything else it has learned in the meantime. Influence functions struggle to track these compound, cascading effects over time, making it nearly impossible to isolate the exact impact of one specific lesson.

Tracking weights to outputs

Once data alters parameter weights, determining how those weights influence outputs from a specific query introduces another layer of complexity:

- **Diffuse Knowledge vs. Verbatim Memorization.** Influence studies show that model output answers rarely stem from a single training document, and may rely on thousands of weakly influential documents that collectively build an abstract concept. Isolating a single data source is straightforward only when the model has verbatim-memorized a unique passage.
- **Tracing weights across layers.** Even if AI clearly recommends a traditional remedy, researchers cannot easily map the path the AI took through its internal machinery to generate that output. It likely used multiple sources, forming a temporary internal "circuit" that blends fragments of Indigenous herbal wisdom from multiple sources, modern pharmacological papers, and general botany texts all at once. If an AI service receives an input on traditional medical knowledge, such as "using powdered bark from Tree X reduces fever when steeped in warm water," it does not save that as a single, readable document, but breaks down the information into abstract mathematical fragments and spreads them across internal layers. One layer might process plant names, another processes symptoms like "fever," and another tracks preparation methods.

Dynamic learning inside the models. Modern AI models are not static, trained once. There are alignment overwrites, which involve learning from human feedback, and (RLHF / DPO): Reinforcement Learning from Human Feedback (RLHF) and Direct Preference Optimization (DPO) significantly shift model outputs and safety behaviors. A pre-training data point might dictate underlying knowledge, but over time, the post-training alignment determines whether or how the model expresses or refuses to state that knowledge. This learning and other fine-tuning instructions can alter or suppress pre-training weights, making it hard to determine whether a capability or fact originated in pre-training or fine-tuning datasets.

The attribute of how a document used to train an AI contributes to a model's weights is computationally near-impossible at scale, and proving a direct line of causation from training data to generated text remains one of the hardest problems in AI governance and computer science.

It is not pointless to propose remuneration schemes, but the technical impossibility of fine-grained, post-hoc attribution fundamentally changes how those schemes can be structured.

Collective management of rights

When one cannot mathematically isolate how much an individual document contributed to a specific model parameter or output, remuneration models cannot function like precision micro-transactions. Instead, policymakers, rights organizations, and AI developers focus on compensation models that partly or largely bypass the attribution problem and treat a group or class of training inputs together through collective management approaches and through voluntary or statutory licensing models.

When considering if and how AI developers can or should compensate parties that create or hold rights to training inputs, it is useful to recognize that collective management approaches have often been implemented to address remuneration or compensation of holders of copyright, in a variety of cases where it is not practical to match uses of works with compensation to individual rights holders.

There are many collective management organizations (CMO) in copyright, some public and some private, some created by or legitimated by statute, and some organized voluntarily.

Mihály Ficsor's *Collective Management of Copyright and Related Rights* (WIPO Publication No. 855)⁴¹ is an in-depth discussion of different CMO systems, legislative options, and practices worldwide, aimed at governments preparing or applying collective management norms, with detailed treatment of competition-law aspects, national treatment, and different models of collective management.

Collective management exists precisely for situations where individual direct licensing breaks down due to high transaction costs. For AI, scraping billions of tokens from millions of distinct authors creates an insurmountable transaction cost if done on an individual basis. CMO models substitute millions of micro-negotiations with blanket licensing structures. They can be organized along different approaches. For example:

Voluntary Collective Management: Rights holders opt in to grant licensing authority to a CMO. Most CMOs operate within a single national boundary. Global federations, such as CISAC (authors and composers), IFRRO (reprographic rights), SCAPR (performers), and BIEM (mechanical rights), play a role in coordinating bilateral reciprocal representation agreements, setting technical standards and lobbying. There are exceptions, for example, driven by online music streaming and digital publishing. Under the EU Collective Rights Management (CRM) Directive (Directive 2014/26/EU), major European CMOs can grant

⁴¹ Mihály Ficsor's *Collective Management of Copyright and Related Rights* (WIPO Publication No. 855), 3rd edition. WIPO, 2022.
<https://www.wipo.int/edocs/pubdocs/en/wipo-pub-855-22-en-collective-management-of-copyright-and-related-rights.pdf>

multi-territorial licenses for online music rights across all EU member states, allowing digital service providers (DSPs) to obtain a single license covering multiple European territories.

Extended Collective Licensing (ECL): In some states, a CMO representing a substantial number of rights holders is authorized by statute to extend the license to works of non-member rights holders. This approach has been proposed in some European jurisdictions as a way to allow AI developers to mine cultural archives without facing copyright infringement claims from unrepresented authors.

Mandatory Collective Management (MCM): By statute, rights holders cannot negotiate licenses or seek injunctions, and their rights must be exercised and managed by a CMO. Common examples include cable television or the Artist's Resale Right (Droit de suite).⁴²

Statutory Remuneration: Some uses of copyrighted works are subject to mandatory licensing. For example, in many jurisdictions, once a musical composition has been commercially released, any recording artist can record and distribute their own version, without consent from the songwriter, subject to payment of a royalty set by law. Rights holders can negotiate lower royalties, and the statutory license sets a ceiling on the royalty.

Some of the copyright collective management regimes do have a precise matching of users and remuneration, for example, cable television, music mechanical licensing fees, and artists' resale right. In each of these cases, there are events which can be accurately monitored at a low cost, relative to the remuneration.

When it is not practical, cost-effective or even possible to track and compensate for all uses of copyrighted works, firms, governments and collection societies use a variety of methods to allocate remuneration funds to individual rights holders. This may involve a variety of proxy mechanisms, statistical sampling methodologies, and flat allocation formulas to distribute collected funds fairly without incurring prohibitive administrative costs. Among them are four primary structural methods to allocate remuneration:

1. Statistical Sampling and Surveying

Instead of tracking every single transaction or user, CMOs monitor a representative subset of usage and extrapolate the results across the entire pool of collected royalties.

⁴² An economic right that entitles visual artists and their heirs to receive a royalty percentage whenever their original physical artworks are resold in the secondary market through art market professionals (such as auction houses, commercial galleries, and art dealers). Under Article 6(2) of the EU Resale Rights Directive (Directive 2001/84/EC), national governments are given the discretion to decide whether management of the Artist's Resale Right (ARR / droit de suite) should be mandatory (compulsory) through a collective management organization (CMO). Over half of the EU's 27 Member States have opted for mandatory collective management in their national copyright statutes. Under these regimes, artists and their heirs cannot collect royalties directly from art dealers or auction houses; they must exercise their claim through an accredited national CMO.

Log-Book Sampling: Used extensively for radio broadcasting and venue background music. A rotating sample of radio stations or venues submits complete logs of works played during a specific week. The CMO uses these logs as a statistical proxy to distribute the total pool collected from all venues in that category.

Point-of-Sale and Inventory Sampling: Used in reprographic rights (photocopying and digital scanning in schools or libraries). CMOs conduct periodic surveys across select universities or schools to determine which book chapters or academic journals are most frequently copied, allocating blanket educational fees accordingly.

2. Market Proxies and Top-Seller Data (Analogy Allocation)

When a specific category of use cannot be sampled directly, CMOs map distribution to data from an adjacent, fully monitored market.

1:1 Market Mapping: For private copying levies (surcharges collected on blank media, smartphones, and hard drives), it is impossible to inspect private consumer devices. CMOs often distribute these levies based on top-100 commercial music sales, streaming logs, or commercial radio airplay data, operating under the assumption that private copying habits mirror commercial market popularity.

Broadcast Proxies: In public performance licensing for small commercial establishments (e.g., cafes, hair salons), fees collected from business owners are often distributed using terrestrial television and radio broadcast logs as the allocation proxy.

3. Title-Based Catalog Inclusion (Pro-Rata / Flat Distribution)

Rather than measuring actual consumption or usage, allocation is calculated based on the availability or volume of works made accessible in a licensed database or archive.

Per-Title / Per-Page Weighting: Blanket licensing fees collected from corporate or academic entities are divided pro-rata among rights holders based on the total number of registered works or pages contributed to the licensed repertoire.

Tiered Catalog Valuation: Works are categorized by domain authority, length, or genre, and rights holders receive fixed baseline payments based on catalog size regardless of individual access metrics.

4. Cultural, Social, and Sectoral Dividends

In many jurisdictions—particularly in Europe and Latin America under statutory rules—a fixed percentage of unallocated or blanket remuneration is legally diverted from individual distribution entirely.

