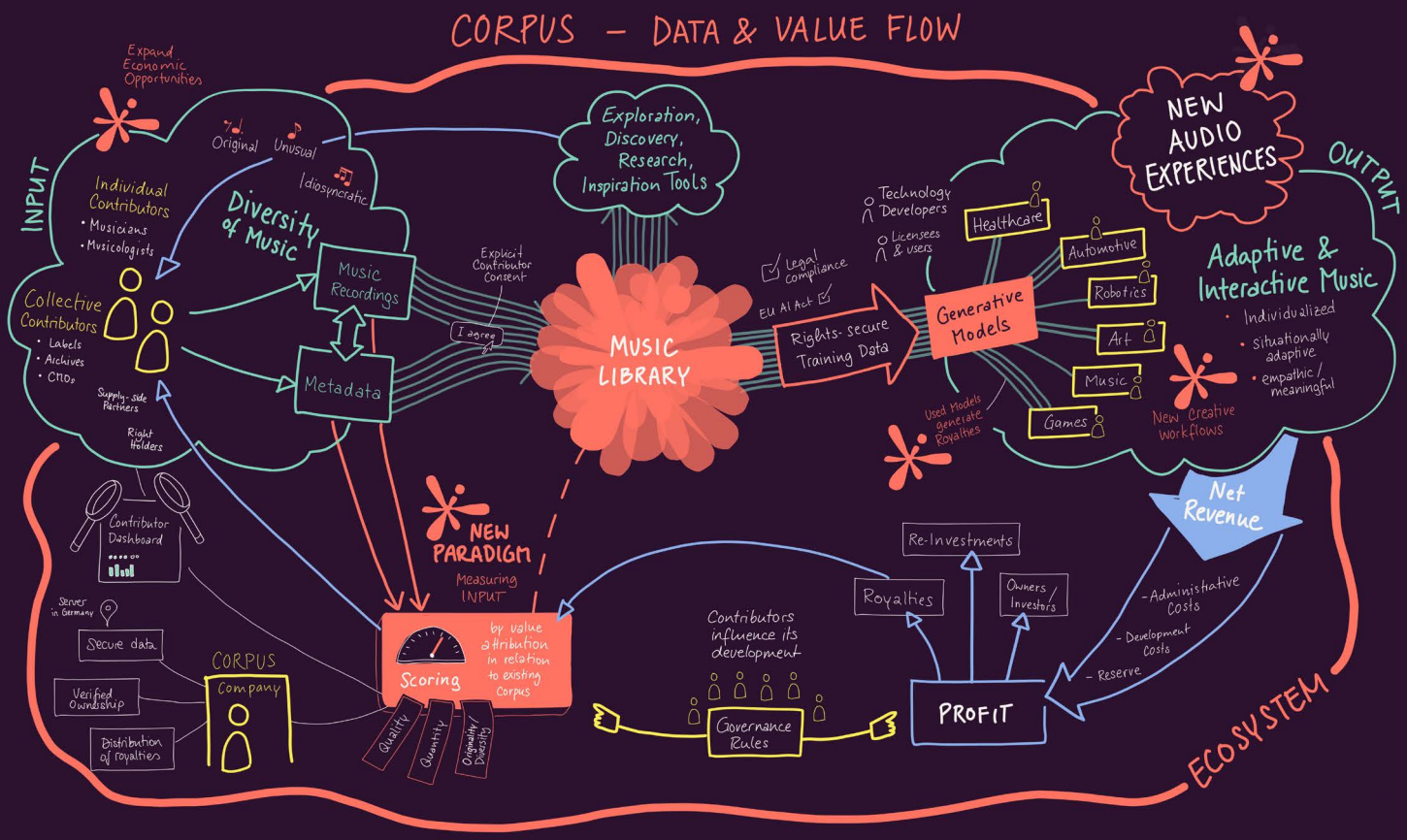
CORPUS

Building the Infrastructure for a New Music Economy in the Age of AI



CORPUS Data and Value Flow: CORPUS links musical contribution, model training, and economic participation in a single system, shifting value attribution from copied outputs to licensed inputs and shared downstream use.

Executive Summary

From Music Licensing to Music Infrastructure

The music industry built its entire economic infrastructure around a single dimension of musical value: the transaction. Music as product - counted, copied, consumed. This logic powered an immense global industry. But it is blind to everything else music does. Streaming pushed this reduction to its extreme. Generative AI now brings it to a structural limit.

CORPUS is designed to reverse this reduction.

It is a licensing and royalty protocol that enables music creators to participate in the AI transition - not as suppliers to a system that extracts their work, but as co-owners of the infrastructure their contributions build. Contributors keep their rights. Each work is evaluated for quantity, quality, and originality relative to the existing library. Originality is rewarded because musical diversity strengthens the models trained on it.

The protocol is open and auditable. The data is protected through federated learning - licensees train models within secure environments; the corpus itself never leaves CORPUS infrastructure. Governance begins at the scoring system itself, where contributor juries calibrate what the protocol rewards. Contributors receive ongoing royalties and accumulate CRPS (Corpus Participation Rights) - a lasting stake in the system their work builds.

For companies, CORPUS provides a compliant, scalable alternative to both scraping and buy-outs. But its role goes beyond licensing. The markets now emerging - adaptive sound in vehicles, therapeutic music in healthcare, responsive environments in games, semantic interaction in robotics - require music that functions not as product but as situated experience. These markets do not yet exist within the music economy. They could surpass it in size. CORPUS is building the infrastructure to make them possible.

Imprint

This white paper is a living document.
Current version: 3 March 2026

Written by
Mathis Nitschke (Founder and CEO, Sofilab and CORPUS)

Edited by
Virginie Berger (Business and brand consultant, artist advocate)
Maria Goeth (Communications director)

Proofreading by
Max Graf (AI Engineer, CORPUS)
Lars Ullrich (CTO, CORPUS)

Map by
Anja von Klitzing (www.dialogstifter.de)

Illustrations by
Jan Stoewe (www.jansteins.de)

Layout by
Anja Gerscher (www.anjagerscher.de)

Published by
Sofilab GmbH
Akademiestr. 3
80799 Munich
Germany

Represented by
Mathis Nitschke

Register court: Munich District Court
Commercial register: HRB233361

Contact
info@sofilab.art
www.sofilab.art

SOFILAB

CORPUS is a project by SOFILAB

Notes on the March 2026 revision

This revision introduces two changes to the White Paper. The first is conceptual. The original document described CORPUS primarily as a licensing and royalty system – a fairer way to distribute revenue from AI-trained music. This revision reframes CORPUS as a response to a deeper structural problem: the music industry's progressive reduction of musical value to a single dimension – the transactional. Music as product, counted, copied, consumed. This logic built an immense global industry but rendered invisible the other dimensions of value that music has always carried: shared experience, situated meaning, emergent resonance between sound and context. AI brings this reduction to its structural limit. CORPUS is designed to reverse it – by building economic infrastructure that can recognize and sustain value beyond the transaction.

This argument is woven into the existing text at five points: the introduction (1.1), the analysis of why copy-based licensing fails (3.2/3.3), the incentive logic that rewards diversity (7.1), and the closing (9).

The second change is structural. Three areas that were deliberately kept vague in the original are now addressed with specificity. Section 7.5 introduces the dual currency design: ongoing royalties in conventional currency plus CRPS (Corpus Participation Rights) that represent a lasting stake in the system itself. To prevent early contributors from permanently capturing a disproportionate share of ongoing revenue, the diversity component of the royalty score is subject to gradual temporal decay — while CRPS, once issued, do not expire. Three legal pathways for CRPS are presented as under evaluation. Section 8 replaces the generic discussion of governance models with a concrete three-layer architecture — an open protocol (transparency), controlled data access through federated learning (protection), and a proprietary semantic pipeline (commercial value) — and introduces a phased governance model: a scoring jury selected by sortition from the contributor base in the early phase, dual-track governance separating operational speed from deliberative oversight post-launch, and a foundation-based institutional structure as the long-term safeguard.

Table of Contents

1 Introduction

1.1 From artistic Frustration to infrastructural Necessity

CORPUS grew out of artistic work. Sofilab, a Munich-based sound design and innovation lab, began experimenting with generative music models in the late 2010s. The technology was advancing fast, but the available training material was either low-quality, poorly annotated, or legally unusable. Without a structured, rights-compliant corpus, artistic exploration remained labor-intensive and yielded limited results. One thing became clear early: the performance of generative models depends at least as much on their training data as on their architecture.

But the more consequential insight came from a different direction. Sofilab's work in automotive, medical, and robotics industries required sound that behaves — not sound that is played. In a vehicle, sound must respond to driving dynamics, driver state, and context. In therapeutic settings, music must adapt to a patient's condition in real time. In each case, the task was the same: translate a complex, evolving situation into meaningful sonic response. Manual preproduction cannot scale to this — every additional sensor or behavioral parameter multiplies the conditions, creating a space that cannot be exhaustively designed by hand. This is where generative AI becomes essential: not as a novelty, but as the only way to make sound interactions flexible and context-sensitive in unforeseen situations.

