Ensuring Traceability and Citation of AI-Generated Sound: A Minimal Metadata Framework for Cultural Heritage and Creative Industries

Prof. Dr. Ángel Manuel OLMOS
Professor of Musicology – Real Conservatorio Superior de Música de Madrid

angelmanuel.olmossaez@educa.madrid.org

ARTinRARE COST Action CA23158

Accademia di Belle Arti di Roma. 25/09/2025

Abstract

The rapid diffusion of large-scale generative audio models is reshaping musical creation, distribution, and scholarship, yet it simultaneously erodes the provenance chain that underpins intellectual credit, legal compliance, and long-term preservation. Current identification standards—ISRC, ISWC, DDEX-ERN, C2PA—were devised for human-authored or fixed-media works and provide, at best, partial coverage of algorithmic authorship, training data lineage, or model-specific parameters. This presentation introduces **MS-AIS (Minimal Set for AI-Sound)**, a lightweight, interoperable metadata schema designed to restore traceability and enable unambiguous citation of AI-generated sonic artefacts across artistic research, commercial catalogues, and memory institutions.

Employing a combined methodology of normative gap analysis, stakeholder interviews (creators, labels, archives), and pilot implementation in Iberian sound repositories, we isolate eight mandatory data points—persistent identifier, acoustic fingerprint, model/version, training corpus reference, prompt/seed synopsis, human operator(s), generation timestamp/location, and licence status—supplemented by optional ethical and technical descriptors.

MS-AIS aligns with FAIR principles, dovetails with existing PID infrastructures (DOI, Handle), and can be serialised in JSON-LD or embedded within Broadcast Wave Format extensions, ensuring compatibility with both scholarly and industry workflows. We conclude by outlining a sectoral adoption roadmap and inviting collaboration toward a formal standard within the COST *Artistic Intelligence* Action. By operationalising transparency, the proposed framework safeguards cultural heritage, fosters responsible creative AI, and equips policymakers with a practical lever for evidence-based regulation.

Table of contents

Introduction....3
Context & Motivation....3

Problem statement: Scope and Limitations of ISRC, ISWC, DDEX-ERN, and C2PA for AI-Generated Music....4
ISRC (International Standard Recording Code)....4
ISWC (International Standard Musical Work Code)....4
DDEX-ERN (Electronic Release Notification)....5
C2PA (Coalition for Content Provenance and Authenticity)....6
Methodology....7
Approach Overview....7
Normative Gap Analysis....7
Stakeholder Interviews....11
Stakeholder selection and representativeness....11
Composition of stakeholders (N = 59)....12
Brief rationale for quotas....13
Thematic saturation....13
Pilot Implementation....13
Results — Group-level synthesis of stakeholder interviews....15
1) Creators (artists/producers) — transparency with control, credit, and low-friction capture 15
2) Independent labels/distributors — compliance, risk signaling, and DDEX alignment....15
3) Archives & memory institutions — durable provenance, authority control, and ethical access....16
4) Cross-cutting consensus and points of tension....17
5) Priority matrix for MS-AIS fields by stakeholder group....17
6) Implementation priorities derived from interviews....18
Limitations and reflexivity....18
Proposed Framework: MS-AIS (Minimal Set for AI-Sound) Metadata Schema....18
Alignment with Standards and Interoperability....19
FAIR Principles (Findable, Accessible, Interoperable, Reusable)....19
Integration with PID Infrastructure (DOI/Handle)....19
Technical Compatibility (JSON-LD and Broadcast Wave)....20
A. JSON-LD serialization (web interoperability)....20
B. Broadcast Wave embedding (industry and archival workflows)....21
C. Profiles and crosswalks....21
Adoption Roadmap and Standardization....21
Sectoral Adoption Steps....21
Community Collaboration (towards formal standardization)....23
Phased Implementation (from early pilots to broad recognition)....24
Conclusion....25
Bibliography....26
Appendix A — Semi-structured interview protocol (content)....28
Appendix B — Participant roster & per-interview key findings (concise)....30
B.1 Creators (Artists/Producers) — N = 30....30
B.2 Independent Labels/Distributors — N = 14....32
B.3 Archives & Memory Institutions — N = 15....33
Appendix C: List of acronyms and abbreviations....35

