INTERNATIONAL JOURNAL OF JUDICIAL LAW

The Access Boundary, not a Complete Legal Shield: Traditional Cultural Expressions, Copyright and Institutional Control of AI Training Data

Prajwal Bhattarai
Comparative Legal Researcher, International Policy Consultant, and Regulatory Governance Advisor, Nepal

* Corresponding Author: **Prajwal Bhattarai**

Article Info

ISSN (online): 2583-6536
Volume: 05
Issue: 06
Received: 10-09-2026
Accepted: 08-10-2026
Published: 06-11-2026
Page No: 48-56

Abstract
Traditional cultural expressions used in generative-AI training raise distinct questions about copyright in recorded expression, lawfulness of acquiring training copies and the terms on which archives publish and license collections. This article argues that institutional release decisions are an important but incomplete governance lever, not a substitute for intellectual-property rights or a legal prohibition on independent scraping. The 2024 WIPO Treaty on Intellectual Property, Genetic Resources and Associated Traditional Knowledge concerns patent disclosure rather than AI training. In *Bartz v Anthropic* and *Kadrey v Meta*, United States district courts reached fact-specific fair-use holdings concerning literary training data; neither judgment creates a general immunity for training on traditional expressions. *Bartz* also distinguished training use from a retained pirated-book library, although that litigation subsequently ended in a court-approved settlement and dismissal that did not decide the merits. New Zealand's Te Papa disclosed established Māori consultation and collection-access practices, alongside reported absences of AI-specific risk assessments and contract clauses. Nepal's Copyright Act expressly excludes folk songs, folklore and proverbs as such from copyright while protecting some original derivative expressions. The article proposes differentiated consent, contract, technical and statutory safeguards, with attention to their enforceability and limits.

DOI: https://doi.org/10.54660/IJJL.2026.5.6.48-56

Keywords: traditional cultural expressions, generative AI, copyright, cultural heritage governance, Indigenous data sovereignty, collection licensing

1. Introduction: Three Different Legal Questions, Not One Gate

A national archive holds recordings of traditional song, textile patterns and oral narrative, collected over decades and digitised with donor funding. A developer assembling a training corpus takes what the archive has published. A model is trained. It subsequently generates, on demand and at negligible cost, material in the idiom of that tradition which circulates commercially without attribution and competes with the output of the communities from which the originals came.

The central question is not whether intellectual property can never matter, but which legal lever addresses which act. Three questions need to be kept apart. The first is what copyright and related international instruments protect and which exceptions apply. The second is the lawfulness of the routes by which training copies are obtained which acquisition law and contract law address. The third is how institutional release decisions, technical interfaces and community protocols govern release and some downstream users, a reach that depends on notice, assent, rights, technology and applicable law. These distinct levers should not be collapsed into a universal 'access term'.

The WIPO Treaty on Intellectual Property, Genetic Resources and Associated Traditional Knowledge, adopted on 24 May 2024, deals with disclosure by patent applicants when a claimed invention is based on genetic resources or associated traditional

knowledge. Its patent trigger does not govern the assembly of a generative-AI training corpus without a relevant patent application (World Intellectual Property Organization [WIPO], 2024) [10].

Two June 2025 United States district-court rulings addressed fair use on records concerning books used for model training. In *Kadrey* the court stressed the inadequacy of proof about market dilution; in *Bartz* the court distinguished transformative training from retaining a library assembled using pirated books. Neither case adjudicated ownership of traditional cultural expressions, all AI training methods or the application of another country's law (*Bartz v. Anthropic PBC*, 2025; *Kadrey v. Meta Platforms*, Inc., 2025) [1, 6].

Alongside these sits the emerging practice of Indigenous data governance, expressed in institutional protocols and in principles developed by Indigenous scholars and organisations. It takes the problem seriously and it is the body of practice this article identifies as operative at the institutional layer. But its operation is not what the protocol documents alone suggest and the demonstration of that is the article's main documentary evidence.

In *Bartz* the acquisition history mattered because the court separated the fair-use analysis of training from reproductions retained as a general-purpose library. That does not mean a museum may impose a legally binding condition on every person who sees a publicly accessible image. It means that provenance, copying, access credentials and the terms of any actual licence may create issues distinct from training use; contractual obligations depend on their legal formation and scope. On 20 July 2026 the court granted final approval of a class settlement and dismissed the action with prejudice, so the piracy issue is no longer an impending trial outcome; that order does not decide the merits (*Bartz v. Anthropic PBC*, 2025, 2026) [1, 2].

An official disclosure from New Zealand's national museum, dated 26 March 2026 (Museum of New Zealand Te Papa Tongarewa, 2026) [9], allows that proposition to be tested against an institution with one of the more developed consent architectures. What it shows is that the architecture governs provenanced taonga and their display and use, while the material actually available to a corpus assembler is determined by the publication decision and the collections interface. In answer to the information request, the museum reported that it holds no assessment that specifically evaluates AI scraping, third-party AI services or digital colonialism as threats to mātauranga Māori, that its Microsoft agreements contain no clause explicitly referencing Te Tiriti o Waitangi or Māori data governance and that it describes no technical measure specific to preventing the scraping of mātauranga Māori, while referring to general hosting and application safeguards and to single sign-on for non-open data. The protocol is real and the reported exposure sits outside it.