These two lines of work — generative music research and adaptive sound design — converged on a single recognition. The music industry has built its entire economic infrastructure around one dimension of musical value: the transaction. Music as product — counted, copied, consumed. This logic powered an immense global industry. But it is blind to everything else music does. When a vehicle's sound environment creates a sense of calm that neither the driver nor the system designed in advance, that is not a transaction. When a familiar melody reaches a dementia patient in a moment of confusion, that is not a copy. Music has always created value through shared experience, through situated meaning, through what emerges when sound meets context. These dimensions never disappeared from human life. They disappeared from the economic system.

The streaming economy pushed transactional logic to its extreme: royalties tied to plays, algorithms optimized for repeat consumption, value concentrated around mass appeal. This made unusual, original, and culturally specific music economically invisible — not because it lacked value, but because the system could not see value beyond the transaction.

Generative AI now brings this reduction to its structural limit. When machines can produce unlimited music at near-zero cost, the transactional dimension alone can no longer sustain

an industry or justify the participation of human creators. But AI simultaneously opens entirely new domains — vehicles, healthcare, robotics, education, architecture — where music functions not as product but as adaptive, context-responsive experience. These markets do not yet exist within the music industry. They could ultimately surpass it in size.

This is the dual condition that defines CORPUS. The old model of musical value is collapsing. A new one is emerging. CORPUS is designed to build the infrastructure for the transition: a licensing and royalty system where contributions are valued not by consumption but by how they enrich a shared generative resource — where every licensing system is understood as an incentive system, and where the incentives finally reward what matters: quality, originality, and diversity — and where contribution generates not only revenue but lasting participation in the system itself.

Figure 2: Independent musicians contribute original recordings to CORPUS

1.2 Who this Protocol is for and how the Ecosystem fits together

CORPUS is conceived as more than a licensing protocol: it is the foundation of a complete ecosystem that connects creators, rights holders, and technology developers. This includes a contributor dashboard, scoring methods, and governance rules alongside the licensing system itself. The system only works if these roles interact transparently and sustainably, aligning incentives across the entire value chain.

Within this ecosystem three primary groups can be distinguished:

- **Individual contributors (e.g., independent musicians)**
They supply the creative material. Their main concerns are: how their works are licensed, how royalties are calculated, and what rights and controls they retain. Transparency and a clearly defined opt-in structure are essential – including the ability to withdraw works from future use while maintaining earned rights from past training cycles.

- **Supply-side partners (e.g., publishers, labels, libraries)**
They manage larger catalogues. Their focus is on integrating CORPUS into existing workflows, scaling value attribution across thousands of works, and ensuring that legal safeguards protect both rights holders and the licensees who rely on them. In practice, such partnerships are typically established through direct agreements rather than via the public contributor platform, which is why they are referenced in this white paper only in selected scenarios.

Figure 3: Publishers, labels, and libraries integrate larger catalogues into CORPUS

- **Licensees and users (e.g., AI developers, game studios, tech companies)**
They need rights-secure training data. Their questions center on reducing legal and reputational risk, tracking and verifying usage, comparing CORPUS with buy-out models, and establishing provenance and permission with confidence.

The white paper is therefore structured around the licensing and royalty infrastructure that connects these groups. Model architectures or training workflows are referenced only where they help illustrate how licensing, attribution, and governance are applied in practice.

Figure 4: AI developers and technology companies license rights-secure training data from CORPUS

2 Current Industry Practices

Despite rapid progress in model architectures and commercial interest, the foundations of generative music AI remain fragile. The datasets used for training are either legally uncertain or financially prohibitive, leaving no sustainable path for artists nor developers.

For example, research datasets such as MAESTRO are licensed for non-commercial use only, making them unsuitable for commercial training, while other large-scale datasets have been criticized for sweeping in copyrighted works without consent.

2.1 Why Scraping, Buy-Outs, and Legal Exceptions Fail

Current practices and legal doctrines all face critical barriers:

- **Scraping** of streaming platforms and archives provides cheap data, but exposes developers and downstream users to lawsuits. For creators, scraping means no consent, no attribution, and no compensation. Models trained this way are nearly impossible to commercialize safely. Downstream adopters – music tools, games, ad platforms – inherit this liability.

- **Buy-outs** offer legal certainty but at prohibitive cost. Even modest catalogues can cost hundreds of thousands of euros, while competitive models require hundreds of thousands of hours of music. Once acquired, catalogues are static, cannot adapt to evolving model needs, and exclude artists from any ongoing revenue.

Both approaches – scraping and buy-outs – create bottlenecks. One excludes rights holders, the other excludes independent developers, startups, and regional rights holders who cannot afford massive catalogue acquisitions. As model complexity and dataset requirements continue to grow, this polarization makes the field increasingly unsustainable.

- In the United States, some developers argue that training qualifies as **Fair Use**. But the four-factor test weighs against this view. Commercial intent is high, the entire work is ingested, music is inherently creative, and the effect on the market is disruptive. Recent cases, such as Warhol v. Goldsmith and ongoing lawsuits led by the Authors Guild, have further narrowed the scope of transformative use in commercial contexts. Relying on Fair Use as a long-term foundation is increasingly precarious.

- Europe follows a different path, but the outcome is similar. Article 4 of the **DSM Directive allows** commercial text- and data-mining only if rights holders have not opted out. Several CMOs, including GEMA, have already exercised such opt-outs, explicitly blocking unlicensed training on their repertoires. At the same time, the EU AI Act requires

providers of general-purpose AI models to disclose the provenance of their training data. This combination makes compliance significantly harder: unlicensed datasets are not only risky but structurally non-compliant with upcoming regulation.

Taken together, these factors create a patchwork of legal uncertainty and escalating disputes. Scraping cannot remain viable, buy-outs cannot scale, and exceptions cannot be relied on. The space for grey-area practices is shrinking fast, leaving both developers and creators without a sustainable path forward.

2.2 What a new Approach must Deliver

The current landscape leaves both creators and companies without a viable path forward. Scraping erodes trust and invites litigation; buy-outs concentrate access in the hands of a few; collective rights systems were never designed for training data.

A sustainable framework must meet three conditions at once:

- **Legal compliance** – music explicitly licensed for training, not swept in by default.
- **Fair compensation** – royalties reflect the contribution and influence of each work in AI training.
- **Economic scalability** – access affordable for startups, SMEs and cultural institutions, not just major players.

CORPUS is built to meet these conditions by shifting licensing to the input side. Contributions are licensed opt-in, evaluated for quantity, quality, and originality relative to the existing corpus, and rewarded accordingly. This creates an incentive system that strengthens the dataset while remaining legally defensible and economically inclusive.

Such a system cannot be realized within the frameworks that dominate today's music industry.

3 Why Existing Rights Systems struggle with AI Music

3.1 CMOs were built for Copies and Performances, not Training

Collective rights management organizations (CMOs) such as GEMA, SACEM, or PRS were created to solve the problem of licensing copyrighted works at scale in the era of mechanical reproduction, broadcasting, and public performance – a framework shaped between the late 19th and mid-20th century. Their model assumes that works exist as fixed units and that value arises when those units are copied or transmitted at scale.

AI music generation does not fit this framework. It does not copy or perform existing recordings for end-user listening, but produces new audio from learned patterns. This raises immediate legal questions: is training itself a licensable act? Do CMOs even hold the relevant rights to issue such licenses? For example, PRS's performance rights terms do not cover training. More broadly, most membership agreements were written long before AI was a factor, covering public performance and reproduction but not machine learning.

Some CMOs have already begun to explore AI-specific licensing, but in very uneven ways. In Sweden, STIM has launched a pilot collective license for AI training in partnership with startups such as Songfox and Sureel, where authors can explicitly opt in and receive attribution and royalties. In France, SACEM has taken an authorization-first stance, requiring explicit consent for AI data-mining of its repertoire. In Germany, GEMA has proposed a licensing model that would require AI developers to pay rights-holders not only for training but also for certain downstream uses of AI-generated music, including proposals for significant revenue shares (e.g. 30% of net income) and minimum royalties. At the same time, GEMA has floated extensions of collective licensing frameworks, including Extended Collective Licensing (ECL), to apply to AI training – an approach whose legal basis in this domain remains uncertain.

These examples show that CMO strategies are not coordinated and sometimes push beyond existing mandates. This inconsistency highlights unresolved questions: Who decides what can be licensed? Who gets paid? And do CMOs legally hold the rights they claim to exercise over AI training?

3.2 The Collapse of "Copy logic" in a World of Generative Outputs

At a deeper level, the problem is not just contractual – it is conceptual. The entire logic of collective rights management is built around the economic regulation of copying: the assumption that value is created when a work is duplicated and used at scale.