Deductions for Cultural and Social Purposes: Under statutes like the EU Collective Rights Management Directive (Directive 2014/26/EU, Article 13) or French copyright law, CMOs are permitted or required to deduct a statutory percentage (e.g., up to 25% of private copying levies) to fund national cultural grants, live performance funds, emergency welfare assistance, or pensions for creators.

Indivisible Reserve Pools: Unallocated funds resulting from untraceable or orphan works are held for a statutory period (e.g., 3 years) and subsequently reallocated to general creator funds or used to offset administrative overhead for the entire membership.

Once transactions based on consent between buyers and sellers are replaced with collective management, the methods of allocation adopted by the collective management system determine the distribution of monetary benefits. Choices often default toward metrics that are easy to measure and are influenced by the relative power of different stakeholders.

Music Streaming and the evolution of collective licensing

In 1851, the Société des Auteurs, Compositeurs et Editeurs de Musique (SACEM), the first performing rights society in the world, was created, uniting 350 songwriters, composers and sheet music publishers as a cooperative. Today WIPO estimates that there are more than 1,000 collective management organizations across all creative sectors: music, film, books, visual arts, and reprography. Nearly all of these CMOs are national, and every country with a copyright law typically has 3 to 8.

The development of new technologies and the lobbying by authors, performers, publishers and distributors have driven changes in the national and international copyright laws, and the operation of CMO and direct licensing of rights.

Digital streaming music services were reportedly \$22 billion in 2025, representing 70 percent of all revenue from recorded music (compared to less than 20 percent of revenues in 2015 and less than 2 percent in 2004).⁴³ This has led to a number of changes at the national, regional and international level.

The intellectual property rights associated with the streaming of recorded music differ from country to country and are complicated everywhere. There are rights associated with the master recordings, the rights for composers, and rights associated with the distribution and making available of those rights. The master recording may have rights assigned to producers and performers, both featured and sessions, and separately for composers, with separate rights for the music, the lyrics and the

⁴³ GLOBAL MUSIC REPORT 2026, IFPI.
https://www.ifpi.org/wp-content/uploads/2026/03/GMR2026_SOTI.pdf

arrangements. The distribution rights will depend upon the technology, with on-demand streaming and non-interactive streaming treated differently in many jurisdictions. Some of these rights are managed through national or regional (in Europe) collective management organizations, and some through direct licensing.

For interactive on-demand services, like Spotify, Apple Music, YouTube Music, Amazon Music Unlimited, Tidal or Deezer, corporate entities or non-profit licensing agencies negotiate directly with the streaming services for performer rights. For radio-style non-interactive streaming, record labels typically work through a national CMO to collect performer royalties. Rights for composers have largely been administered through CMOs, but there has been a shift toward direct licensing.

National CMOs are generally prohibited from taking an equity interest in digital streaming platforms (DSPs),⁴⁴ but music labels, publishers and performers, and direct licensing organizations face no such restrictions.

The recorded music market has been highly concentrated at the ownership level recordings. In 2025, the Universal Music Group (UMG) had 32.5 percent of the combined global digital/physical revenue. Sony Music Entertainment (SME) had 22.7 percent, and the Warner Music Group had 14.8 percent. The collective independent sector was roughly 30 percent.

Merlin Network

In 2007, the Merlin Network was created as a non-profit, member-owned licensing agency. Merlin operates globally as a collective bargaining agency for the independent label sector, allowing thousands of independent artists and smaller labels to negotiate digital licensing terms as a bloc rather than each cutting individual deals with streaming services. It has several features that are different from a typical CMO. Merlin's role is focused on rights negotiation for sound-recording rights, and not the distribution to the artists or labels. It can take equity in a DSP, and its fees are low, 1.5 percent of licensing revenues. Participation in Merlin is strictly voluntary, with relatively easy entry and exit paths, and licensing through Merlin is non exclusive, in the sense that an artist or label can use Merlin to negotiate with some services, but not others. Today the Merlin Network represents roughly 15 percent of the global music market share, leading it to be described as a 4th major, and is credited with increasing the influence of independent artists.

Pro-rata, artists-centric and user-centric allocations of music streaming royalties

Various studies have found that a small number of artists generate the majority of music streams (1 percent of artists accounting for 90 percent of all streams, in a Rolling Stone/Alpha Data analysis of 2019–2020 streaming data.)⁴⁵

⁴⁴ Due to conflict of interest provisions in some national laws.

⁴⁵ Cite needed.

Pro rata distributions or royalties

Several services, like Spotify, use a “**pro rata**” approach for distributing royalties. In this approach, Spotify and other DSPs consider the usage of songs across the platform to allocate royalties. Under pro-rata, every subscriber's fee goes into one pool, then gets divided by each artist's share of total platform streams. A subscriber who exclusively listens to niche or independent artists effectively subsidizes the most popular artists that dominate overall stream volumes, creating a mismatch between what a listener actually pays for and where their money ends up.

Among the alternatives that have been considered or implemented are the “artists-centric,” “user-centric” and “Blur Banff” approaches.

Artist-centric models

The **artists-centric** model was first pioneered by the service Deezer to provide extra weight on streams when users search by artists by name, or artists have minimum stream and unique-listener thresholds. This was proposed as a filter against background noise and streaming-farm fraud, not a switch to a per-listener allocation.

User-centric models

The **user-centric** approach would distribute payments to artists that a specific subscriber listened to that month. Soundcloud, an online audio streaming and music sharing platform founded in 2007, has implemented this on a limited scale for some of its services. The Warner Music Group was the first major label to buy into this approach.

Blur-Banff Model

The **Blur-Banff Model**, which has been proposed⁴⁶ but not implemented, is an even more user-centric approach. In this proposal, subscribers could direct their payments to artists be managed by an intermediary of their choice, with the intermediaries regulated for transparency, and offering the users different pay-off models, from a pro-rata system, to more payoffs that favored certain music genres, favored session musicians, flatter lower payments to most popular acts and relatively large payments to independent or less known artists, or any anything else, an intermediary was offering..

⁴⁶ Notes on the Blur Banff proposal, KEI Research Note 2013:1. <https://www.keionline.org/blur-banff>, and discussion in: William W. Fisher III, Promises to Keep: Technology, Law, and the Future of Entertainment, Stanford University Press, August 2004, Chapter 6: An Alternative Compensation System. <https://opencasebook.org/casebooks/424-copyrightx-bocconi-university/resources/11.1-william-w-fisher-iii-promises-to-keep-technology-law-and-the-future-of-entertainment-chapter-6-an-alternative-compensation-system/>

The artist-centric and user-centric approaches are both formulas for splitting an existing royalty pool differently than the pro-rata approach. The Blur-Banff model is different; it decentralizes the choice of formula itself to competitive intermediaries selected by listeners, who could in turn implement pro-rata, user-centric, genre-weighted, or any other model.

Access and benefit sharing regimes and collective management of rights II.

National, regional or multilateral access and benefit sharing (ABS) regimes can be implemented as collective management systems.

The WHO PIP Framework and the newer WHO pandemic agreement both provide the WHO with monetary payments from some users to defray the cost of operating the system. Manufacturers using the system contract to provide a combination of free or reduced-priced vaccines or other products. The donations and price regulated products are managed by the WHO, to be used in countries determined by the WHO, which are not necessarily the same as the countries that share pathogens. There is also optional licensing of patents and manufacturing know-how, and a set of possible non-monetary benefits.

The International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) Multilateral System (MLS) for access and benefit-sharing covers a set number, initially 64, of food and forage crops and provides for both monetary and non-monetary benefits, as well as a robust set of exceptions to benefit sharing obligations.

The Nagoya Protocol's Article 10 anticipates the possibility of delinked, multilateral ABS mechanisms, where genetic resources or associated traditional knowledge occur in transboundary situations, or where prior informed consent cannot be granted or obtained.

The Cali Fund, built later under a separate CBD decision, and linked to negotiations on digital sequence information, is an attempt to implement an ABS system where there is no direct link between the obligation for monetary contributions to a global fund, and the use of a specific genetic resource or traditional knowledge. Additionally, half of the monetary contributions are earmarked for the benefit of indigenous peoples, even if no traditional knowledge is involved in a company's products. The delinking was a deliberate negotiating outcome after an alternative was rejected.

The BBNJ Treaty (the Agreement Under the United Nations Convention on the Law of the Sea on the Conservation and Sustainable Use of Marine Biological Diversity of Areas Beyond National Jurisdiction) is primarily focused on genetic resources for which no government can claim sovereignty, but it also makes references to the rights in traditional knowledge held by indigenous peoples and local communities. The BBNJ Treaty anticipates monetary benefits to a global fund that will benefit developing countries, and indigenous peoples and local communities.