The study employs doctrinal and institutional comparison rather than empirical analysis of training corpora, models, archive catalogues or Nepalese institutional practice. It does not measure the prevalence of traditional material in training datasets or the economic effects of model outputs on cultural communities. The opening archive scenario is illustrative. Nepal's institutional position has a narrower documentary foundation than the United States and New Zealand comparisons.

The material is selected as a functional stress-test of distinct legal levers, not as a representative country sample or a comparison of equivalent national systems. United States judgments illuminate the copyright and acquisition questions; New Zealand's Te Papa disclosure illustrates institutional decisions about collection access; and Nepal supplies the statutory and policy setting in which safeguards might be evaluated. The WIPO treaty and the European Union text-and-data-mining rule supply further contrasts at the instrument level. Because these sources operate at different institutional layers, the comparison tests the reach and limits of each lever rather than treating the jurisdictions as matched observations or ranking their actual protection of cultural communities.

2. Why Copyright and Institutional Governance Have Different Jobs

Copyright and institutional governance do different jobs in four respects. The first is the scope of protectable subject matter. Copyright is an incomplete tool for protecting longstanding community traditions, but describing it as structurally incapable of ever applying is too broad. Original recordings, arrangements, translations, adaptations, collections and other protected subject matter require separate analysis under the applicable domestic statute.

Nepal's Copyright Act, 2059 (2002), section 4, excludes folk songs, folk tales and proverbs as such from copyright protection; section 3(2), however, protects original derivative works based on folk expression, together with original translations, arrangements and collections, without prejudicing rights in an underlying work. The distinction makes it unsafe to treat all recordings, new arrangements or derivative expressions as unprotected. The analysis relies on the English extract of sections 3 and 4 published in WIPO Lex; the Nepali text published by the Nepal Law Commission was not independently reviewed and the analysis does not resolve the effect of amendments or institutional rules (Copyright Act 2059 (Nepal), ss. 3–4).

The second is the identity of the right-holder. Where a traditional expression has been recorded, the copyright that exists may sit in the recording and may belong to the recordist, performer, producer, employer or institution rather than to the community, depending on the applicable law, employment terms, licences and contracts. Where that is so, the practical consequence can be perverse: the only enforceable right in the material may be held by the party whose release decision created the exposure.

The third is the character of the harm. The injury that may follow from mass generation of imitative material is not that a particular work has been copied but that the tradition's distinctiveness is diluted and the market for authentic community production is displaced by cheap substitutes. That is a hypothesis about harm, not an empirical finding of this study. Infringement through the copying of training inputs and alleged infringement through generated outputs are separate questions: the first engages the reproduction right and the applicable exceptions, while the second asks about similarity to a specific protected work. Neither captures such harm: a model trained on a thousand textile patterns and generating a thousand-and-first may reproduce none of them.

The fourth is the direction of the new international instrument which is where the WIPO treaty belongs in the analysis. Its disclosure duty is a real advance and its recognition of the providing community is precisely the conceptual move this field needs. But it attaches to patent applications, where there is an applicant, a claim and an examination process at which

a disclosure can be required. Model training offers none of those hooks: there is no application, no claim and no examiner. The instrument's architecture presupposes a gatekeeping moment that generative AI development does not contain.

The resulting legal position is plural: copyright may reach original protectable expression; moral, cultural and community interests can be wider; contractual and technical controls can govern some distribution channels. None of these devices is automatically effective against all downstream actors and their limits are part of the policy design rather than reasons to discard them.

Two distinct bodies of scholarship sharpen the boundary between legal permission and governance legitimacy. Lemley and Casey (2021) [8] analyse the fair-use implications of machine learning, providing a doctrinal account that predates the record-specific *Bartz* and *Kadrey* decisions. Carroll *et al*. (2020) [3] articulate the CARE Principles for Indigenous Data Governance: collective benefit, authority to control, responsibility and ethics. Their framework concerns Indigenous communities' interests in data stewardship and secondary use; it is not itself a copyright statute or a universally enforceable ban on model training. Kukutai and Taylor (2016) [7] provide an edited collection on Indigenous data sovereignty that situates such principles within a wider agenda. The present article brings these discussions together by identifying which obligations copyright, contract and institutional access controls could respectively impose. It does not assume that a museum's unilateral notice binds a scraper who has not assented to its terms.

3. What the 2025 Copyright Decisions Actually Decided

The relevant decisions are *Bartz v Anthropic PBC*, No 3:24-cv-05417 (ND Cal, order of 23 June 2025) [1] and *Kadrey v Meta Platforms Inc*, No 3:23-cv-03417 (ND Cal, order of 25 June 2025) [1]. They are first-instance US rulings on fact-specific summary-judgment records, not globally applicable interpretations of copyright law.

In the first, decided on 23 June 2025, the court held that training the defendant's models on books was exceedingly transformative and therefore fair use and that digitising print books the defendant had purchased was also fair use because the scanned and compressed copies replaced the originals. It held, however, that the defendant had no entitlement to use pirated copies for its central library and that building a permanent general-purpose library did not excuse the piracy. A trial was left to be held on the pirated copies used to build that library and on resulting damages, actual or statutory, including for willfulness. The plaintiffs did not allege that any output infringed their works, so the ruling concerns inputs (training and library copying) and not output-based claims.