In the context of generative AI, there may be no copies at all. A therapeutic robot generating calming music for a patient, or a game engine producing a one-time soundtrack for a player, illustrates the point: the output exists only in that moment and may never be heard again. There is no original and no copy – only individualized, ephemeral experience, with no intermediary platforms, publishers, or playback channels. Concepts like public performance, mechanical reproduction, and broadcast no longer apply in a meaningful way.

But the problem runs deeper than a technical mismatch. Copy-based licensing was always a system for administering one specific dimension of musical value: the transactional. It counted units. It tracked reproductions. It measured consumption. When music functions as an adaptive, situated response – to a driver's state, a patient's condition, a player's movement – the value it creates is not transactional. It is relational and emergent. No copy-count can capture it, because there is nothing being copied. Retrofitting transactional infrastructure onto these use cases will produce the same structural errors the industry made with downloads and streaming: overreach, friction, and a further loss of trust.

3.3 CORPUS's input-side Paradigm as a Response

CORPUS takes the opposite approach. The decisive moment is not the detection of copies but the licensing of contributions for training. Each contribution is assessed for how it enriches the existing library – its quantity, quality, and diversity – and receives an input weighting that remains fixed once assigned. This weighting does not generate payment on its own. It determines how contributors participate in the royalty pool once models trained on their works are licensed and deployed. Payouts remain tied to real economic use: model licensing, device deployments, API calls.

This is more than a different counting method. It shifts how musical value is constituted. In the transactional model, a work's value is determined by how many times it is consumed as a discrete product. In CORPUS, a work's value is determined by how it strengthens a shared generative system – how its distinctiveness expands the capability of models that enable new forms of musical experience. The individual contribution gains economic meaning – both as ongoing revenue and as a stake in the system – through its role in a larger whole. And the larger whole gains its capability precisely from the diversity of its parts.

Where current collective rights systems measure outputs, CORPUS measures inputs – and builds an economic logic that can recognize value beyond the transaction.

4 Applications: Where Licensed, Diverse Training Data is Essential

CORPUS is designed to serve domains where conventional approaches to music cannot meet the demands of scale, responsiveness, or diversity – and where AI can unlock new possibilities. These use cases show where generative music creates the highest leverage: expanding creative workflows for musicians, enabling adaptive sound in industrial contexts, and opening cultural and educational applications. Crucially, AI models trained on unlicensed data cannot be deployed ethically and commercially in regulated or brand-sensitive sectors such as automotive, healthcare, or advertising. A licensed and diverse corpus is therefore not just desirable, but the only viable foundation for lawful, scalable, and trustworthy adoption.

Figure 5: Generative music unfolds as an adaptive, responsive environment rather than a fixed asset.

Music production tools and Digital Audio Workstations

AI-assisted composition tools and web-based DAWs are proliferating, lowering barriers for musicians who lack access to education, budgets, or professional networks. This segment is expected to see strong growth, with new platforms appearing almost weekly. In these environments, users do not simply consume AI output but actively refine and correct it. That human-in-the-loop process is precisely what generates value: each interaction strengthens creative outcomes and, when built on CORPUS-licensed models, feeds back into a virtuous cycle where musicians/producers benefit both as contributors and as users of ever-improving tools.

Interactive Artistic Environments
Games, XR, and installation art increasingly demand music that behaves like a living environment rather than a fixed soundtrack. CORPUS-trained models allow scores to unfold in real time, responsive to player movement, narrative, or audience interaction. While adaptive music already exists, it is bounded by pre-produced assets. With CORPUS, these boundaries expand: developers could, for instance, integrate a Wwise plugin into Unity or Unreal Engine, enabling AI-trained scores that adapt continuously without exhausting variation. This creates artistic experiences that are not just reactive but genuinely co-evolving with their users.

Adaptive soundscapes in mobility
Vehicles process a constant stream of inputs: driving behavior, speed, location, weather, and even driver state. Designing sound responses for all these conditions manually is impossible. CORPUS-trained models enable adaptive soundscapes that respond intelligently – supporting driver concentration, comfort, and emotional balance while maintaining consistency with the vehicle's brand identity. Each vehicle can host its own licensed model, generating feedback and music that adjust in real time to context – creating a coherent yet individualized auditory experience across an entire fleet.

Healthcare robotics and therapy
Hospitals increasingly explore humanoid robots and assistive devices that provide guidance, companionship, and emotional support, while also helping to reduce staff workload. For dementia patients, for example, music and responsive listening are key: familiar songs, regional repertoires, or even improvised singing can create comfort and connection. Research shows that music therapy can reduce agitation, improve mood and memory, and strengthen communication and social bonds for people living with dementia. CORPUS emphasizes diverse, global contributions, making such culturally attuned applications possible – something standard datasets, dominated by Western commercial music, cannot provide. Beyond robotics, hospital devices can also draw on music-therapy principles to address alarm fatigue – replacing uniform, high-stress alerts with sound cues that differentiate urgency levels and convey calm where appropriate. This allows critical alarms to remain clear while reducing unnecessary stress for patients and staff.

Advertising and brand communication
Brands increasingly invest in distinctive sonic identities – from carefully composed sound logos to curated music palettes – but today these assets remain largely static. CORPUS-trained models make it possible to extend such brand references into adaptive experiences that stay recognizably on-brand while responding to context. A retail store, for example, could move beyond fixed playlists: background music might shift with the time of day, then adapt in real time to live sensor data such as visitor numbers, demographics,

or overall atmosphere. At the same time, lawful machine-generated music enables rapid, low-cost production of branded material – useful for campaign prototyping, regional adaptations, and highly segmented user groups. This allows brands to maintain aesthetic coherence across thousands of micro-contexts without relying on unlicensed or generic content.

Exploratory Cultural Mediation and Education
Apart from highly interactive and personalized instrumental lessons, music education still relies largely on static examples and fixed recordings – limiting active engagement and experimentation. CORPUS-trained tools make music itself interactive: learners could trace the lineage of a folk melody across cultures, reshape motifs on the fly, or dialogue with a model trained to explain harmonic structures. In public institutions such as libraries or museums, this enables exploratory interfaces where visitors not only listen to regional repertoires but transform them – remixing, extending, or cross-pollinating across traditions. Such forms of cultural mediation depend on detailed metadata that connects works across styles, origins, and histories – another reason CORPUS's annotated diversity is essential. These tools make musical literacy embodied, playful, and deeply satisfying, opening new pathways for institutions to connect heritage with contemporary creativity.

Taken together, these applications demonstrate a decisive fact: in many domains – from music production to artistic interactive environments, from mobility to healthcare, advertising and cultural mediation and education – the scale and adaptivity required cannot be achieved with traditional approaches to music creation. AI expands creative workflows by lowering barriers and adding new possibilities. Only AI models trained on a rich, diverse, and deeply annotated music library can generate sound that adapts continuously to context, scale, and interaction – and only if that corpus is built on trust, licensing, and contributor participation.