All of these agreements also have provisions to provide non-monetary benefits to developing countries, and in some cases, to indigenous peoples and local communities.

In all of the multilateral ABS negotiations the use of materials and digital sequence information by AI services is considered a significant issue, and also, the question of how the monetary and non-monetary benefits are managed.

For global funds that are to benefit Indigenous peoples and local communities, the challenge of choosing beneficiaries has often been at least partly unresolved or deferred.

The negotiations on the implementation of the Cali Fund are potentially an important precedent, although it remains a work in progress. The monetary contributions to the Cali Fund are voluntary, and the initial targets, of 1 percent of global company wide profits or 0.1 percent of global company wide revenues, for entire sectors of the global economy, are widely considered unrealistic, and expectations have not been matched with company buy-in.

The 2024 decision on "Digital sequence information on genetic resources," adopted at the Conference of Parties 16 to the CBD (Cali, Colombia, 2 November 2024),⁴⁷ includes these features:

1. Funding should support the realization of the objectives of the CBD in developing country Parties, especially for the conservation and sustainable use of biodiversity, and contribute to scientific research on biodiversity, and benefit indigenous peoples and local communities, including women and youth within those communities.
2. Funding will be allocated according to an overall allocation formula, being developed by an Ad Hoc Technical Expert Group, based upon biodiversity richness, geographic origin of the genetic resources from which digital sequence information was derived, and capacity needs.
3. Where appropriate, and subject to national circumstances and national legislation, at least half of the funding of the global fund should support the self-identified needs of indigenous peoples and local communities, including women and youth within those communities.
4. The fund, based on the overall allocation formula, will be provided "through government authorities or by direct payments through institutions identified by indigenous peoples and local communities.

Decision adopted by the Conference of the Parties to the Convention on Biological Diversity on 1 November 2024

16/2. Digital sequence information on genetic resources

18. Funding should support the realization of the objectives of the Convention in developing

⁴⁷ CBD/COP/DEC/16/2, available at <https://bd.int/doc/decisions/cop-16/cop-16-dec-02-en.pdf>.

country Parties, in particular the least developed countries and small island developing States, and Parties with economies in transition, especially the conservation and sustainable use of biodiversity, including through the delivery of activities described in national biodiversity strategies and action plans; contribute to scientific research on biodiversity; benefit indigenous peoples and local communities, including women and youth within those communities; and support the building of capacity, in accordance with Article 16 of the Convention, to generate, access, use, analyse and store digital sequence information on genetic resources according to capacity needs. Funding will also be available for those purposes to indigenous peoples and local communities in developed countries, where appropriate. In the event that any other intergovernmental forums decide to make use of the multilateral mechanism to share the benefits from the use of digital sequence information on genetic resources, the funding should also support the realization of their objectives.

19. Funding will be allocated taking into account the overall level of funding available in the global fund and an indicative list of criteria, as set out in enclosure II. A formula will be determined by the Conference of the Parties at its seventeenth meeting on the basis of the work of a group established with the terms of reference provided in enclosure III.

20. Funding to Parties will be disbursed through direct allocations to countries, as described in paragraph 19. Each recipient Party is invited to designate or establish, as appropriate, a national entity, such as a national biodiversity fund, to receive funds and distribute them in a transparent manner to support the activities described in paragraph 18. Such entities may allocate resources, in a transparent manner, on the basis of projects developed through a country-driven or community-driven process and should be accountable for ensuring that the funds are used for the self-identified purposes for which they are distributed. They should operate according to internationally accepted fiduciary standards and provide reports on the activities undertaken under the fund and on their impacts. Recipient Parties, at their own discretion, may alternatively designate an international, regional or subregional entity to fulfil those functions.

21. Where appropriate, and subject to national circumstances and national legislation, at least half of the funding of the global fund should support the self-identified needs of indigenous peoples and local communities, including women and youth within those communities, through government authorities or by direct payments through institutions identified by indigenous peoples and local communities.

22. The Conference of the Parties may set aside a proportion of funds to support capacity-building and capacity development, ensuring that developing country Parties, in particular the least developed countries and small island developing States, and Parties with economies in transition have access to the tools and expertise necessary to fully participate and benefit from digital sequence information on genetic resources.

Identifying indigenous peoples or local community beneficiaries

Although half of the funding is earmarked for indigenous peoples and local communities, nothing in the decision defines indigenous peoples or local communities, or ties the funding to traditional knowledge associated with the genetic resources (or their digital sequence information). The

identification and recognition of indigenous peoples or local communities will be the responsibility of national governments that are Parties to the CBD.

The overall allocation formula, which will determine how much money goes where, across countries and communities, will be decided by an Ad Hoc Technical Expert Group, and may be adopted at COP 17, in October of 2026.

Indigenous peoples and local communities will self-identify their needs (what the funding should support), and governments or institutions identified by indigenous peoples and local communities will spend the money.

It remains to be seen how this will work, and as of August 2026, the Cali Fund has only \$1,000.

There are a large number of indigenous communities, and even a larger number when local communities are considered. These involve different leaders and governance traditions, and may involve different languages.

- **Australia.** The 2006 Census of Housing and Infrastructure in Aboriginal and Torres Strait Islander Communities, which identified 1,187 discrete Indigenous communities, and 167 Aboriginal and Torres Strait Islander languages were reported in use at home in the 2021 Census.
- **New Zealand.** Commonly estimated to have 124 iwi (tribes).
- **Canada.** There are 634 recognized First Nations bands/communities, representing over 50 distinct Nations and 50+ Indigenous languages, according to Indigenous Services Canada and Crown-Indigenous Relations. In addition, there are 51 Inuit communities across four regions (Nunatsiavut, Nunavik, Nunavut, and the Inuvialuit Settlement Region).
- **Brazil.** The 2022 IBGE census recorded 391 distinct Indigenous ethnic groups/peoples, up sharply from 305 in the 2010 census, with the increase attributed mainly to methodology changes and growing self-identification, speaking 295 Indigenous languages, totaling roughly 1.69 million people. At the settlement level, the same census identified 8,568 Indigenous localities (villages, communities, or other forms of settlement with 15+ people), about 72 percent inside officially demarcated Indigenous lands.
- **Peru:** There are 55 officially recognized Indigenous peoples, speaking 47 Indigenous languages, according to Peru's Database of Indigenous or Original Peoples, including 51 in the Amazon and 4 in the Andean region. Nearly 6 million people self-identified as Indigenous in the 2017 census (about a quarter of the population). At the settlement/governance level, the 2017 agricultural censuses counted 2,703 titled or recognized native communities (comunidades nativas, mostly Amazonian) and 6,682 peasant communities (comunidades campesinas, mostly Andean/coastal), roughly 9,385 combined, though a substantial share still lack formal land title.

- **Thailand:** Officially, only 9 "hill tribe" groups are recognized by the state (Hmong, Karen, Lisu, Mien, Akha, Lahu, Lua, Thin, Khamu), with a government-counted hill-tribe population of about 925,825 as of 2002. Broader, non-government estimates put Thailand's actual Indigenous population, including coastal Chao Ley and other unrecognized highland and forest groups, at around 5–6 million people, or roughly 7–9.7% of the population, spanning more than 20 distinct ethnic groups.
- **India:** There are 705 ethnic groups officially recognized as Scheduled Tribes under Article 342 of the Constitution. Of the 705, 75 are further designated Particularly Vulnerable Tribal Groups.
- **Nigeria:** Nigeria has no domestic legal category equivalent to India's Scheduled Tribes or Peru's native communities. Ethnic-group counts (commonly cited as 250 to 371, depending on the source and how sub-groups are split) describe general ethnic diversity, not a recognized "Indigenous peoples" status. International advocacy bodies do apply "Indigenous" framing to specific marginalized minorities, Niger Delta groups like the Ogoni, and pastoralist/hunter-gatherer groups, but that's an external/activist designation, not a Nigerian statutory one.
- **South Africa:** The San and Khoikhoi (jointly "Khoisan") self-identify as South Africa's Indigenous peoples and are organized into six main groups: San communities (ǀKhomani San, Khwe, !Xun) and several Khoikhoi communities (Nama, Koranna, Griqua, Cape Khoekhoe). They were not formally recognized under national legislation until the Traditional and Khoisan Leadership Act of 2021.
- **Philippines:** 110 officially recognized ethnolinguistic groups (the National Commission on Indigenous Peoples' figure, though other sources range 101–140 depending on how sub-groups are counted).
- **Chile:** 10 legally recognized Indigenous peoples under Indigenous Law 19.253, including: Mapuche (the large majority), Aymara, Rapa Nui, Diaguita, Atacameño (Likan Antai), Quechua, Colla, Kawésqar, Yagán, and Chango. About 2.1 million people self-identified as Indigenous in the 2024 census, 11.5 percent of the population.
- **Argentina:** 35 officially recognized Indigenous groups per the Complementary Survey of Indigenous Peoples, with roughly 955,000–1.3 million people self-identifying as Indigenous depending on the census year (about 2.2–3% of the population). The largest is Mapuche.
- **Bolivia:** 36 recognized Indigenous groups, representing 41 to 48 percent of the population by recent estimates, dominated by Quechua and Aymara speakers.
- **Mexico:** 68 officially recognized Indigenous Peoples (pueblos indígenas) per the National Institute of Indigenous Peoples (INPI), corresponding to 68 recognized Indigenous

languages. Around 23.2 million people identified as Indigenous in the 2020 census (about 19%), one of the largest Indigenous populations in the hemisphere in absolute terms. Mexico has ratified the International Labour Organization Indigenous and Tribal People's Convention No. 169 and, notably, amended its constitution in 2024 to recognize Indigenous peoples and communities as subjects of public law — a significant, quite recent legal upgrade.