In the second, decided on 25 June 2025, the court granted summary judgment to the defendant on the record before it, holding the training of its model to be fair use while describing the ruling as limited and case-specific. The first statutory factor favoured the defendant because the training was highly transformative; the second favoured the authors because their books are highly expressive; the third favoured the defendant; and the fourth, market harm which the court treated as the most important, also favoured the defendant. The model had been trained on works downloaded from shadow libraries. The court considered and rejected three theories of market harm: regurgitation, on the ground that the record showed no more than about fifty words from any book even under adversarial prompting; loss of an AI licensing market which it held irrelevant to the factor, lost licensing fees not being cognisable harm; and market dilution which it rejected on this record as too weakly argued, having not been raised in the complaint or the summary judgment motion, while indicating that a properly developed dilution argument could have succeeded given the technology's capacity to generate large volumes of secondary works. The court also wrote that in many circumstances it will be illegal to copy copyright-protected works to train generative AI models without permission, while stressing that its ruling did not hold Meta's use lawful in general.

Three propositions for this article follow and they are the reason the decisions are discussed at length.

First, a fair-use defence succeeded on the specific training records before these courts. It cannot be extrapolated as a categorical rule that all training on traditional expressions is lawful. Copyrightable output, acquisition practices, different markets and the applicable jurisdiction may yield different outcomes; *Kadrey* expressly left room for evidence of market substitution or dilution that the plaintiffs did not present.

Second, the lawfulness of acquiring and retaining material is analytically separate from a fair-use defence concerning particular training activity. In *Bartz*, the fair-use ruling did not protect a retained library of pirated books; on 20 July 2026 the court granted final approval of a class settlement and dismissed the action with prejudice, so the remaining claims were not tried and that order does not decide the merits or fair use. The proceedings support careful attention to acquisition and retention but do not establish that an archive's posted conditions automatically bind a third party who has not assented to them (*Bartz v. Anthropic PBC*, 2025, 2026) [1, 2].

Third, market dilution was discussed as a potentially important evidentiary issue in *Kadrey*, not judicially established as an actionable claim for cultural harm. Claims about displacement of community cultural producers require separate statutory analysis and evidence that this article has not collected.

4. Research Method and Source Selection

The method is layered rather than comparative in the conventional sense. Three legal layers are examined for their capacity to control the use of traditional expressions in training: an international intellectual property instrument, domestic copyright doctrine and institutional access governance. The jurisdictions enter as the sources of the best available material on each layer rather than as parallel systems being compared.

The documentary units are a copyright rule, a judicial holding on a defined litigation record, an international instrument or an institutional statement about collection governance. Their functions are compared by asking who owns a relevant right which act is regulated, against whom an obligation can be enforced and where the legal boundary ends. The study does not inspect AI training corpora, test the effectiveness of access controls or survey archives. A museum's reported practices and the courts' fact-specific decisions are evidence about those sources, not proof of general institutional behavior or measurable cultural-market harm.

The United States supplies the principal judicial evidence on copyright and AI training. New Zealand supplies the institutional-governance example through published information about Te Papa's collection practices. Canadian

and international Indigenous data-governance scholarship informs the principles discussed in section six, but a parallel institutional case study of Canada is not undertaken. Nepal is the principal setting for the policy recommendations, although evidence concerning its institutional practices is limited. The comparison is therefore functional across different legal layers rather than a claim that each jurisdiction provides directly equivalent data.

The European Union provides a further legal-design comparison. Article 4 of Directive (EU) 2019/790 creates a text-and-data-mining exception for lawfully accessible works when rights have not been reserved in an appropriate manner, including machine-readable means for material made public online. The rule is implemented through national law in each Member State and it is neither a global Indigenous cultural-rights regime nor automatically applicable in Nepal. Japan's copyright training exception and Australian Indigenous collections protocols remain useful extensions not examined here.

The documentary sources include the available United States district-court decisions, the July 2026 order approving the settlement, WIPO treaty materials, the European Union's text-and-data-mining exception and the WIPO Lex extract of Nepal's Copyright Act. Te Papa's published institutional response supplies a practice-based account, read here as a statement about the records the institution holds, not as an independent assessment of its effectiveness. No judgment establishing enforcement of these particular museum access terms against an independent AI developer was identified. The study did not undertake a survey of Nepali collecting institutions or establish every relevant aspect of Nepal's current treaty position. The resulting recommendations concern possible legal and technical safeguards rather than demonstrated enforcement outcomes.

5. The International Instrument and the Gatekeeping Problem

The WIPO treaty adopted on 24 May 2024 deserves examination in its own right, because its design reveals why intellectual property instruments struggle with this subject.

Its central obligation is a disclosure duty. Where a claimed invention is based on genetic resources, a patent applicant must disclose the country of origin or source; where it is based on associated traditional knowledge, the applicant must disclose the Indigenous Peoples or local community that provided it. Where the information is unknown, the applicant must disclose the source and where nothing is known, must declare so. Patent offices must provide certain guidance but are not required to verify disclosures.

Its limits are as instructive as its content. The treaty applies to patents and does not currently extend to other areas of intellectual property or to derivatives, although a review four years after entry into force may consider both. It is not retroactive: its obligations do not apply to applications filed before entry into force, subject to existing national disclosure laws. Entry into force requires fifteen eligible Parties to deposit instruments of ratification or accession, with effect three months later. It was open for signature until 23 May 2025. Whether the fifteen-Party threshold has been reached was not verified for this study and the argument does not depend on it.