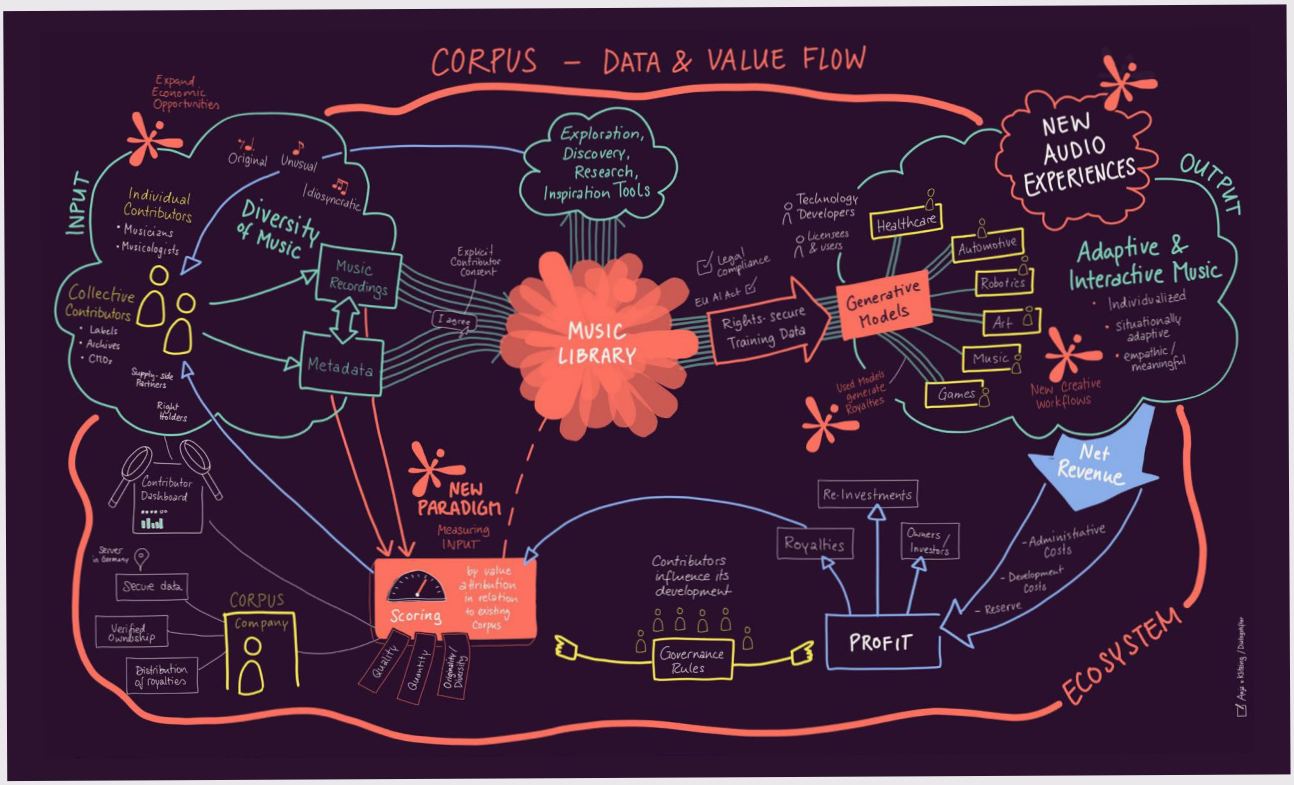


Figure 6 CORPUS – Data and Value Flow: Musical works enter CORPUS, are evaluated for their contribution, and serve as licensed training data. Revenues from model use return to contributors, linking creative input to shared economic value.

5 The CORPUS Royalty Protocol

CORPUS is in development. What follows is not a description of an existing system, but the design principles guiding its implementation. The protocol is conceived as a licensing and royalty infrastructure that incentivises the global musician community to contribute their music to a shared library, designed to serve as a training dataset for generative AI and other downstream AI applications.

Security and rights management are integral to this process: uploads will be verified for provenance, licensed works tracked across training and deployment, and royalty flows made auditable and transparent. The principle is clear: businesses and their users gain legal certainty, and music contributors maintain trust and long-term participation. Later chapters describe the mechanisms in detail – from verification procedures to auditing and governance.

In its initial phase, CORPUS will operate as a platform to establish the dataset and validate the licensing model. Over time, it is intended to evolve into an open protocol that can be embedded into existing services and creative tools, providing the foundation for an ecosystem that extends beyond any single marketplace or application.

Future integrations could make this process seamless. For example, musicians uploading tracks via distribution services such as MusicHub, TuneCore, or DistroKid might see an additional option: “also register with CORPUS.” The same could apply to community platforms like Bandcamp, where independent artists could license their works for AI training alongside sharing them with audiences. With a connected wallet, the contribution would be registered in one step. Depending on the integration, CORPUS could reuse existing rights metadata from the distributor or import structured data via standards such as DDEX, with only a lightweight additional declaration for AI-specific consent to ensure all parties agree. The same principle extends to creative tools: in DAWs or music apps, CORPUS could appear as a native publishing or export target.

The protocol borrows the structural logic of Web3 systems – automated and trustless execution of agreements – while remaining technology-neutral at this stage. Whether blockchain becomes part of the technical foundation will depend on scalability and adaptability; for now, CORPUS prioritises flexibility in the early stages of development.

Figure 7: By designing licensing, provenance, and auditability into the system from the outset, CORPUS makes regulatory compliance— including alignment with the EU AI Act—a structural outcome rather than an added constraint.

5.1 Core Design Principles: Dynamic Licensing, Traceability, Governance

CORPUS balances creator incentives with industry compliance, ensuring that contributions remain usable, secure, and commercially deployable in sensitive markets. These principles define the framework for licensing, attribution, and governance, making CORPUS not only fair for contributors but also the only viable path for lawful, scalable adoption in regulated industries.

- **Dynamic licensing and royalties:** Instead of buy-out catalog purchases, CORPUS licenses works under ongoing agreements. Contribution value is weighted on the input side, evaluated for quantity, quality, and originality relative to the existing library, so that rewards favor material that expands diversity and depth. Royalties follow the monetization logic of downstream applications – for example, models licensed per device in a car fleet, or API usage in a game engine – so that rewards scale with real economic activity.

- **Fully licensed, traceable, and opt-in:** All music in CORPUS enters through explicit contributor consent. Works remain owned by their creators and can be traced across models and applications through a central registry that links dataset usage back to individual contributions. In practice, contributors gain visibility through their dashboard: once a dataset has been compiled for a model, they can see where their music was included and how it generates value.

- **AI as a creative partner and market enabler:** Models trained on CORPUS are not designed to replace musicians but to expand both artistic workflows and economic opportunities. They can function as responsive collaborators in creative contexts, while also powering entirely new markets for adaptive sound in vehicles, healthcare, robotics, advertising, and other interactive domains. These emerging markets have the potential to outweigh the losses caused by generative AI in traditional music distribution by a large factor. Equitable structuring is essential; without it, the value created by these markets bypasses creators.

- **Community-governed evolution:** The protocol includes mechanisms for contributors to influence its development: auditing revenue flows, adjusting attribution rules, and participating in governance decisions. Section 8 describes a phased governance model – beginning with a scoring jury selected by sortition and evolving toward institutional safeguards – designed to ensure that those who provide the data also shape the system's direction.

- **Open and engaging discovery:** CORPUS-native tools are designed to make both creation and discovery adaptive and alive. The library is not a closed archive but a living

resource: contributors' music remains accessible for exploration, learning, and interaction. Discovery is intended to be playful and serendipitous rather than static or purely transactional.

Beyond search, CORPUS enables new creative workflows: musicians and producers can explore music by mood, story, or emotional arc, uncovering connections that spark ideas rather than reproducing clichés – and in doing so, reward the diverse contributions that feed these exploratory tools.

Tools like our Story2Music link narrative prompts directly to musical ideas, reviving collaborative workflows that stock music has displaced and opening new spaces for artistic and professional creation.

5.2 The Licensing Framework: Connecting Contributors and Licensees

The licensing framework connects three key actor groups:

• **Contributors:** Musicians provide recordings, and they can also add or refine annotations such as metadata or quality checks. While creators are encouraged to enrich their own works, CORPUS recognizes that only a minority will consistently do so. To ensure coverage and quality, other participants can contribute and edit annotations as well and receive points for this value addition. In this way, both music and metadata become part of the shared corpus and are rewarded through the same incentive system.

• **The CORPUS dataset**: Licensed works and annotations form a central dataset that supports multiple AI training pipelines. A single contribution can generate value across many contexts – for example, a track included in training a generative engine for cars may also be part of subsets later licensed for adaptive music in games, healthcare applications, cultural projects or mobility systems. In this model, contributors are not compensated only once, but share in multiple downstream revenue streams each time the dataset or its subsets are licensed for training. This makes every contribution a recurring source of income, aligned with the evolving applications of the corpus.

• **Model users, developers, and service providers:** AI developers license access to the CORPUS dataset, either the full library or specific subsets tailored to their application. Revenue from these deployments flows back into CORPUS and is redistributed to contributors according to the system's dynamic value attribution.

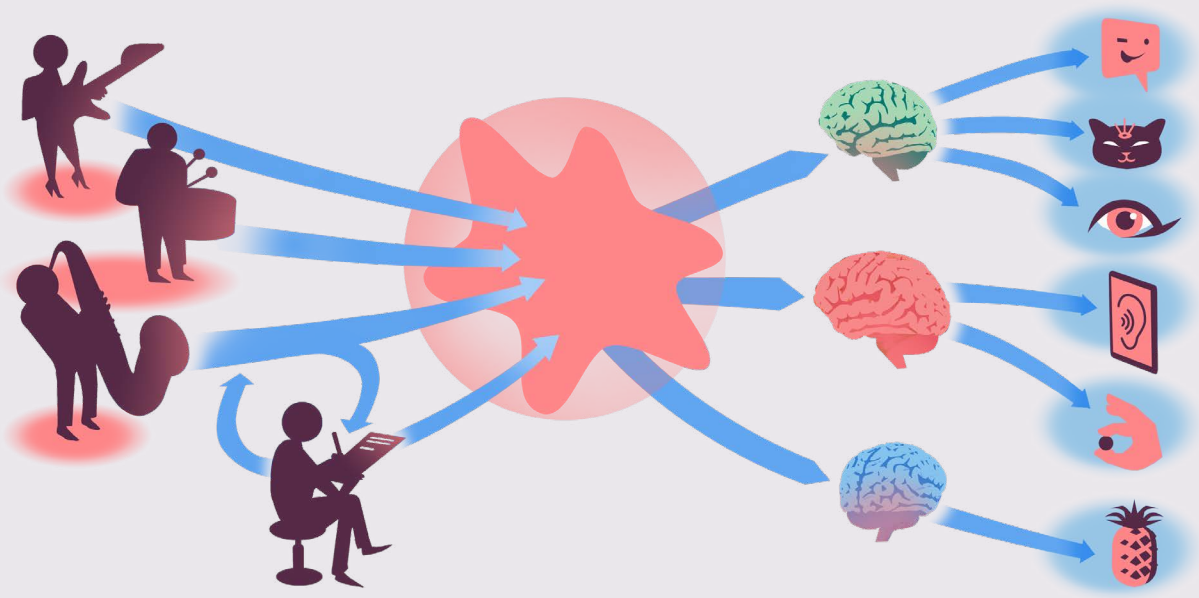


Figure 8: Musicians contribute recordings to a shared corpus. This corpus is used to train multiple AI models with different goals. Each model can then power a wide range of applications and behaviors.