- **Russia:** 47 officially recognized "Indigenous Small-Numbered Peoples" (40 of the North/Siberia/Far East, plus 7 in European Russia), out of more than 160 distinct peoples in the country overall. The "small-numbered" threshold is capped at 50,000 members, which is why much larger non-Russian peoples like the Sakha, Buryat, and Volga Tatars don't qualify despite being Indigenous in the ordinary sense. Together the recognized groups number just over 315,000 people, about 0.2% of Russia's population.
- **Finland and Sweden:** Both fall under a single Indigenous people, the Sámi, who inhabit Sápmi across northern Norway, Sweden, Finland, and Russia's Kola Peninsula, with no other domestically recognized Indigenous group in either country. In Finland, the Sámi split into three linguistic groups (Northern, Inari, and Skolt Sámi). In Sweden, the Sámi are organized historically around siida (traditional collective land-use units) rather than a fixed count of "communities."
- **Ireland:** No Indigenous peoples in the international or International Labour Organization sense. A recognized ethnic minority is Irish Travellers (Mincéirí), formally recognized as a distinct ethnic group in March 2017.
- **Italy:** Also no Indigenous peoples category. Italy recognizes 12 historical linguistic minorities under Law 482/1999 — Albanian (Arbëresh), Catalan, Croatian, French, Franco-Provençal, Friulian, German, Greek, Ladin, Occitan, Sardinian, and Slovenian, but this is a minority-language protection.
- **China:** 55 officially recognized ethnic minority groups (shaoshu minzu) alongside the Han majority, 56 groups total (roughly 125 million people as of the 2020 census), including well-known groups like Zhuang, Uyghur, Tibetan, and Mongol. China does not use an "Indigenous peoples" legal category domestically and has generally taken the position internationally that all its citizens are equally indigenous to Chinese territory, rejecting UNDRIP-style Indigenous-specific rights claims. China passed a new "ethnic unity" law on March 12, 2026, which took effect July 1, 2026, mandating a shared national identity framework across schools, government bodies, and enterprises.
- **Indonesia:** No formal government recognition of "indigenous peoples" as a distinct legal category; officials argue nearly all Indonesians (except ethnic Chinese) are indigenous, so specific rights claims are contested. The government does recognize 1,331 distinct ethnic groups nationally. AMAN, the main Indigenous advocacy federation, represents 2,596

customary communities with about 21 million members, and estimates the true Indigenous population at 50–70 million.

- **Colombia:** 115 Indigenous peoples recognized by the national statistics agency (DANE). Although ONIC, the main Indigenous organization, counts 102 and says only 82 are officially recognized by the government.
- **Ecuador:** 14 officially recognized Indigenous nationalities plus 18 constituent pueblos within the largest nationality, the Kichwa, which represents about 7.7 percent of the population. The 2008 Constitution declares Ecuador a plurinational, intercultural state.
- **Venezuela:** 51 distinct peoples recorded in the 2011 Indigenous Census (the most recent), representing about 2.8% of the population.
- **Malaysia:** Recognition is split across three legal regimes. Peninsular Malaysia: 18 officially recognized Orang Asli subgroups (less than 1 percent of the peninsula's population). Sarawak: roughly 25–28 native (Dayak/Orang Ulu) groups, over 70 percent of the state's population. Sabah: about 39 native ethnic groups (Anak Negeri), nearly 60% of the state's population. Altogether, Malaysia's Indigenous population (Orang Asal) is around 13–14% of the national total.
- **Papua New Guinea:** No formal recognition list. Estimates range from roughly 600 to over 1,000 distinct tribes/ethnic groups, with 800–840 living languages (about one-eighth of all human languages on Earth). Land in Papua New Guinea is instead governed almost entirely through customary tenure recognized at the clan level rather than through any state-maintained ethnic registry.
- **Democratic Republic of Congo (DRC):** No precise count of "Indigenous" groups exists. Government estimates around 600,000–700,000 (about one percent of the population); civil society organizations argue for up to 2 million (three percent). The recognized category covers three peoples collectively, Mbuti, Baka, and Batwa (the Congo Basin "Pygmy" peoples).
- **The United States of America:** There are 575 federally recognized tribes, American Indian tribes and Alaska Native Villages in the United States.
- **Japan:** Officially recognizes only the Ainu as Indigenous. UN bodies and Indigenous-rights organizations also treat the Ryūkyūans as Indigenous. IWGIA explicitly describes Japan as having two Indigenous Peoples. The Ryūkyūans comprise six major Indigenous language groups, so if defined as culturally/linguistically distinct Indigenous communities rather than "peoples," the count would be higher.
- **Switzerland:** Romansh speakers are a historic linguistic minority, but not classified internationally as an Indigenous people.

- **Germany:** Recognized national minorities include Sorbs, Danes, Frisians, and Sinti/Roma, but these are not generally treated as "Indigenous Peoples" under UN terminology.
- **France, including overseas territories:** Bretons, Corsicans, and Basques have been described as historic/native peoples, but France does not classify them as Indigenous Peoples. Within the overseas territories, French Guiana alone has six recognized Indigenous/Amerindian peoples, including Kali'na, Lokono (Arawak), Palikur, Wayana, Wayãpi and Teko. New Caledonia has the Kankak. There are also Indigenous Polynesian populations in French Polynesia and Wallis and Futuna, and France has acknowledged Indigenous peoples in Mayotte.
- **Egypt:** Nubians are the strongest candidate for Indigenous status, and some Bedouin communities are sometimes characterized as Indigenous, although referring to all Bedouins as indigenous would be inaccurate.
- **Morocco:** The International Work Group for Indigenous Affairs (IWGIA) describes the Amazigh as the Indigenous Peoples of Morocco and North Africa. Morocco's 2024 census reported Tamazight speakers to be 24.8 percent of the population. Amazigh populations are traditionally divided geographically and linguistically, including Riffians/Tarifit speakers, Central Atlas Tamazight groups, and Shilha/Tachelhit peoples of the Souss and Anti-Atlas, among others. The IWGIA reports more than 900 Amazigh associations, although associations aren't equivalent to indigenous communities.
- **Saudi Arabia:** There is no generally accepted UN-style enumeration of Saudi Arabian Indigenous peoples. Bedouin tribes have characteristics associated with Indigenous/local communities, but treating all Bedouin as a single Indigenous people would be problematic.
- **Turkey:** Kurds, Armenians, Laz, and Circassians are ethnic/national minorities or peoples, but aren't normally enumerated as Turkey's "Indigenous Peoples" in the UN system.
- **Armenia:** Armenia has national/ethnic minorities, including Yazidis, Assyrians and others, but they are generally treated as minorities rather than Indigenous Peoples.

Local Communities

References in ABS regimes to "indigenous peoples and local communities" are sometimes discussed as though they refer only to Indigenous peoples and Indigenous communities. They do not. The terms identify two overlapping but distinct categories.

There is no universally agreed definition of Indigenous peoples. The term may be associated with peoples who have historical continuity with the inhabitants of a territory, and who have maintained distinct social, cultural, linguistic, political, or economic institutions and identify themselves as Indigenous.

Local communities, on the other hand, are a broader category and do not require a community to be Indigenous or to establish priority of occupation of a territory. Local communities are generally understood as groups whose members share distinctive cultural, social, economic, or ecological practices and identities, which may, in some contexts, be associated with a particular place, territory, resource, or traditional way of life.