The sanction structure repays attention because it illustrates how carefully the instrument's reach was bounded. Failure to disclose attracts appropriate, effective and proportionate measures. An applicant may correct a failure unless there was fraud or fraudulent intent. Where fraudulent intent exists, post-grant sanctions or remedies may apply. But apart from fraud, no Party may revoke, invalidate or render unenforceable a patent solely because of a non-disclosure. The disclosure duty is therefore procedural: it generates information and, absent fraud, does not affect the validity of what was granted.

For the present subject the treaty is significant in two ways and irrelevant in a third. It is significant because, once in force and for the Parties bound by it, it would establish in an international instrument that the providing Indigenous People or local community is a legally cognisable entity whose identification can be required which is a conceptual foundation that copyright does not supply. It is significant because its disclosure mechanism is a model capable of transplantation: a requirement to declare provenance, enforced by proportionate measures at a gatekeeping moment, is exactly the architecture that a corpus disclosure obligation would need.

It is irrelevant to training because there is no gatekeeping moment. The patent system works as a vehicle for a disclosure duty because an applicant must come to an office, make a claim and submit to examination. Model development involves no application and no examination. A corpus is assembled privately, training occurs privately and the model is released as a product. Unless a jurisdiction creates a registration, filing or audit requirement for training corpora, there is no occasion on which a disclosure duty could attach. That observation is this article's principal contribution to the international layer: the treaty's mechanism is transplantable only into a regime that first creates the gate.

6. The Institutional Layer, as Actually Disclosed

The institutional layer is where this article locates an important lever and an official disclosure from New Zealand's national museum dated 26 March 2026, responding to a request under the Official Information Act 1982 dated 26 February 2026 (Museum of New Zealand Te Papa Tongarewa, 2026) [9], permits it to be examined in unusual detail.

The consent architecture is substantial and of long standing. It rests on the Mana Taonga principle, endorsed by the institution's board in 1992 which the response states extends to digital environments and gives iwi, hapū, whānau and descendant communities a say in decisions about their taonga and associated knowledge. An iwi clearance process, managed by dedicated relationships staff, is the mechanism through which iwi indicate support or otherwise for the use and display of provenanced taonga and associated mātauranga. Release forms, loan agreements, deeds of gift, standard clauses in taonga management agreements and draft relationship agreements with iwi supply the documentary architecture and a strategy of 2004 records the expectation that Māori be treated as partners in developing and presenting mātauranga Māori. Access guidelines and licence templates exist, alongside a practice for releasing datasets where the institution owns the rights and a procedure for assigning re-use rights statements.

The institution also has an artificial intelligence governance framework. A policy of 2025 and accompanying guidelines state that Māori data governance, including Māori data sovereignty, is implicit within Mana Taonga and the guidelines call for recognition of Māori rights and interests in

AI-related decisions. Staff requests to use unapproved AI tools pass through a service desk and an internal panel that includes the Principal Advisor Kaupapa Māori, who considers Māori data governance.

Against that, the response records a set of absences and they are reported here in the terms the institution used. The response should not be read as saying that the museum has no controls of any kind over its digital collections: it describes consent, governance and security arrangements alongside the absences recorded here and every conclusion in this section is specific to Te Papa and to the date of the disclosure. It states that it holds no assessment that specifically evaluates threats to mātauranga Māori or Māori data from scraping, third-party cloud or AI services or digital colonialism; two general information-security and ICT risk assessments, from 2020 and 2022, exist but, it says, do not assess those categories. It has developed no business case or budget request specifically for mitigating AI-related risks to mātauranga Māori. Those parts of the request were refused under section 18(e) of the Official Information Act 1982 which applies where the information does not exist or cannot be found, so the response is a statement about records held and not an independent finding about institutional practice. It states that it does not hold contractual clauses explicitly referencing Te Tiriti o Waitangi or Māori data governance in its Microsoft agreements, whose detailed terms were withheld and it points to the supplier's public statement that customer data is not used to train foundation models. It describes the application of Indigenous Knowledge labels and Local Contexts notices as exploratory only and not current practice. Asked about technical measures, it describes none specific to preventing the scraping of mātauranga Māori; it refers to single sign-on protection for data not shared under an open access policy and to unspecified hosting-provider and application-level safeguards for its website and Collections Online. A draft data sovereignty analysis of its digital ecosystem dated June 2025 was released, with provisional scoring withheld and among the documents released were its 2004 mātauranga Māori strategy and its 2017 intellectual property policy and strategy.

What governs access in practice emerges from two further details. The collections interface requires an application for a key, with terms of use linked and access through it is limited to data already public on the collections website. Data not shared under an open access policy is protected by single sign-on.

Four conclusions follow and they are the article's core findings on this layer.

The practical lever is the publication decision, but availability does not mean unrestricted lawful copying and nor does every posted condition automatically become binding. The institution can choose whether to publish a file, under what licence, behind what authentication controls and with which requests for attribution or culturally appropriate use. Whether a recipient is legally bound turns on protected rights, actual contractual assent, statutory restrictions and enforceable technical controls. An unlicensed crawler that never agreed to posted terms poses a different legal problem from an API subscriber that accepted a licence.

Institutional consultation over provenance and display does not automatically regulate third-party model training. Consent to collection, culturally restricted access, derivative uses, public display, bulk API access and third-party AI training are separate questions and a protocol or clearance addressed to one does not supply authority for the others. Release procedures need to ask separately whether public web delivery, bulk API access and commissioned dataset transfers carry the same cultural risks and whether communities have agreed to the relevant release. These processes can improve governance even where legal exclusion remains uncertain.