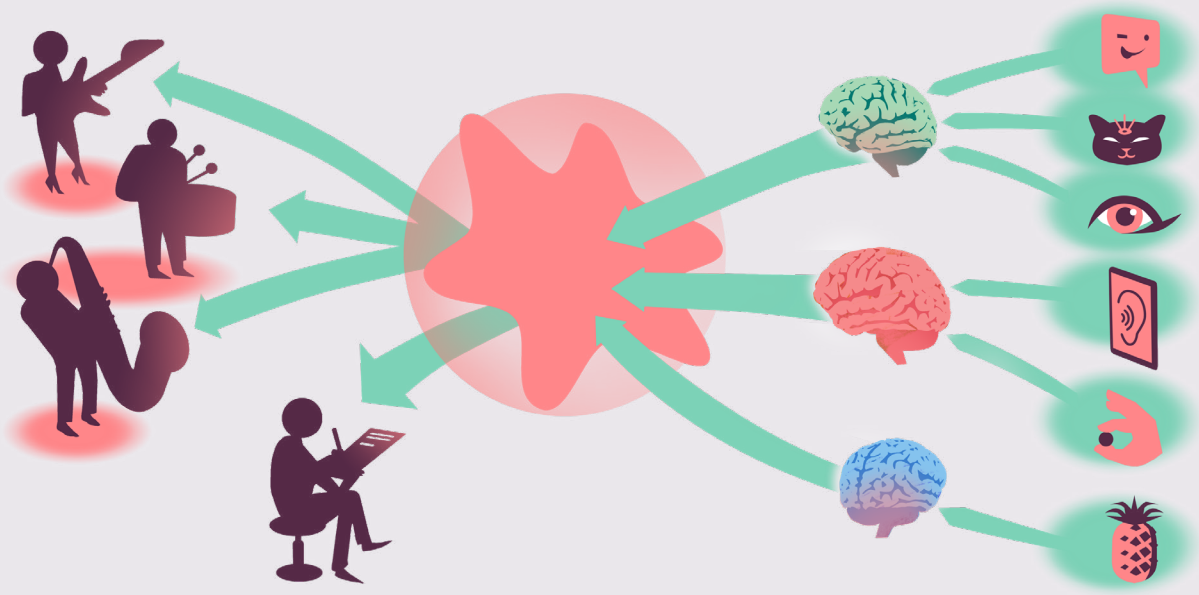


Figure 9: AI developers pay for model usage; downstream applications generate revenue; and royalties flow back to contributors based on their weighted participation in trained models.

6 Securing Contributions and Rights Integrity

The credibility of CORPUS depends on whether rights holders, industry partners, and regulators can trust that the system secures data, verifies ownership, and distributes royalties in a legally defensible way. Security and rights integrity are not add-ons but the backbone of the protocol. This chapter outlines how contributions are uploaded, verified, licensed, and traced – some measures already in place, others phased in as CORPUS scales. Together, they form the roadmap for a transparent, enforceable licensing infrastructure.

6.1 Protecting Uploads and Storage from the Start

All contributions enter CORPUS through encrypted upload channels and are stored on self-administered servers located in Germany under EU data protection law. This avoids reliance on opaque cloud hyperscalers and ensures data is handled under clear governance. As the corpus grows, this will evolve into a federated global server network with nodes in strategic locations, each operating under the same standards.

Infrastructure is aligned with recognized security frameworks, including ISO 27001 (risk and information security management) and SOC 2 (audited controls for secure operations). While formal certification will follow at commercial scale, alignment from the outset ensures compatibility with partner expectations and with upcoming regulatory obligations under the EU AI Act, such as Article 52 provenance requirements for foundation models.

For model training, CORPUS plans to integrate confidential computing technologies such as NVIDIA's secure enclaves, which ensure that even during active computation, training data cannot be inspected or exfiltrated by providers – closing one of the most sensitive attack vectors in AI workflows.

6.2 Verifying Rights and ensuring Provenance

Ensuring that contributions are legally usable is central to CORPUS's value. The system combines contributor consent, ownership safeguards, and transparent verification procedures.

Contributions are currently governed by provisional, non-commercial terms. These permit internal model training for R&D, prototypes, and partner demos, but not the public release of trained models. At no point does contributor data become platform-owned. Once commercial licenses are finalized, contributors will explicitly opt in, ensuring that all future use is revenue-sharing and transparent.

• **Contributor ownership and consent**
Contributors retain ownership, and works are licensed only through explicit opt-in. Withdrawal is possible at any time, stopping future use of a work in new training runs. However, if a work has already contributed to model training, its influence remains embedded in those weights. CORPUS therefore continues to treat the work as royalty-eligible, ensuring that contributors are compensated for past contributions even after withdrawal.

This approach respects withdrawal rights for future uses while recognizing the technical impossibility of "untraining" existing models. It also sets CORPUS apart from today's industry norm, where once a work is ingested, contributors lose both control and economic participation. CORPUS instead guarantees residual value – a unique mechanism that keeps compensation flowing even when rights holders later change their participation status.

• **Split-contribution attribution**
CORPUS supports split attribution for works with multiple rights holders (e.g., composer, performer, producer). Each contributor is acknowledged, and commercial licensing requires explicit confirmation or indemnification to manage disputes transparently.

• **Peer review and metadata validation**
Metadata and quality checks are partly community-driven. To prevent bias or neglect, initial deployments will rely on invite-only expert reviewers. From 2026 onward, CORPUS will introduce structured systems – for example, reviewer tiers or a strike system – supported by audit logs and escalation paths to ensure transparency and accountability.

• **Detection of AI-generated uploads**
As of now, CORPUS actively screens uploads for synthetic origin to prevent unauthorized AI-generated material from entering the corpus. Submissions are checked using dedicated detection systems as part of the ingestion pipeline, and flagged material is reviewed before acceptance. Because detection methods and generation techniques continue to evolve, this remains an adversarial and moving target.

Looking ahead, CORPUS plans to complement detection with additional safeguards, such as watermarking or hash-based provenance tools where they prove reliable and practical. In parallel, the protocol is being developed with emerging regulatory frameworks in mind. Several jurisdictions, including the EU under the AI Act, are moving toward mandatory provenance disclosure, and CORPUS is preparing to adapt as such requirements become concrete and enforceable.

- **Handling infringements and disputes**
  If a contribution is challenged, CORPUS is providing an appeal and mediation process. Works may be quarantined during review, but contributors will not be penalized without due process.

6.3 Building Trust through Auditability and Transparency

Auditability is both a governance tool and a legal necessity. Every contribution, license, and royalty flow in CORPUS is logged in an append-only, tamper-evident registry, ensuring past records cannot be altered without detection. These logs provide the evidentiary basis for compliance with EU law – for example, DSM Directive Article 4 on text- and data-mining opt-outs and AI Act Article 52 transparency obligations – while also giving contributors confidence in the transparency of corpus management.

Looking ahead, CORPUS will also support embedding attribution metadata into model outputs – for example, through ISCC codes or provenance hashes. This ensures links between training data and generated material remain visible across the ecosystem. A game studio or mobility partner, for instance, could verify the provenance of a licensed model before integration into its engine or device.

In sum, CORPUS treats security and rights management not as features but as the foundation of its licensing protocol. With append-only audit logs, explicit contributor consent, and defensible verification procedures, it builds an infrastructure that can withstand regulatory scrutiny, protect creators, and give industry partners confidence in the corpus's legitimacy.

7 From contributions to Royalties: How Value flows in CORPUS

The attribution system translates contributions into stable scores that determine long-term royalty flows. Scoring happens once — at the moment a contribution enters CORPUS — and remains fixed except for metadata improvements or fraud/error corrections. This creates clear incentives. But the deeper function of the scoring system is to encode a specific economic logic: one in which the value of each contribution is defined by its relationship to the whole, not by its performance in a market of discrete products.