A community may qualify as a local community for purposes of an ABS regime even when its members would not be considered indigenous. The fact that Germany, Switzerland, Republic of Korea, Turkey, and Saudi Arabia have few or no recognized Indigenous peoples does not establish that they have no relevant local communities.

Given the extent of voluntary and non-voluntary migrations and relocations, reference to local communities is important. More than 10 million Africans were forcibly transported across the Atlantic during the transatlantic slave trade and would be excluded from a policy limited to indigenous peoples. The descendants of these enslaved people formed numerous culturally distinct communities in the Americas. And there is also considerable voluntary migration within Africa and to other continents.

Brazil is an example of a country that formally recognizes both Indigenous peoples and local communities. These include Indigenous peoples, who have pre-colonial Indigenous populations, Quilombolas, who are principally descendants of Africans brought during slavery, and Pomeranian communities, which are descendants of European immigrants. All are communities possessing culturally transmitted knowledge and practices involving biological resources.

The ILO/UN estimate of the global number of Indigenous peoples is just under 500 million persons, or roughly 6 percent of the world's population, and this is based on fairly arbitrary definitions of who is indigenous. There are no comparable definitions of how many people qualify as members of local communities, in the CBD sense. The World Resources Institute and the FAO both have estimated that up to 2.5 billion people depend on indigenous and community lands. This would imply that roughly 2 billion persons could be considered part of a local community that was not indigenous, but all of these estimates are imprecise and subject to nuances in definitions and context.

Local communities in South Africa

South Africa has a relatively small population generally identified internationally as Indigenous Peoples. The Khoi-San peoples have inhabited southern Africa for thousands of years and today represent no more than about one percent of South Africa's population.

Much of South Africa's Black population descends from Bantu-speaking populations whose ancestors migrated southward over a much longer historical period and subsequently developed distinct societies, including the ancestors of today's Zulu, Xhosa, Sotho, Tswana, Venda, Tsonga, Ndebele, Swazi and other peoples.

In South African usage, the Zulu, Xhosa, Sotho, Tswana, Venda, Tsonga, Ndebele and Swazi include communities that could be characterized as “local communities.” A 2025 South African government Indigenous Peoples Plan identifies and describes, for example, the Venda, Tsonga and Northern Sotho (Pedi) communities in Limpopo, as well as the Ndebele, Tsonga and Swazi communities in Mpumalanga, their centuries-long residence, distinctive languages, customs, traditions, agriculture and community-based ways of life. South Africa's Constitution likewise recognizes Sepedi, Sesotho, Setswana, siSwati, Tshivenda, Xitsonga, isiNdebele, isiXhosa and isiZulu as indigenous languages whose historically diminished use and status must be advanced by the state. Some South African government materials refer to the Khoisan, AmaXhosa, AmaZulu, AmaNdebele, AbeSotho and VhaVenda as indigenous peoples of South Africa. The South African Human Rights Commission has noted that Nguni, Sotho-Tswana, Venda and Tsonga speakers identify as indigenous, while using the narrower term “indigenous peoples” for Khoi and San peoples in the context of its inquiry.

More recent international migrations have also brought substantial numbers of people from other African countries to South Africa; approximately two million Black African immigrants were living in South Africa at the time of the 2022 census.

Permanent European settlement at the Cape began in 1652. Over the following centuries, Afrikaners developed distinctive cultural practices and bodies of traditional knowledge relating to agriculture and medicine, including knowledge concerning the medicinal uses of plants. B.-E. van Wyk documented the medicinal ethnobotany of the Cape and Karoo, examining Khoi-San and Cape Dutch traditions together and noting the difficulty, in some cases, of determining whether particular knowledge, uses and practices originated with the Khoi-San, the Cape Dutch, or developed through exchanges between the communities.⁴⁸

This history raises difficult questions concerning the beneficiaries of traditional knowledge protections. If traditional knowledge concerning a genetic resource has been held and transmitted for generations by both Afrikaner and Khoi-San communities, and its original source cannot be determined, does an ABS regime recognize both communities as holders entitled to participate in decisions concerning access and benefit sharing? Does it matter whether one community is recognized as Indigenous Peoples while the other might qualify only as a local community? If the same traditional knowledge is held by both communities, could it be treated as being in the public domain when obtained from the Afrikaner community but as protected traditional knowledge when obtained from the Khoi-San community? Or should both communities be regarded as holders of the same traditional knowledge? The answers depend, in part, upon how “local communities,” “holders,” “traditional knowledge,” and the “public domain” are defined, and whether protection depends upon the historical origin of the knowledge, its present holders, or both.

⁴⁸ Van Wyk, B. E., de Wet, H., & Van Heerden, F. R. (2008). An ethnobotanical survey of medicinal plants in the southeastern Karoo, South Africa. *South African Journal of Botany*, 74(4), 696-704. <https://doi.org/10.1016/j.sajb.2008.05.001>

Definitions of local communities

The current WIPO negotiating text on the Protection of Traditional Knowledge (WIPO/GRTKF/IC/52/4) has 43 references to local communities but no definition or guidance on what they are. The WIPO IGC has been negotiating for more than 25 years without agreeing on a definition. This is not unique to WIPO. In international law, a definition is not considered a prerequisite for protection. That said, the ambiguity over what constitutes a local community remains a practical problem in reaching consensus and implementing agreements.

The CBD Secretariat's position is that although there is no set definition of "local communities," the CBDy uses the term "indigenous and local communities" with reference to communities that have a long association with, and depend on, the lands and waters that they have traditionally lived on or used.⁴⁹ A 2011 Secretariat Guidance document on the issue states, in part:

5. Many communities may be considered local and may also be described as traditional communities. Some local communities may include peoples of indigenous descent. They are culturally diverse and occur on all inhabited continents. For example, small farming communities in France, who have occupied and farmed their lands for many generations acquiring useful environmental knowledge including specialist knowledge about a variety of activities including sustainable agriculture, cheese making and wine making or even animal husbandry may represent a local or traditional community. Long term established rice and fish farmers in Asia may represent another type of local community. In Brazil, local and traditional communities are an important segment of the national population and are a legal entity in the national constitution.
6. However, "Local community" remains, to some extent, an ambiguous term. It can refer to a group of people which have a legal personality and collective legal rights and this is considered a community in the strict sense. Alternatively, a "local community" can refer to a group of individuals with shared interests (but not collective rights) represented by a non-governmental community-based organization (NGO). For example, many traditional communities act through NGOs, which are social rather than community organizations.
7. The issue of cultural identity remains a multidimensional and complex issue.
8. Self-identification is the most appropriate way to establish who may be indigenous and/or a local and/or a traditional community representative

In 2011, the CBD convened an Expert Group Meeting of Local Community Representatives within the context of the CBD, which had, as one objective, to identify common characteristics of local communities. The report of the meeting included an Annex that provided a list of "common characteristics" of local communities, in the context of Article 8(j) of the CBD.⁵⁰

⁴⁹ Guidance For The Discussions Concerning Local Communities Within The Context Of The Convention On Biological Diversity. UNEP/CBD/AHEG/LCR/1/2, 7 July 2011.

⁵⁰ REPORT OF THE EXPERT GROUP MEETING OF LOCAL COMMUNITY REPRESENTATIVES WITHIN THE CONTEXT OF ARTICLE 8(j) AND RELATED PROVISIONS OF THE CONVENTION ON BIOLOGICAL DIVERSITY. September 4, 2011. UNEP/CBD/WG8J/7/8/Add.1
<https://www.cbd.int/doc/meetings/tk/wg8j-07/official/wg8j-07-08-add1-en.doc>

ADVICE AND RECOMMENDATIONS ARISING FROM THE EXPERT GROUP MEETING OF
LOCAL-COMMUNITY REPRESENTATIVES

The following recommendations were adopted by consensus and are transmitted for the consideration of the seventh meeting of the Working Group on Article 8(j) and Related Provisions:

I. COMMON CHARACTERISTICS

The experts recommended that a working definition may be possible based on the following characteristics, some of which could be considered essential. The group agreed that self-identification or the right to self-identify should be foremost and essential in any list of characteristics and that, because of the diversity of local communities, a possible list of characteristics should be broad and inclusive. Furthermore, from the list of possible characteristics below, a local community could possess a cluster of characteristics, reflecting its own unique cultural, ecological and social circumstances. The list below appears in no particular order of priority.