The absence of contractual clauses referencing Māori data governance in the Microsoft agreements is the most readily remediable gap identified. An institution that has developed release forms, licence templates and taonga management agreements has demonstrated the capacity to draft; that the same institution reports no such clauses in that agreement and points instead to the supplier's public statements, identifies a drafting task rather than a conceptual problem.

The reliance on supplier statements rather than negotiated contract terms deserves emphasis because it may generalise. A public statement that customer data is not used to train foundation models is the supplier's own representation about a commercial practice. It is not reported to be a term that Te Papa negotiated, the communities concerned cannot themselves enforce it and the supplier may revise it; whether the supplier's published data protection addendum gives the institution contractual rights is a question the response does not resolve. An institution whose assurance to its communities rests on another party's published policy may therefore have transferred part of the governance question to a party with no accountability to those communities.

Traditional-knowledge labels can communicate expectations and information about culturally appropriate uses, but a label is not itself equivalent to copyright, an effective opt-out or an enforceable contract. Rights-reservation rules matter: Article 4(3) of Directive (EU) 2019/790 provides an instructive model under which a text-and-data-mining exception for lawfully accessible works is conditional on the absence of an appropriate rights reservation, including machine-readable means for online content. The EU rule does not apply automatically in Nepal or New Zealand, but it illustrates why a legal entitlement and its technical expression must be examined together.

The Canadian contribution to this layer is noted briefly and with the limitation recorded in section four. Principles for Indigenous data governance developed through Canadian and international Indigenous scholarship and organisations and widely cited in the research data community, supply the normative framework on which institutional protocols of the kind described above draw. No Canadian institutional disclosure comparable to the New Zealand response was obtained for this study and the article therefore does not describe Canadian institutional practice.

7. Nepal: Copyright Law and Institutional Governance

Nepal is the principal setting for the article's policy recommendations, but published evidence about contemporary institutional practices is limited. The analysis accordingly distinguishes copyright doctrine from proposed access-governance measures.

Nepal is not a legal blank slate. Sections 3 and 4 of the Copyright Act 2059 distinguish protected original and derivative expression from categories such as folk songs, folk tales and proverbs that the statute does not protect as such. That is a meaningful statutory baseline, although it does not settle who may give culturally valid permission, how original recordings are owned, what exceptions apply to machine

learning or whether a collecting institution's online terms are enforceable. No actual Nepali institutional access policy has been obtained for this article. Those are remaining questions, not findings of absence (Copyright Act 2059 (Nepal), ss. 3–4).

What can be said is structural and follows from the analysis rather than from Nepali sources. If the access term on which a holding institution releases material is an important lever over the use of traditional expressions in training, then a first question for Nepal is institutional rather than legislative and it is answerable by examining what its collecting institutions publish and on what terms. That is a tractable inquiry requiring no legislative change and no litigation and it is the first thing a researcher or policymaker in this field should do. The recommendations in section eight are framed accordingly: they are addressed to institutions and are available to an institution acting alone.

A second structural observation concerns sequencing and is favourable to a jurisdiction in Nepal's position. An institution that has not yet digitised and published at scale retains a choice that institutions further along have already made. The publication decision which section six identifies as an important lever, is prospective: material not yet published can be published on terms, with labels, at resolutions and through interfaces chosen with corpus assembly in view. An institution that has already published extensively cannot recall what is in circulation. This is the one respect in which arriving late to digitisation is an advantage and it is time-limited.

8. What an Institution Should Do

The recommendations follow from the finding that the access term is an important though incomplete lever and they are ordered by expected effect relative to cost, a ranking that is a judgement and not an estimate.

Table 1

What Different Governance Measures Can and Cannot Reach

The first is to treat the publication decision as a governance decision. An institution that asks, before publishing a digitised collection, what the community of origin understands about aggregate reuse, has converted an operational step into a consent occasion. This requires no new instrument, only the recognition that publication is the moment at which control is surrendered.

Table 1:

Actor	What may bind or reach the actor	Principal limits
Releasing institution	Its own consent practices, protocols, funder and legal obligations and the terms it chooses to publish or license.	Governs its own release decisions; does not by itself bind third parties who have not assented.
Authorised licensee	A signed licence with use restrictions (for example, no training, retention limits, audit rights), enforceable as a contract under the governing law.	Depends on contract formation, the licence terms, the remedies available and jurisdiction; cannot recall earlier lawful copies.
Contractual user (click-through or API key)	Terms of use brought to the user's attention and accepted as a condition of access may bind the user.	Turns on notice, assent and applicable law; limited where data is already public; limits on contracting out of statutory exceptions may apply.
Independent scraper (no assent)	Copyright, where a protected work is copied and no exception applies; other law that varies by jurisdiction.	Posted cultural-use notices alone may be only policy signals; expression excluded from copyright as such (for example under Nepal's section 4) is not reached by copyright.
AI developer using a third-party dataset	Acquisition law (the retained pirated library in *Bartz* was treated differently from training), rights reservations where the local law recognises them (for example EU Article 4) and contractual flow-down terms.	Fair-use rulings are fact-specific and not a universal rule; reservations apply only where the relevant law recognises them; flow-down depends on the dataset provider's rights.