7.1 Incentive logic: rewarding Quantity, Quality, and Diversity

The attribution system translates contributions into stable scores that determine long-term royalty flows. Scoring happens once – at the moment a contribution enters CORPUS – and remains fixed except for metadata improvements or fraud/error corrections. This design creates clear and predictable incentives.

Three dimensions determine a contribution's weight:

- **Quantity**: every contribution expands the breadth of the corpus. A baseline score ensures participation itself is rewarded.
- **Quality**: works must meet technical and descriptive standards. Clean recordings, accurate metadata, and consistent annotation raise weighting.
- **Diversity**: contributions that expand the corpus into underrepresented areas earn extra weight. Originality is measured through relational analysis, not by surface-level tags: the system evaluates how a work positions itself relative to existing data points, capturing structural, stylistic, and cultural distinctiveness. This rewards genuine expansion rather than saturation, and helps counter historical biases such as the overrepresentation of Western commercial music in training datasets.

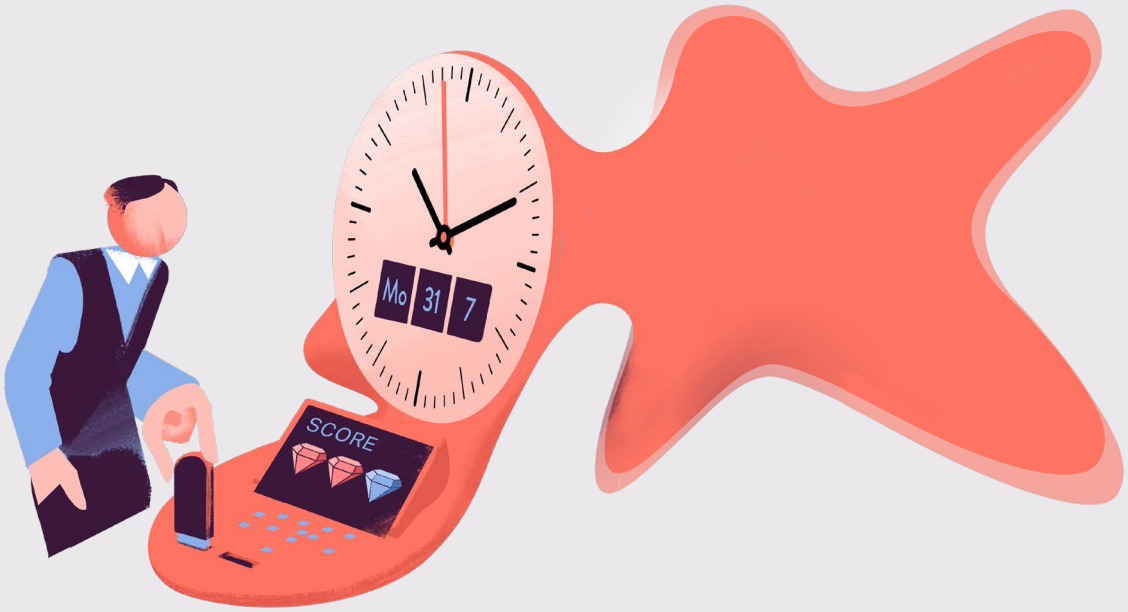


Figure 10: The contribution of each work is assessed at the moment it enters the corpus, and an input weighting is assigned based on how it enriches the library in terms of quantity, quality, and diversity.

Technical attribution structure

Figure 11: Baseline Contribution Scoring

• **Quantity: Base points for contribution**
Each work receives a baseline score depending on input format and completeness. For example, an uncompressed WAV may earn 100 points, an MP3 50, with additional points for stems, MIDI files, or detailed metadata.

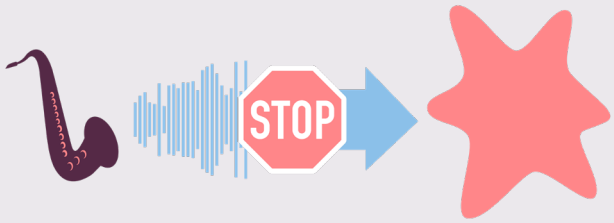


Figure 12: Contribution Integrity Checks

• **Quality: Automated filters and edge cases**
Non-musical content, illegal samples, or corrupted files are excluded. Filters may include spectral anomaly detection, provenance hashes, similarity searches against databases of copyrighted songs, and AI-detection tools. Flagged works earn zero points until reviewed, with appeal options for contributors.

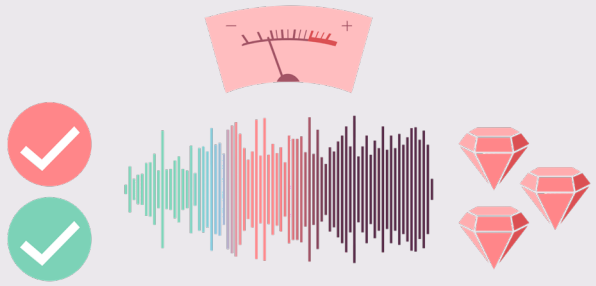


Figure 13: Production Quality Assessment

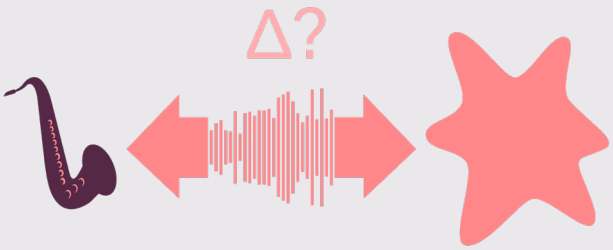


Figure 14: Relational Originality Scoring

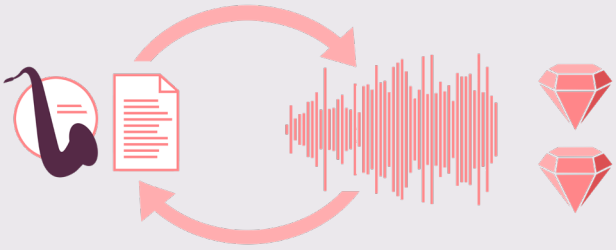


Figure 15: Metadata as First-Class Contribution

• **Quality: Production quality**
An evaluation system using Music Information Retrieval (MIR) benchmarks assesses production quality (spectral balance, dynamic range, noise levels, stereo field). High-quality recordings can earn bonus points; flawed material may earn none or be excluded. To avoid frustration, contributors can revise or resubmit flawed works, supported by automated feedback that highlights specific issues.

• **Originality: Diversity and Novelty**
Contributions are compared against the existing music library at the moment of ingest. Works that fill gaps or introduce new material earn additional points, while oversaturated areas are weighted lower. Originality is never judged in absolute terms but always in relation to what already exists. Crucially, this extends beyond genre tags: CORPUS applies musical analysis that detects distinctiveness in structure, timbre, and style across genres. Originality can thus arise from unusual harmonic language, experimental production, or cultural specificity just as much as from genre novelty. Because the library evolves, the diversity component is subject to gradual temporal recalibration, described in Section 7.5.

• **Annotation and metadata contributions**
Annotations – such as genre, mood, instrumentation, or cultural context – are a critical part of corpus value. Musicians can add or refine these for their own works, but others may also contribute, either voluntarily or incentivized. All verified annotation work is rewarded with points, recognizing that metadata quality is as vital to training as the recordings themselves.

Together, these mechanisms discourage mass-uploading and reward contributions that genuinely expand the corpus. Originality becomes progressively harder to achieve as the dataset grows — a dynamic that pushes contributors toward new territory rather than repetition.

This inverts a central assumption of the streaming economy. There, algorithms promote what resembles what has already succeeded; economic gravity pulls the entire catalogue toward a narrowing center. Difference is a disadvantage. In CORPUS, difference is the engine. An Anatolian bağlama recording, a Tuvan throat-singing session, an experimental prepared-piano improvisation — these are not niche curiosities. They are high-value assets, because they expand the expressive range of every model trained on them. Individual authenticity and collective capability reinforce each other: originality strengthens the whole, and the whole gives each contribution a reach it could never achieve alone.

7.2 Translating Points into Revenue shares

Contributor points translate into revenue shares through a transparent two-step process.

- **Revenue pool and costs**
  All revenue from sublicensing and from CORPUS-owned products and tools flows into a central pool. A defined share covers infrastructure, legal, staff, R&D, and reserves for growth. The remaining net revenue forms the royalty base – ensuring contributors are paid transparently while CORPUS retains resources to scale and create long-term value.

- **Model-level attribution**
  Royalties are tied to the training stage, not to individual outputs. Contributors are paid when their works form part of the dataset for a monetized model. For each distribution round:

  - Points per contributor are summed across all works included in that model.
  - A global point total is calculated across all contributors in the model.
  - Revenue is allocated proportionally.

- **Example**
  If Model A generates €1M net revenue, and Contributor X holds 0.5% of the model's total points, they receive €5,000.