- (a) Self-identification as a local community;
- (b) Lifestyles linked to traditions associated with natural cycles (symbiotic relationships or dependence), the use of and dependence on biological resources and linked to the sustainable use of nature and biodiversity;
- (c) The community occupies a definable territory traditionally occupied and/or used, permanently or periodically. These territories are important for the maintenance of social, cultural, and economic aspects of the community;
- (d) Traditions (often referring to common history, culture, language, rituals, symbols and customs) and are dynamic and may evolve;
- (e) Technology/knowledge/innovations/practices associated with the sustainable use and conservation of biological resources;
- (f) Social cohesion and willingness to be represented as a local community;
- (g) Traditional knowledge transmitted from generation to generation including in oral form;
- (h) A set of social rules (e.g., that regulate land conflicts/sharing of benefits) and organizational-specific community/traditional/customary laws and institutions;
- (i) Expression of customary and/or collective rights;
- (j) Self-regulation by their customs and traditional forms of organization and institutions;
- (k) Performance and maintenance of economic activities traditionally, including for subsistence, sustainable development and/or survival;
- (l) Biological (including genetic) and cultural heritage (bio-cultural heritage);

- (m) Spiritual and cultural values of biodiversity and territories;
- (n) Culture, including traditional cultural expressions captured through local languages, highlighting common interest and values;
- (o) Sometimes marginalized from modern geopolitical systems and structures;
- (p) Biodiversity often incorporated into traditional place names;
- (q) Foods and food preparation systems and traditional medicines are closely connected to biodiversity/environment;
- (r) May have had little or no prior contact with other sectors of society resulting in distinctness or may choose to remain distinct;
- (s) Practice of traditional occupations and livelihoods;
- (t) May live in extended family, clan or tribal structures;
- (u) Belief and value systems, including spirituality, are often linked to biodiversity;
- (v) Shared common property over land and natural resources;
- (w) Traditional right holders to natural resources;
- (x) Vulnerability to outsiders and little concept of intellectual property rights.

This is a long list, and as presented, “common characteristics” is not the same as “characteristics in common.” Recognized or identified local communities do not need to have each characteristic, and self-identified appears to be the key requirement, given that none are described as essential.

The CBD Secretariat subsequently collected a compilation of views on the use of the term Indigenous Peoples and Local Communities, which it published in 2014 (UNEP/CBD/COP/12/INF/1).⁵¹

The government of Brazil made the following comment:

With regard to the use of the term "indigenous peoples and local communities", Brazil has no objection to the expression, since it is already used in our current national legislation. According to the Decree 6040/2006, which established the National Policy for the Sustainable Development of Traditional Peoples and Communities (PNPCT), “Traditional Peoples and Communities means culturally diverse groups who recognize themselves as such, have their own forms of social organization, occupy and use territories and natural

⁵¹ Compilation Of Views On Use Of The Term “Indigenous Peoples And Local Communities”, Note by the Executive Secretary, UNEP/CBD/COP/12/INF/1, June 26, 2014.
<https://www.cbd.int/doc/meetings/cop/cop-12/information/cop-12-inf-01-en.pdf>

resources as a condition for their cultural, social, religious, ancestral and economic reproduction, using knowledge, innovations and practices generated and transmitted by tradition”.

The Finnish government’s comment expressed concern of the ambiguity of the term indigenous peoples and local communities, recognizing the desire by many for a broad definition, and noting the lack of guidance from the earlier negotiating history of the CBD.

The term “indigenous peoples” is clearly established internationally (e.g. UNDRIP), as well as in Finnish national legislation. However, Finland would like to draw attention to the expression “local communities”. This term is not clearly defined, which may cause some problems, for instance, concerning to which local communities Article 8(j) is applied. . . . Finland understands that it is important to keep the definition broad because of the globally great diversity of human communities and their traditional knowledge. However, Finland would like to continue the discussion and encourages experts and representatives of local communities, if possible, to further clarify the term. Finland thinks that it would also be useful to study the history of the Convention concerning this particular question. The negotiations that took place over 20 years ago, which led to the Convention, are documented and published, but it is not clearly documented how the wording of Article 8(j) actually developed and what was the meaning of the term “local communities” at the time.

Bolivia’s comment endorsed the term “Indigenous peoples” and focused on the importance of recognizing their rights. A translation of their comment states that “While we welcome the mandate to continue considering a change in terminology, we are concerned by the lack of progress on this issue and by attempts to regress—specifically by referring to these groups as “Indigenous communities” within this Convention on Biological Diversity--thereby disregarding the decades-long struggle of Indigenous peoples to gain recognition as such in the multilateral arena.”

In addition to the 1992 CBD and the 2010 Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from Their Utilization, the term ‘local communities’ is found in several other multilateral treaties, including:

- United Nations Convention to Combat Desertification (UNCCD) (1994)
- International Treaty on Plant Genetic Resources for Food and Agriculture (2001)
- United Nations Framework Convention on Climate Change (UNFCCC) and the Paris Agreement (2015)
- Agreement under the United Nations Convention on the Law of the Sea on the Conservation and Sustainable Use of Marine Biological Diversity of Areas Beyond National Jurisdiction (BBNJ) (2023)
- WIPO Treaty on Intellectual Property, Genetic Resources and Associated Traditional Knowledge (WIPO GRATK treaty) (2024)
- WHO Pandemic Agreement (2025)

None of these agreements define local communities, deferring to national laws.

Finally, it is worth noting that the WHO Global Traditional Medicine Strategy, adopted in 2025 by the WHA, has 17 references to Indigenous peoples, and no references to local communities.⁵²

Self-identified

The CBD COP 16 Cali Fund decision describes "self-identified needs", or institutions "identified by Indigenous peoples and local communities." The diversity of communities within countries presents significant practical hurdles in doing this.

A state may (or may not) have a functioning indigenous-consultation apparatus (free, prior and informed consent processes, land-title registries, recognized councils), and even then, the state may have no comparable apparatus for local communities that don't fit indigenous criteria. A Cali Fund disbursement mechanism will then likely default to whatever national indigenous-recognition apparatus exists (because that's the only apparatus on hand), while non-indigenous communities may be excluded by administrative path-dependence.

In some countries, it will not be practical or possible to present anything that represents a consensus or approved position on funding needs.

Decision 16/2 has a "subject to national circumstances" qualifier. The Cali Fund mechanism could produce disbursements that flow relatively smoothly and visibly to indigenous communities in countries with strong pre-existing federations (e.g., Ecuador, Colombia, Mexico, the Philippines, and Indonesia-via-AMAN), while countries with the more fragmented, least-institutionalized indigenous populations or local communities, will have funds absorbed into general government administration, if there is no other entity capable of receiving it.

The Ad Hoc Technical Expert Group's country-level allocation formula (biodiversity richness, geographic origin, capacity needs) will decide how much money a country gets, but does not address the sub-national problem of who within that country actually receives and directs the funding.

Possible approach to benefit sharing that anticipates TMK used to train GenAI services

As noted above, the WHO policy objectives for traditional medicine include the development of evidence-based policies and the use of research to improve health care outcomes, and the need to recognize the rights of Indigenous peoples. The use of TMK to train AI services will deepen our

⁵² Global traditional medicine strategy 2025-2034, WHO, October 30, 2025.
<https://www.who.int/publications/i/item/9789240113176>

understanding of how traditional medicine works, its benefits, limitations and risks, and provide insight into the development of new medical products.

Much traditional medicine knowledge is effectively in the public domain, widely diffused and without obvious holders of sui generis intellectual property rights. There is also a need to document many practices before knowledge is lost.

The dilemma facing researchers, from the non-profit, commercial and technology sectors, is to resolve the existing or proposed set of intellectual property rules regarding the collection, dissemination and use of TMK.

There is also considerable debate over how ABS systems should be designed, given the expanding and important use of digital research tools and GenAI services.