Note. The table is a conceptual summary of the analysis in this article and not a statement of the law of any single jurisdiction.

Second, institutions should use differentiated licensing. Access agreements may address bulk downloading, machine-learning training, retention, provenance reporting and audits where the institution has authority and a contractual relationship exists. Public notices and labels can retain value as cultural governance signals, but do not necessarily create enforceable rights against everyone who encounters publicly available files.

The third is contractual. An institution procuring technology services, cloud storage or digital platform capability should include clauses addressing whether its collections data may be used for model training, whether it may be retained after termination and what audit rights the institution has. The disclosure examined reports the absence of such clauses in the Microsoft agreements and reliance on the supplier's public statements instead, at an institution with long-established consent practices; it is a drafting task within any institution's capacity.

The fourth concerns the interface. Where an institution operates a collections interface requiring an application for a key, the application may be a contracting occasion: where terms of use are brought to the applicant's attention and accepted as a condition of access, they may bind the applicant and a restriction on use for model training may then be enforceable as a contract, subject to the law governing formation and enforceability and to any limits on contracting out of statutory exceptions. The disclosure examined shows an interface so gated but limited to already-public data which means the key controls convenience and traceability for that channel rather than the availability of the material and it cannot revoke earlier lawful downloads or bind independent datasets. Aligning the two, so that bulk or structured access is available only on terms, is a technical and administrative change rather than a legal one and it promises accountability rather than prevention.

The fifth is to conduct the risk assessment whose absence the disclosure records. An institution that holds no assessment of AI scraping exposure cannot know which of its collections are most at risk and the assessment is a precondition of prioritising the other four measures.

Technical controls should be assessed realistically and layered. Authenticated access, rate limits, robots directives, API controls and provenance logging can each deter, slow, trace or document bulk collection. But robots directives are

voluntary conventions, rate limits can be evaded, an authorised user can copy and re-host material and none of these measures recalls what has already been copied. They support accountability and evidence; they should not be described as preventing scraping.

Culturally legitimate consent also needs a defined process. Which person or body can give consent for a particular item and who can revise or withdraw it, is a question for the relevant community's own governance protocols, not one an institution should settle alone. A model institutional process would record the community body or bodies with authority over the material and the basis of that authority; consult them before release and before any change to the terms of access; provide a procedure for resolving disagreement among claimants; and state that a withdrawal or revision operates prospectively on terms the institution controls, because it cannot recall copies already lawfully made. This is a model process, not a claim that one universal Indigenous consent rule exists.

For Nepal, a principles-based framework is the most that the sources examined support, because the legal authority of its collecting institutions was not examined. It would distinguish four tiers: material that communities identify as sensitive or restricted which is not released online or is released only with community authorisation; material for open educational access, released at display resolution with notices and without bulk interfaces; commercial or bulk licensing, available only under a written licence with training, retention and audit terms and, for community-derived material, community participation in setting the terms; and community-authorised uses, where the community decides the conditions. The tiers are a proposal for consultation and their enforceability depends on the matters summarised in Table 1.

Two cautions attach to the whole programme. Restricting access to cultural material has costs that fall on the public and frequently on the communities themselves, whose members may be the most frequent users of digitised collections and who may be geographically dispersed. A regime designed to prevent corpus assembly can obstruct a grandchild in a diaspora and the measures above are therefore framed as terms and notices rather than as closure. And an institution acting alone is of limited effect where the same material is held in several institutions which is the ordinary condition of collections assembled during colonial periods, so coordination among holding institutions is a precondition of any of this working.

9. Economic and Development Dimensions

Three observations follow and all are reasoned rather than measured.

The first concerns where the value is captured and is the economic heart of the problem. The material may have little direct market value as individual items; its value in training is aggregate and lies in the distinctiveness it contributes to a model's range. That value may be captured by the model's developer and distributed to its users, while the communities of origin receive no benefit and may face displacement of their own production by generated substitutes, a possibility this study did not measure. This is better described as a transfer of value than as an externality, since the model's capability partly embodies the contribution, but the claim is reasoned and not measured.

The second concerns why licensing is unlikely to solve it. A licensing market requires an identifiable right-holder, a transaction cost low relative to the value of the individual contribution and a legal background under which the user needs the licence. All three are absent: the right-holder is frequently indeterminate or is the holding institution rather than the community, the per-item value is negligible and the 2025 decisions indicate that, on the records before those courts, the user did not need a licence for training on lawfully acquired works, with one court holding expressly that lost licensing fees are not cognisable harm. Proposals for collective licensing of traditional expressions in training should confront the third point which is not a transaction cost problem but a doctrinal one.

The third concerns the development asymmetry in who develops the dilution theory. Market dilution was judicially identified in June 2025 as a theory that could succeed on a properly developed record. The parties with the resources to develop it are likely to be large rights-holders in major markets. The precedent they establish will be shaped around commercially published works with identifiable owners and its transposition to collectively held traditional expressions is not automatic. A community or state concerned with traditional expressions has an interest in the theory's development that is not represented by the litigants who will develop it which is an argument for intervention, amicus participation or scholarly development of the specific application rather than for waiting.

The study does not provide empirical measures of training-corpus composition, model revenues, displacement of community production, licensing values or digitisation costs. Its conclusions therefore concern legal doctrine and institutional governance rather than measured economic effects.