Works that score highly but are not yet included in a monetized model remain eligible for future rounds. As new models are trained, these contributions may begin generating revenue,

ensuring that value can be realized over time rather than only at the moment of upload.

7.3 Oversight and Dispute handling

Trust in the protocol requires more than formulas. CORPUS builds governance and transparency into payout operations so contributors can verify how their works generate value.

- **Contributor dashboard**: each contributor has a dashboard showing which of their works are included in monetized models and which remain pending for future eligibility. It also displays point balances, royalty shares, and payout history, making attribution and revenue flows visible in real time.
- **Distribution round**: initial payouts may be annually or quartely, with frequency increasing as revenue grows. Small balances roll over to reduce transaction costs.
- **Payout processing**: contributors complete Know-Your-Customer procedures before payments; each payout is accompanied by an invoice or credit note that the contributor can use in their own tax declaration.
- Auditability: all financial flows are logged in tamper-evident records. Contributor-elected boards will review reports and audit trails, with the option for external audits as the system matures.
- **Dispute resolution**: contributors can appeal point allocations, fraud flags, or payout calculations. Governance mechanisms define escalation steps and mediation procedures, ensuring disputes are handled fairly and transparently.

Beyond formal audits, CORPUS will establish participatory feedback loops. These may include contributor surveys, score simulations, and public comment periods, giving musicians and partners a direct role in testing attribution rules and refining payout practices. This ensures that oversight is not only top-down but actively shaped by the community that powers the corpus.

7.4 Balancing Complexity with Usability

Any system that balances fairness with resistance to abuse will sound complex. The danger is opacity: rules pile up until even insiders no longer understand them, as seen in tax codes or legacy royalty schemes.

CORPUS addresses this tension in two ways. First, complexity exists only where it increases fairness or prevents abuse – not where it obscures flows. Second, interfaces absorb the

complexity. Contributors don't need to understand formulas (but they can if they want to); they see a clear breakdown in their dashboards: which works, which models, which revenues, which payouts.

Community input will remain central. Contributors will test, debate, and refine attribution rules to keep the balance between fairness and usability.

7.5 Beyond Royalties: Ownership and Long-term Participation

The royalty system described above distributes revenue. But distribution alone does not make contributors participants in the system they power. Streaming demonstrated this asymmetry starkly: Spotify's market capitalization exceeded $80 billion while the musicians who supply its catalogue debate whether streaming income covers their recording costs. The revenue flows — but the value accumulates elsewhere.

CORPUS is designed to address this structural problem. From its inception, the project has pursued the principle that contributing to the corpus should generate not only income but ownership — a stake in the system itself, not just a share of its periodic revenues.

The dual currency

This leads to a dual economic structure:

The first layer is **revenue participation** — the royalty system described in the preceding sections. Contributors receive ongoing payouts in conventional currency, proportional to their input weighting, whenever models trained on their works generate licensed revenue. This is the transactional layer: predictable, auditable, and tied to real economic activity.

The second layer is **system participation**. Each contribution generates not only royalty points but participation rights — provisionally designated CRPS (Corpus Participation Rights) — that represent a stake in the value of the shared corpus itself. Unlike royalties, which reflect current income, CRPS reflect the contributor's role in building a long-term asset. As the corpus grows, as more models are trained, as more industries license those models, the value of the system increases — and CRPS holders participate in that appreciation.

This dual structure aligns incentives in a way that pure revenue sharing cannot. A contributor who receives only royalties is a supplier. A contributor who holds CRPS is a co-owner. The difference is not semantic; it determines whether the system reproduces the extractive logic of existing platforms or genuinely implements the part-whole relationship described earlier — where originality strengthens the shared resource, and the shared resource gives each contribution lasting value.

What CRPS represent

CRPS are designed to carry two dimensions of participation:

Economic participation: a claim on the long-term value of the corpus – including, in a liquidity event, the appreciation of the underlying asset. This is the critical difference to streaming: contributors participate not only in the flow of revenue but in the stock of value their contributions help build.

Governance participation: the right to influence how the corpus is managed, how licensing terms evolve, and how the protocol develops. This can take the form of voting rights, representation on advisory bodies, or veto mechanisms on decisions that affect contributor interests.

Early contribution, temporal dynamics, and the separation of flows

The scoring system's diversity bonus creates a natural time dynamic. In the early phases of corpus assembly, the novelty premium is highest: every contribution is more likely to fill a gap. As the library grows, achieving high originality scores becomes progressively harder. Early contributors therefore accumulate more CRPS relative to later ones – a structural reward for trusting the system before its value is proven.

This creates a tension that the protocol must address openly. Early accumulation of participation rights is necessary to incentivize the Contributors who take the greatest risk. But left unchecked, it reproduces the dynamic that concentrates value at the top of existing platforms.

CORPUS resolves this through a separation of the two currencies:

CRPS do not decay. Participation rights, once issued, are permanent. They represent the historical fact of having built the system – a fact that does not change regardless of how the corpus evolves. A contributor who was the first to bring Tuvan throat singing into the library performed a structural function that shaped the models trained in that era. That function does not expire.

The diversity component of the royalty score does decay. A contribution's diversity bonus is assessed relationally – against the existing library at the moment of ingest. But the library evolves. A track that filled a significant gap in 2026 may occupy well-represented territory by 2031. Its quality has not diminished. Its function within the current system has shifted.

To reflect this, CORPUS introduces a gradual temporal recalibration of the diversity component. For the first five years after ingest, the diversity bonus remains unchanged.

After this period, it decays asymptotically toward a permanent floor of 30% of its original value – never reaching zero. The decay is exponential: after ten years, approximately 70% of the original bonus is retained; after twenty, approximately 37%. The precise parameters – protection period, decay rate, and floor – are governed by the scoring oversight bodies and can be adjusted in response to corpus dynamics.

This separation is the key design decision. Royalties reflect a contributor's evolving role in the current system. CRPS reflect the lasting fact of having built it. One is dynamic. The other is permanent. Together, they ensure that early contributors are rewarded for their risk without permanently capturing a disproportionate share of ongoing revenue.

Legal implementation: current evaluation

The legal form of CRPS is under active evaluation. Several pathways are being assessed, each with distinct trade-offs:

Profit participation rights (Genussrechte) under German law offer a well-established mechanism for economic participation without altering the ownership structure of the entity that holds the corpus. They can include profit participation and participation in a liquidation event, and they can be structured as transferable. Critically, when issued as consideration for a contribution – not as a gift – they do not trigger gift tax, resolving a key concern in earlier design phases.

Tokenized participation rights under Swiss DLT legislation (the Swiss Ledger-based Securities framework, effective since 2021) offer a regulated pathway for issuing CRPS as digital securities with full legal recognition. This would enable broader transferability and international scalability, but requires a Swiss entity or subsidiary and FINMA-compliant issuance.

Cooperative membership remains conceptually closest to the vision – every contributor becomes a member – but current cooperative law limits transferability and may not accommodate the speed and flexibility required in the AI sector. A hybrid approach, where cooperative membership provides governance while Genussrechte provide economic participation, is also under consideration.

No final structure has been selected. The decision will depend on regulatory developments, contributor feedback, and the outcome of ongoing legal consultations. What is fixed is the design principle: contributing to CORPUS must generate lasting participation, not merely periodic payment.

8 Organizational Structure and Legal Evolution

8.1 From Platform to Protocol

CORPUS faces a structural decision that will shape its long-term role in the music-AI ecosystem. The question is not simply how to organize a company. It is whether CORPUS becomes a platform – a centralized entity that holds data, builds models, and sells products – or a protocol: an open infrastructure that others can build on, with CORPUS operating specific commercial services on top.

The difference is consequential. A platform concentrates control. A protocol distributes it. A platform captures value at the center. A protocol enables value creation at the edges. For a system that claims to reverse the extractive logic of the existing music economy, the choice is not neutral. The architecture must be consistent with the argument.

CORPUS is pursuing a three-layer design that resolves this tension.

8.2 Layer 1: The Open Protocol

The CORPUS protocol – the rules for how contributions are evaluated, how licenses are structured, how provenance is tracked, and how participation rights are allocated – is designed to be transparent, auditable, and ultimately open. Openness is the precondition for contributor trust, and contributor trust is the precondition for the corpus to grow. If the rules by which music is evaluated and compensated are opaque, the system cannot credibly claim to serve contributors.

Concretely, this means the scoring methodology, the audit framework, and the data standards for contributions will be published and independently verifiable.

8.3 Layer 2: Controlled Data Access

The corpus – the music itself – requires a fundamentally different approach. Raw audio files, once released, cannot be recalled. Traditional dataset licensing, where audio files are shipped to a licensee, offers no structural protection against misuse.

CORPUS is evaluating federated learning as the primary mechanism for dataset access. In this model, the data never leaves CORPUS infrastructure. A licensee – for example, an automotive OEM developing an adaptive sound model – submits its model architecture to a secure training environment hosted on CORPUS servers or within hardware-secured enclaves (such as NVIDIA Confidential Computing). Training is executed within this environment. Only the resulting model weights are exported. The original audio files are never exposed.