EU, China and California rules for transparency of training data

Given the speed at which GenAI services are developing, there will be a short window of time before a significant corpus of TMK is incorporated into GenAI services. Policies by database managers and GenAI services to provide transparency and traceability of the dissemination and uses of TMK to train GenAI services are important. Norms on this include evolving standards and government policies, including for example:

- The IEEE Standard 2890-2025, on the Recommended Practice for Provenance of Indigenous Peoples' Data;
- The European Union AI Act, which requires, in Article 53(1)(d), a "sufficiently detailed public summary of the content used for training" plus confidential technical documentation for regulators, including a Model Documentation Form for the AI Office that must include provenance, source, composition of training data, and a less detailed public summary using the Commission template published;
- The State of California Generative Artificial Intelligence Training Data Transparency Act (AB 2013), which came into effect on January 1, 2026;
- Various proposals before the U.S. Congress, such as the AI Foundation Model Transparency Act; and
- China's Interim Measures for Generative AI Services, which provides that the Cyberspace Administration of China (CAC) can request information on the "descriptions of the source, scale, type, quality, etc." of such Training Data.⁵³

In these three jurisdictions (the EU, California and China), AI laws almost entirely focus on copyright and trade secrets. For example,

⁵³ Mayer Brown, Gen AI: China Proposes Draft Measures For Regulating Generative Artificial Intelligence, Mondaq. June 9, 2023.
<https://www.mondaq.com/hongkong/new-technology/1327608/gen-ai-china-proposes-draft-measures-for-regulating-generative-artificial-intelligence>

- Article 53(1)(c) of the **EU AI Act** specifically refers to EU copyright law and its recital 107 on training data which states that “While taking into due account the need to protect trade secrets and confidential business information, this summary should be generally comprehensive... to facilitate parties with legitimate interests, including copyright holders, to exercise and enforce their rights.”
- **California AB 2013** requires disclosure on “Whether the datasets include any data protected by copyright, trademark, or patent, or whether the datasets are entirely in the public domain” with no mention of rights in traditional knowledge.
- The **Chinese** Interim Measures references intellectual property rights, in practice, copyright, trademark and patent rights, and does not recognize sui generis rights in TK or genetic resources.

The EU database *sui generis* right and regulatory data protection for drug or agricultural products are not treated as distinct intellectual property categories.

Copyright and trade secrets, as well as confidential business information are areas where there are legal norms controlling access to knowledge. Patents, on the other hand, are designed to disseminate information, and patent grants and patent applications, would generally have no restrictions on their use to train AI services.

The existing legal architecture for GenAI data transparency relies almost entirely on copyright, trade secrets, and statutory regulatory data protections. These tools are fundamentally mismatched with the nature of TMK, leaving a regulatory vacuum during a rapid window of AI ingestion.

Standards like IEEE 2890-2025, alongside the CARE Principles for Indigenous Data Governance (Collective Benefit, Authority to Control, Responsibility, Ethics), remain voluntary technical benchmarks or soft law. Without explicit statutory cross-references in binding AI transparency laws, GenAI developers will lack legal incentives to track or disclose TMK provenance.

Specific policy mechanisms could be integrated into evolving AI governance:

- **Expanding Mandatory AI Data Summaries:** Amending templates (such as the EU AI Office's model form or state-level disclosure mandates) to explicitly require disclosure of datasets sourced from ethnobotanical databases, traditional knowledge registries, or Indigenous lands, independent of copyright status.
- **Integrating Soft-Law Standards into Hard-Law:** Conditioning regulatory approvals or government procurement of medical AI tools on adherence to standards like IEEE 2890-2025 when processing traditional health and biological data.

- **Consistency with the WIPO GRATK treaty:** Government AI transparency requirements should be consistent with and supportive of the obligations and objectives of the WIPO GRATK treaty.

Formalities

The Swakopmund Protocol on the Protection of Traditional Knowledge and Expressions provides that the protection of traditional knowledge shall not be subject to any formality.....

Check on other regimes, and discuss implications.

Brazil's Biodiversity Law (Law No. 13,123/2015)

As noted above, the legislative changes in Brazil have shifted the emphasis from time consuming obligations to establish consent to use TK to a combination of disclosing research activities and enforcing benefit sharing obligations at the point of commercialization, particularly for TK that already has wide diffusion of use.⁵⁴ Researchers no longer need advanced government licenses to conduct initial studies or training data aggregation on genetic heritage or associated TK.

Research activities are logged in an online database, the National System for the Management of Genetic Heritage and Associated Traditional Knowledge (SisGen), which issues an automatic compliance registration. The Act recognizes that much TK is widely diffused across multiple communities or published in historical literature, making it impossible to identify a single, specific Indigenous origin. When TK is widely diffused or of unidentifiable origin, user institutions are exempt from seeking individual Prior Informed Consent (PIC). However, the commercial use of that knowledge is not exempt from benefit-sharing.

The benefit-sharing obligations are deferred to the point of commercial exploitation of the final product. Intermediate research, basic science, and pre-commercial R&D are not required to make monetary contributions, while finished commercial products incorporating TK are required to pay royalties on annual net revenue, from 1 to 0.1 percent.⁵⁵

Indigenous Peoples, Local Communities

As noted above, local communities are a separate legal category from Indigenous Peoples and may have the same or different rights, under various international or national legal systems, as regards TK.

⁵⁴ Brazil's Biodiversity Law (Law No. 13,123/2015).

⁵⁵ James R. Welch, Brazil's New Biodiversity Law, *Ethnobiology Letters* 6(1):216-217, December 2015, DOI: 10.14237/ebi.6.1.2015.562.

Proposal for downstream benefit sharing obligations

Among the profound issues regarding ABS regimes for traditional knowledge is the fact that *sui generis rights* for GR or TK are sometimes fashioned as perpetual rights, which, when implemented as exclusive rights and associated with a medical product, could result in never ending monopolies on products. At the same time, if benefit sharing is tied to upstream uses of knowledge to train AI programs, the result could be a one-time exhaustion of rights, once the AI is trained using the inputs, without regard to the value of the knowledge, for specific products or uses, and ending any benefit to future generations.

As discussed above, while there is considerable effort to implement Persistent Unique Identifiers (PIDs) and other methods of tracking data stored in digital format, there are technical and practical reasons why these efforts will have limited success in establishing connections between AI inputs and outputs, regardless of the regulatory or database manager policies on transparency and tracking.

The WIPO GRATK treaty places obligations on patent holders to disclose the country of origin or the source of genetic resources when a claimed invention in a patent application is based on that resource. And for traditional knowledge associated with that genetic resource, the Indigenous Peoples or local community who provided the traditional knowledge, resources, or if not known, the source of the traditional knowledge.

There are two proposals relating to benefit sharing that can be implemented that would benefit from tracking upstream uses of knowledge, but would not depend on it, and would also address concerns over endless monopolies and intergenerational equity.

One of the proposals is to enable holders of rights in GR or TK to trigger a time-limited obligation on developers of specific products. The second is the use of a jury system to allocate shares of a modest royalty for the use of knowledge, data, materials or technology that is not protected by patents and that was useful in the development of a medical product.

Access plans

The WHO PIP Framework and the WHO Pandemic Agreement both provide opportunities for manufacturers to satisfy benefit sharing obligations with a variety of donations, reduced prices, technology transfer or or access and equity measures. The US NIH now has a policy that requires that access plans be attached to the exclusive licensing of NIH owned patented inventions⁵⁶, and there have been efforts to more systematically require access plans for biomedical products in the

⁵⁶ NIH Intramural Research Program Access Planning Policy Implementation Guidance, NOT-OD-25-137. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-137.html>.

European Union⁵⁷, although recent efforts to require access plans be included in Horizon Europe funding proposals for biomedical research were rejected by the European Parliament.

Time-limited obligations for benefit sharing

The Bellagio compulsory cross-licensing proposal

As in previous sessions of the WIPO IGC negotiations on *sui generis* regimes for genetic resources or traditional knowledge, CPTech had proposed a system of mandatory cross-licensing *sui generis* rights in GR to TK with patent rights.⁵⁸ The proposal borrowed from a European Union Directive on the legal protection of biotechnology inventions, which created a mandatory cross licensing regime between *sui generis* plant breeder rights and patent rights. CPTech proposed to design the *sui generis* GR or TK rights as perpetual, but only to have effect when the GR or the TK was associated with a product or service protected by a patent. The benefits from the cross-license, which would be royalty bearing, would be temporary. In the sense that the royalty would end when the patent expired, but the holders of the *sui generis* right would have perpetual rights to obtain cross-licenses to subsequent products under similar terms. This proposal also provided the holders of the GR or TK rights the ability to use the patented invention.

As has been the case under the EU Biotechnology Directive, parties would have the option of negotiating voluntary licensing agreements without involving the mandatory cross-license. The cross-licensing approach is similar to the provisions in Article 31(I) of the TRIPS Agreement, concerning dependent patents, which involves compulsory cross-licenses of patents, when “the invention claimed in the second patent shall involve an important technical advance of considerable economic significance in relation to the invention claimed in the first patent,” and when “the second patent’ cannot be exploited without infringing another patent ‘the first patent.’”