9.1. Enforceability and the Limits of Access Governance

Access governance should begin with a chain-of-control map. The first link is the community's authority over the particular expression and any conditions for culturally appropriate use. The second is the archive's ownership or licence to digitise and publish a given sound recording, photograph, manuscript or collection record. The third is what the archive has actually supplied to a user. The fourth is whether that user accepted restrictions on bulk copying or training. Rights and duties do not automatically travel across all four links. A culturally legitimate community request can exceed the archive's copyright powers; a dataset licence cannot create rights the licensor does not hold. The value of the map is that it makes those gaps visible before any digital release.

The publication channel also matters. A user who obtains a controlled download after accepting a written licence is differently positioned from a person viewing a publicly indexed web page or a developer whose crawler never clicked through any conditions. Contractual enforceability ordinarily turns on the applicable law of formation, notice and assent, while copyright remedies require a protected work, a restricted act and no applicable defence. Website notices alone are not necessarily accepted contracts: courts in several jurisdictions distinguish between terms a user actively accepts and terms merely posted on a page and enforcement depends on a genuine offer, conspicuous notice, assent, authority and jurisdiction-specific law, a body of case law this article does not survey. A restrictive notice can be useful as public evidence of institutional expectations without necessarily being a contract. An archive can strengthen its

position by separating ordinary public browsing from licensed bulk access rather than treating an undifferentiated website footer as complete permission management.

Technical measures cannot substitute for law or community judgment, but they influence whether the chosen governance boundary exists in practice. Authenticated collection APIs can log dataset access and permit institution-specific rate limits; publicly displayed metadata can be supplied separately from high-resolution recordings; and dataset exports can carry records of provenance and licence conditions. Such systems may be costly, exclusionary or circumvented. Before adopting them, an institution should assess accessibility for researchers and community users, cybersecurity risks, privacy and whether an alternative permission channel is available. The goal is accountable differentiated access rather than automatic closure of all public cultural collections.

The European Union offers a narrowly relevant statutory contrast. Under Article 4 of Directive (EU) 2019/790, the text-and-data-mining exception for lawfully accessible works is conditional on rights not being expressly reserved in an appropriate manner; for online material the Directive identifies machine-readable reservation as an example. This statutory framework can give a reservation legal significance that a cultural-use label alone does not possess. But the rule concerns covered copyright and related rights in the EU legal order; it should not be cited as a general worldwide cultural veto and it does not determine how Nepal or New Zealand treats the same website. In comparative drafting, the essential insight is to pair lawful authority to reserve rights with a technical method that communicates the reservation.

Nepal's Copyright Act 2059 illustrates the additional problem of identifying subject matter precisely. Sections 3 and 4, read together, require a distinction between folk material as such and original derivative expression. In one archival object a traditional melody, a contemporary recorded performance, editorial notes and digitised descriptive metadata may have distinct legal positions. The existence of a protected recording does not automatically give the institution moral or legal authority to disregard the community's wishes; conversely, community provenance does not by itself resolve the ownership of every new recording or adaptation. A defensible institutional register would record these categories separately, identify who was consulted, disclose uncertainty and provide a process for corrections by affected communities.

For third-party technology contracts the institution has more direct bargaining power than it does over unknown crawlers. A collection-management, OCR, cloud-storage or search supplier can be required by negotiated terms not to retain or train on supplied files beyond the commissioned service; logs, subprocessor approvals, deletion certification and audit rights can make such promises more testable. These clauses remain subject to applicable law and the supplier's technical architecture. They also cannot undo public circulation that preceded the agreement. This reinforces the importance of preserving a release history: when was an item first made openly accessible, under which notice or licence and what copies were already distributed?

Finally, cultural governance requires accountability to communities rather than only to the institution's board or legal department. Release decisions should identify which community representatives were asked, whether permissions were restricted to particular uses and how objections will be handled when an artefact has already been published. A meaningful process should allow correction of mistakes without making routine scholarly, educational or community access impracticable. The Te Papa disclosure shows that longstanding consent infrastructure and reported gaps in newer AI procurement or scraping controls can coexist. Its value is therefore diagnostic: a mature institution should periodically compare the permissions it records with the technical pathways by which people and machines actually receive its files, without claiming that the resulting policy eliminates every legal or cultural risk.

10. Counterarguments and Limitations

The strongest objection is that controlling access to heritage may itself harm community members and the public. The proposed answer is not blanket enclosure: institutions should distinguish sensitive items, high-volume machine access, ordinary educational access and community-authorised use. Notices and traditional-knowledge labels may express wishes but often lack independent enforceability. Contractual restrictions should be imposed with appropriate community authority and legal consequences should not be inferred where assent or rights are absent.

A second objection is that the article generalises from a single institutional disclosure. One response to one information request at one institution is a thin basis for a claim about the institutional layer generally. This is correct. The disclosure is used because it is specific, dated, official and unusually candid about absences and because the institution reports consent practices of long standing which makes its reported gaps informative about the limits of a consent architecture standing alone. But the inference that other institutions are similarly placed is an inference and a survey of collecting institutions' AI governance would be the obvious way to test it.

A serious limitation is the developing and jurisdiction-specific nature of fair-use and text-and-data-mining law. The two US judgments and the July 2026 order approving the *Bartz* settlement illustrate these limits, but district-court reasoning is not universal law and neither case concerned the special interests of an Indigenous or local community. The settlement also meant that the remaining piracy claims in *Bartz* did not proceed to the contemplated trial and the approval order does not decide the merits.