This resolves the core tension between openness and protection. The protocol is open; the data is controlled – technically, not only contractually. Federated training also simplifies audit and attribution: every training run is logged with its subset, model, and licensee, creating the traceability that both contributors and regulators require.

8.4 Layer 3: The Proprietary Semantic Pipeline

CORPUS's ability to generate high-resolution semantic descriptions of music – covering narrative function, emotional arc, structural dynamics, timbral character, and contextual fit – is proprietary intellectual property and the primary source of competitive differentiation. It is what makes the training data valuable beyond its raw audio content.

The three layers are distinct in their logic: the protocol is open because trust requires transparency. The data is controlled because protection requires enforcement. The pipeline is proprietary because innovation requires incentive.

8.5 Governance: From Lightweight to Institutional

The most consequential design question is not which governance structure CORPUS adopts, but when. A foundation with a Golden Share – the model currently under evaluation with legal counsel – provides robust long-term protection. But a foundation requires substantial capital to establish, and its governance mechanisms only become meaningful once there are significant revenue flows to govern. If CORPUS waits for the foundation to be operational before implementing contributor governance, the most formative decisions – scoring calibration, licensing terms, CRPS design – will have been made without contributor input.

The protocol must therefore govern itself before the institution is ready.

CORPUS is investigating a phased approach in which governance begins lightweight and becomes progressively more institutional as the system scales. Three mechanisms are under evaluation, in sequence:

Phase 1 – Governance pilot (from beta onward). Before CORPUS can implement formal governance, it must learn what contributor governance actually looks like in practice. Which decisions benefit from collective deliberation? What formats produce meaningful input rather than opinion noise? How does a diverse, global contributor base engage with questions about scoring, licensing, or system design?

CORPUS plans to begin with a structured pilot: a small contributor panel, selected by stratified lottery (sortition) from the contributor base — stratified by region, genre cluster, catalog size, and time of entry to ensure no single demographic dominates. Sortition rather

than election is deliberate: elections in networks with power-law distributions reproduce those distributions. Lottery breaks this topology.

The pilot's first and most consequential application is scoring calibration. Whoever defines "quality" and "originality" defines which music becomes economically visible. Rather than delegating this to algorithms with occasional community feedback, the panel would review scoring outcomes in aggregate — examining whether the system produces systematic biases, and recommending adjustments to scoring dimensions or their weighting. If, for example, lo-fi field recordings from West Africa consistently score low on production quality metrics that encode a specific studio ideal, the panel can flag this for recalibration. The temporal recalibration parameters described in Section 7.5 — protection period, decay rate, and floor — would also fall within its mandate.

But scoring is not the only domain. The pilot will test contributor deliberation on questions such as: which categories of downstream use are acceptable? How should the distribution ratio between infrastructure costs and the royalty pool evolve? What transparency standards should apply to licensees? The goal is to discover which governance questions are suited to deliberative panels, which require specialist expertise, and which can be resolved through simpler mechanisms.

The panel does not modify the algorithm. It sets constraints within which the algorithm optimizes. Panel members would be compensated from the revenue pool, recognizing governance as a contribution on par with annotation or metadata work.

Phase 2 – Dual-track governance (post-launch). As commercial activity grows, CORPUS must serve two radically different time scales. Licensing negotiations, partner onboarding, and API pricing require decisions in days. Scoring philosophy, CRPS design, and data usage boundaries require deliberation over months.

A single governance body for both is either too slow for the first or too fast for the second. CORPUS is therefore evaluating a dual-track structure:

An executive track – the commercial entity – operates with a small professional board that can make operational decisions in real time, within clearly defined parameters. It may act, but it may not change the parameters themselves.

A constitutional track – composed of the scoring jury, external advisors (in law, AI ethics, musicology), and a representative of the institutional guardian – sets the parameters. It convenes quarterly and decides on scoring dimensions, CRPS terms, licensing principles, and the distribution ratio between infrastructure costs and the royalty pool.

The boundary between the two tracks is the critical design element. It must be enforced

technically – as a protocol constraint – not only contractually. The executive track cannot issue a license that violates constitutional principles. The system enforces this by design.

Phase 3 – Foundation as institutional safeguard (at scale). Once revenue flows cross a threshold that makes institutional governance both necessary and financially viable, CORPUS plans to establish a dedicated foundation as the long-term guardian of the protocol and of contributor interests.

In consultation with legal advisors, a structure is being evaluated in which:

A dedicated foundation (Corpus Stiftung) would own 100% of the entity that holds the corpus, and hold a Golden Share – a veto right over strategic decisions – in the commercial entity. The foundation has no shareholders and exists solely to fulfill its purpose. Its governing council (Stiftungsrat) would include representatives elected by the contributor community.

An IP entity (Corpus IP GmbH), fully owned by the foundation, would hold all music contributions, manage the dataset, execute scoring and attribution, and administer CRPS. Contributors' legal relationship is with this entity.

A commercial entity (Corpus Marketing GmbH) would develop and operate the AI technology, the semantic pipeline, and all commercial products. It licenses access to the corpus from the IP entity. This is where external investment is placed and where commercial speed is required. The Golden Share ensures that commercial pressure cannot override contributor protections.

This separation of concerns – trust (foundation), data (IP entity), commerce (marketing entity) – mirrors the three-layer technical architecture. But the foundation is not the starting point. It is the destination. The scoring jury and dual-track governance ensure that contributors have real influence from day one – not from the day the foundation is funded.

8.6 What Remains Open

The precise legal form of CRPS – whether as Genussrechte under German law, tokenized securities under Swiss DLT legislation, or a hybrid of cooperative membership and participation rights – is being evaluated in parallel with the institutional structure (see Section 7.5).

The scope and timeline of protocol publication – which components are opened first, under what license, and how independent verification is organized – will be defined in consultation with the contributor community and technical partners.

The federated learning infrastructure – including the choice of secure compute providers, the design of the training API, and the audit logging architecture – is in active development

and will be piloted with initial industry partners.

No structure is final. But the direction is clear: CORPUS is building an institution whose architecture enforces the values its protocol describes – where openness, protection, and commercial viability are not competing priorities but complementary layers of a single design.

9 From Artistic Desire to Legal Protocol

CORPUS began not as a rights-management concept, but from a musical desire: to enable experiences that traditional methods could never realize. The protocol described here is the legal and technical foundation for that vision.

The path ahead is complex. Many features described here – from automated quality evaluation to community governance – will take years to develop, test, and refine. But the direction is clear: assembling the dataset under provisional licenses, piloting real-world applications with industry partners, and refining the attribution system with community input.

During an industry event, a lawyer from Buma/Stemra listened to a short explanation of CORPUS. His reaction was immediate: “The logic is compelling,” he said with a smile, “but I wouldn’t want to be the one implementing it.” The remark captures the core challenge: entrenched systems resist transformation, even when the rationale is clear. CORPUS is designed precisely for this task – to turn compelling logic into workable infrastructure.

Our goal is to bring together allies – contributors, investors, industry partners, and institutions – who share a conviction: that the coming transformation of music is not a threat to be managed but an opportunity to recover something that was lost.

For most of its modern history, the economic system of music could see only one dimension of its value: the transactional. It counted copies, tracked plays, cleared licenses. It built an immense global industry on this single axis – and rendered invisible the dimensions that had sustained music for millennia before the first recording. The capacity of music to create shared meaning. To resonate with context. To generate experiences that exceed what any individual participant could have designed.

These capacities are not historical artifacts. They are the foundation of the markets now emerging: adaptive sound in vehicles, therapeutic music in healthcare, responsive environments in games and architecture, semantic interaction in robotics. In these domains, music does not function as a product. It arises as a situated, emergent phenomenon. And

for the first time, we can build economic infrastructure that makes this kind of value visible, traceable, and sustainable.

If we succeed, CORPUS will demonstrate that an economy built on shared contribution – where originality strengthens the whole, and the whole gives meaning back to the individual voice – is not idealism. It is better engineering.

9.1 Next Steps

CORPUS is currently moving from its closed alpha stage into a public beta phase. Since December 2025, selected contributors have been onboarded on an invite-only basis to test the platform's submission and attribution workflows. A broader rollout as a public beta is planned for March 2026.

In parallel, Sofilab is developing REEF, CORPUS's real-time adaptive model for interactive sound applications, and is establishing partnerships for industrial pilot projects in sectors such as automotive, healthcare, and music technology. To accelerate corpus growth, collaborations with existing music libraries, labels, and archives are being prepared.

The legal and institutional architecture described in Section 8 – including the foundation structure, the federated learning infrastructure, and the legal form of CRPS – is under active development with external legal and technical advisors. Initial design decisions are expected in the course of 2026.