Some existing ABS regimes have time limits on benefit share:

The Swakopmund Protocol

The Swakopmund Protocol on the Protection of Traditional Knowledge and Expressions of Folklore within the Framework of the African Regional Intellectual Property Organization (ARIPO) makes a distinction between traditional knowledge that is transmitted in an intergenerational context and associated with a local or traditional community, and traditional

⁵⁷ Press Release. “More than 50 organizations urge MEPs to back access provisions in the EU’s next research programme,” DNDi.Org. <https://dndi.org/press-releases/2026/more-than-50-organizations-urge-meps-to-back-access-provisions-in-the-eus-next-research-programme/>

⁵⁸ The proposal is described in this presentation. James Love and Manon Röss, Access to Genetic Resources & Traditional Knowledge, The Bellagio compulsory cross-licensing proposal for benefit sharing consistent with more competition and openness. Consumer Project on Technology, 14 December 2002, Distributed at the Roundtable on Intellectual Property, Traditional Knowledge and Folklore: Sponsored by the Science and Human Rights Program of the American Association for the Advancement of Science (AAAS) and Franciscans International. <https://web.archive.org/web/20260314202428/https://www.keionline.org/wp-content/uploads/bellagio.cross-lic-ense-gr-tk-patent-14dec2002.pdf>

knowledge that “belongs exclusively to a person.” For the latter, the term of protection for TK is 25 years.

Peru: Law No. 27811 (Collective Knowledge Regime)

Peru’s legal regime for the protection of indigenous collective knowledge associated with biological resources includes a specific temporal boundary regarding public domain TK. Under Article 13, if collective knowledge enters the public domain, users are required to pay a percentage of gross sales into the Fund for the Development of Indigenous Peoples, but only if the knowledge entered the public domain within the previous 20 years. Once that 20-year window expires, the benefit-sharing obligation terminates.

South Africa and Kenya: Fixed Post-Commercialization Terms

Standard ABS contract clauses in countries such as South Africa (under NEMBA Bioprospecting Regulations) and Kenya (NEMA ABS Regulations) frequently have fixed benefit-sharing windows (e.g., 5, 10, or 15 years from the date of first commercial sale), after which the royalty obligation lapses regardless of ongoing sales. These are the outcome of negotiations.

Other regimes take a different approach. In Brazil, for example, the benefit-sharing obligations involving royalties on products derived from associated traditional knowledge are tied to the commercial life of the finished product and, unlike patents, copyrights or *sui generis* plant breeder rights, do not have a time-limited term.

The Open Source Dividend Proposal

The Open Source Dividend (OSD) is an incentive mechanism designed to reward researchers, institutions, or entities that openly share knowledge, data, biological materials, or technology on a royalty-free, non-discriminatory basis during the R&D process.

The OSD traces back to 2007, developed during Médecins Sans Frontières' work on innovation inducement prizes for low-cost tuberculosis diagnostics. One concern was that a prize reward to induce investments in R&D, when implemented as an alternative to the temporary monopoly typically associated with patents, could perversely encourage secrecy. The proposal was to set aside some of the prize money to reward any person, community or entity that voluntarily shared knowledge, data, materials, or technology that meaningfully contributed to the development of the winning result, regardless of who actually won.

But this was not the only motivation. The concern is that researchers have financial incentives to make research secret or to license in a restrictive manner, and no financial incentives to share information freely. While for society, knowledge shared more widely has a higher value. One Nobel prize-winning scientist, Sir John Sulston, once conjectured that medical knowledge in the public domain was nine times more valuable to society than knowledge held in secret. The open source

dividend proposal was seen as correcting a flaw in the current system, which only rewards secrecy and restrictive licensing of knowledge and not open sharing.

This OSD proposal was subsequently included in a number of proposals considered by various UN bodies, including the WHO, WIPO and the United Nations Secretary-General's High-Level Panel on Access to Medicines, and in a number of proposals in the US Congress. For example, S.1801, in the 116th Congress, included an open source dividend component of a proposed prize for the development of new antibiotic drugs, which proposed to "reward persons that openly shared on a royalty-free, not-for-profit and non-discriminatory basis, materials, technology, data, and knowledge that contribute in a significant way to the successful development of a qualifying product or significantly advanced the field of antibiotic research."⁵⁹ For a more general approach, see S.495, 115th Congress.

Proponents of the open source dividend have suggested that it can be implemented outside the prize system, as a general obligation on all drug development, by requiring a royalty on each product, during the period when a product benefits from patent protection, or for a fixed term, like 5 to 10 years, for a new product, in order to induce more sharing of knowledge, materials, technology and data.

The allocation of the dividend to those sharing knowledge, materials, technology and data was proposed through a jury system. Once a product was introduced into the market, there would be a time-limited period of nominations for persons or communities whose contributions were considered useful, including evidence to support such nominations and descriptions of the significance of the contribution. The reference to communities was deliberate, to be inclusive of communities associated with traditional medical knowledge, but also groups collaborating on open science, software or technology, and the institutions that supported their work.

The jury of experts would be appointed randomly from a large pool of candidates, identified by professional societies or other trusted actors. The jury of experts would review applications and divide the royalty, based upon transparent criteria.

The implementation of the open source dividend for drugs, vaccines, gene or cell therapies, or diagnostic devices was designed to benefit a broad range of knowledge, material, technology or data inputs, and not just TMK. As such, a fixed open source royalty, from 1 to 5 percent, made sense. An open source dividend of 1 percent on US patent protected drug sales in 2025 would have likely yielded more than \$5 billion in revenue, and if Europe and the rest of the world followed, this would be larger. For just the US and Europe, a 2 percent open source dividend would have generated perhaps \$16 billion in 2025. This money would be distributed to researchers, communities, and businesses that open sourced their knowledge, materials, data or technology, an amount that would have an impact and expand access to important inputs towards innovations. Companies paying into the open source dividend would also receive a benefit from the system, because it would reduce their costs of licensing or buying intellectual property rights.

⁵⁹ SEC. 409K. Innovation incentive fund for new and more effective treatments of bacterial infections.

How might an open source dividend be applied to TMK? It is an appealing approach for an ABS system that does not depend upon contracts with researchers, restrictive access to knowledge, or the technological tracing of the access, use and dissemination of knowledge. However, if done as a separate dividend for TMK only, a one size fits all royalty obligation on every product would not be appropriate. The Cali Fund is attempting to obtain ABS monetary payments from companies in this way, but without success so far.

The approach in Brazil is the more appropriate approach.

First, a determination is made based on whether an economic exploitation of a finished product or reproductive material, and the accessed genetic heritage or TK, added significant value to that product. Meaning the genetic heritage or TK either determined the product's functional characteristics or its market appeal.

Second, a benefit sharing obligation is attached, which has the following structure.

For genetic heritage only, with no TK involved, the obligation is a royalty of 1 percent of net revenue. Article 21 of the Brazilian law allows the federal government, at industry's request, to negotiate sectoral agreements that reduce the genetic heritage benefit-sharing rate down to a floor of 0.1%.

TK of unidentified origin (no traceable specific community), has the same structure as the GR case, of 1 to 0.1 percent of revenue (Article 23).

When there is TK of identifiable origin, Article 24 paragraph 1 requires the amount to be freely negotiated between the user and the specific identified TK provider/community, subject to general fairness and transparency standards. In addition to that negotiated agreement, Article 24 paragraphs 2-5 require a separate, mandatory deposit into the National Fund for Benefit-Sharing (FNRB), specifically to cover other, unnamed co-holders of the same knowledge, since the law presumes co-holders always exist even if only one community is party to the negotiated agreement. That deposit is set at half of the rate listed in Article 20, or 0.5 percent of net revenue by default, or half of of the sectoral-agreement-reduced rate (e.g., 0.05 percent if reduced to 0.1 percent) — and this amount is owed regardless of how many co-holders actually exist (paragraph 4).

Brazil illustrates a non-adjudicated, low-overhead alternative to jury-style apportionment, at the cost of not distinguishing the magnitude of the relevance of genetic resources or TK.

Recommended practice for downstream benefit sharing

There is a limited amount of leverage that can be attached to access to databases or knowledge in general. Additionally, there are significant and well documented challenges in tracing, proving and justifying links between access to knowledge inputs and commercial outputs.

Governments can play an important role in enforcing ABS obligations downstream, particularly when products are sold.

Time-limited ABS obligations on specific products or services are recommended, so that a perpetual *sui generis* right in GR to TK is not used to create a permanent legal monopoly.

The Open Source Dividend approach, for all openly shared knowledge, material, technology and data inputs, is a preferred benefit sharing model. Specifically for drugs and other high value products, the Open Source Dividend jury system is appealing. The Brazilian approach, modified with time limits on benefit sharing for specific products, is appropriate for a GR and TK only regime, and when a system that is simpler and less costly to administer is preferred.

Governments could also require access plans be provided for any product for which TK added significant value, for the benefit of communities that were the source of the TK.