A further limitation concerns Nepal's statutory and institutional baseline. The cited copyright provisions distinguish traditional material from potentially protected derivative expression, but the materials analysed do not establish how Nepalese courts would apply these provisions to AI training. Current practices of Nepalese museums and archives were also outside the study, so no general inference about those institutions is drawn.

The 2024 WIPO instrument addresses patent disclosure concerning genetic resources and associated traditional knowledge, not consent for AI training datasets. It is relevant to this comparison by contrast and is not treated as an operative source of training-data consent obligations.

The proposed account is falsifiable. If legally enforceable institutional conditions have no detectable effect on scraping or training, their practical role is weaker than the policy design assumes. If stronger cultural rights or copyright claims emerge in particular jurisdictions, those may outperform access conditions. The need for genuine contracts and

differentiated collection governance is more defensible than a claim of exclusive reliance on access terms.

11. Conclusion

Traditional cultural expressions in AI training are not governed by a single body of law. The WIPO treaty addresses patent disclosure; copyright may attach to original recordings and derivatives even where traditional songs or narratives as such lack protection; contract and collection governance apply to particular routes by which copies reach developers. The contribution of this analysis is to distinguish these legal questions, not to deny the significance of intellectual-property rights.

Bartz and *Kadrey* illustrate the limits of broad extrapolation from US fair use. The courts granted context-dependent defences for training, while *Bartz* treated a retained pirated-book library differently; a court-approved settlement in July 2026 ended that case without a decision on the merits. Neither decision authorises wholesale training on all culturally sensitive materials or proves that posted museum terms create obligations for non-consenting third parties.

Te Papa's March 2026 response gives a concrete institutional example: meaningful iwi engagement and collection-governance tools coexist with reported gaps around AI-specific risk assessment and technology contracts. The appropriate response is to treat the decision to release, the method of access and any contractual grant as distinct decisions supported by community governance and technical safeguards.

Institutions should document community consultation, distinguish online display from bulk machine access, condition authenticated deliveries on genuine licences, implement rights reservations where the local law recognises them, negotiate training and audit clauses in vendor contracts and test whether those controls work. They should also distinguish enforceable obligations from aspirational cultural-use notices.

Nepal's Copyright Act provides a starting point for legal analysis; claims about Nepal's actual museum and archive practices remain untested. Further research should examine sample collection policies and contracts, the law of electronic assent and whether community representatives can enforce the resulting conditions. Until those questions are answered, release governance is a promising strategy, not a complete legal remedy.

References

1. *Bartz v. Anthropic PBC*, No. 3:24-cv-05417 (N.D. Cal. June 23, 2025) (order on fair use); 2025. Available from: https://docs.justia.com/cases/federal/district-courts/california/candce/3:2024cv05417/434709/231
2. *Bartz v. Anthropic PBC*, No. 3:24-cv-05417, ECF No. 680 (N.D. Cal. July 20, 2026) (order granting final approval of class settlement and dismissing the action with prejudice); 2026. Available from: https://law.justia.com/cases/federal/district-courts/california/candce/4%3A2024cv05417/434709/680/
3. Carroll SR, Garba I, Figueroa-Rodríguez OL, Holbrook J, Lovett R, Materechera S, *et al*. The CARE principles for Indigenous data governance. Data Science Journal. 2020;19:43. doi:10.5334/dsj-2020-043
4. Copyright Act 2059 (2002) (Nepal), ss. 3–4. English extract: World Intellectual Property Organization, WIPO Lex traditional knowledge laws database, https://www.wipo.int/en/web/traditional-knowledge/w/tklaws/article_0131. Nepali text (not independently reviewed): Nepal Law Commission, https://repository.lawcommission.gov.np/np/category/documents/law-archives/statutes-acts-law-archives/
5. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, art. 4. *Official Journal of the European Union*, L 130, 92–125. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019L0790
6. *Kadrey v. Meta Platforms*, Inc., No. 3:23-cv-03417-VC (N.D. Cal. June 25, 2025) (order on cross-motions for summary judgment); 2025. Available from: https://law.justia.com/cases/federal/district-courts/california/candce/3:2023cv03417/415175/598/
7. Kukutai T, Taylor J, editors. Indigenous data sovereignty: Toward an agenda. ANU Press; 2016. doi:10.22459/CAEPR38.11.2016
8. Lemley MA, Casey B. Fair learning. Texas Law Review. 2021;99(4):743-785. Available from: https://texaslawreview.org/fair-learning/
9. Museum of New Zealand Te Papa Tongarewa. Response to Official Information Act request: Protection of mātauranga Māori and Māori data. 2026. Available from: https://www.tepapa.govt.nz/assets/76067/1774567964-response-to-oia-request-protection-of-matauranga-maori-and-maori-data-26-03-2026.pdf
10. World Intellectual Property Organization. WIPO Treaty on Intellectual Property, Genetic Resources and Associated Traditional Knowledge. 2024. Available from: https://www.wipo.int/treaties/en/ip/gratk/

How to Cite This Article

Bhattarai P. The access boundary, not a complete legal shield: traditional cultural expressions, copyright and institutional control of AI training data. Int J Judicial Law. 2026;5(6):48-56. doi:10.54660/IJJL.2026.5.6.48-56.

Creative Commons (CC) License

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.