Introduction

Context & Motivation

Large-scale generative audio models have rapidly emerged as influential tools in music creation, distribution, and research. In the past few years, research on AI-driven music generation has seen *"considerable attention and growth"* (Lerch et al., 2025), fueled by advances in deep learning and the availability of massive audio datasets. Tech industry leaders and academic labs alike have introduced powerful music generators – for example, Meta's MusicGen was trained on 20,000 hours of audio to produce new songs from text or melody prompts (Wiggers, 2023). Likewise, Google's MusicLM and OpenAI's earlier Jukebox model demonstrated that AI can create high-fidelity musical pieces in various genres from simple descriptions. These systems are increasingly being integrated into creative workflows and commercial platforms (Berardinis et al., 2025). Musicians and hobbyists now experiment with AI co-composers, and entirely AI-generated tracks have begun appearing on streaming services and social media. A striking example occurred in 2023, when an AI-generated song mimicking the vocals of Drake and The Weeknd went viral online – it racked up millions of streams before being pulled from Spotify, TikTok, and YouTube due to copyright complaints (Snapes, 2023). Such incidents underscore how quickly generative music technology has moved from the lab to widespread public use, blurring the lines between human and machine creativity in music distribution.

However, this trend also erodes the provenance chain of musical works, raising serious concerns for intellectual credit, legal compliance, and long-term preservation. In traditional music production, it is usually clear who created a piece and what source material was used, forming a traceable provenance chain (authorship and ownership history). By contrast, generative models operate as black boxes: once vast libraries of songs are ingested into a model, *"all of [the source metadata] is lost, you can't trace back the original"* in the outputs (Bulger et al., 2024, p. 410). The lack of transparent source attribution means AI-generated music often arrives devoid of context about which artists or works influenced it (Berardinis et al., 2025). This undermines the ability to credit original creators – indeed, proper recognition of artists' contributions in AI-generated music is *"critical, yet often neglected" (Choi et al., 2025)*. For instance, the viral Drake/Weeknd mimicry not only violated copyright, but also threatened to deny those artists their *"due compensation"* and recognition. More broadly, researchers have warned that generative models can inadvertently reproduce copyrighted material, posing risks of IP infringement. With no provenance data, it becomes difficult to ensure an AI-produced song is legally compliant or to determine who should be paid royalties if it significantly draws on someone's style or content. This opacity also complicates preservation and authenticity in the long run – archives and music libraries rely on provenance metadata to catalog works, attribute authorship, and maintain cultural history. When music is generated without a clear lineage or credited origin, it challenges archivists' ability to preserve the work's context and to authenticate it for future generations. In summary, the rise of generative music AI creates exciting new possibilities, but it also breaks the traditional chain of provenance that underpins the music ecosystem's intellectual property norms and memory. Addressing this gap is crucial to ensure that innovation in music AI proceeds hand-in-hand with attribution, legal integrity, and the long-term stewardship of musical works (*Musical AI - Our Manifesto*, n.d.).

Problem statement: Scope and Limitations of ISRC, ISWC, DDEX-ERN, and C2PA for AI-Generated Music

As artificial intelligence becomes increasingly involved in music creation, questions arise about whether existing music identification and metadata standards can capture the unique challenges of AI-generated content. The standards examined here – ISRC, ISWC, DDEX-ERN, and C2PA – were established to serve traditional music industry needs. We summarize each standard's scope and purpose and highlight their limitations in accounting for *algorithmic authorship*, *training data provenance*, and *model-specific parameters*. These aspects of AI transparency, data lineage, and rights attribution were largely outside the design considerations of these standards.

ISRC (International Standard Recording Code)

The **ISRC** is an international code defined by ISO 3901 and administered by the International Federation of the Phonographic Industry (IFPI) to uniquely identify sound recordings and music videos. An ISRC is a 12-character alphanumeric code that serves as a permanent identifier for a specific recording, regardless of where or how that recording is distributed (*Home — International Standard Recording Code*, n.d.). Its primary purpose is to avoid ambiguity among recordings and to simplify rights management across different formats, services, and licensing deals. Once assigned (typically by the recording's rights owner), an ISRC stays with that recording for its lifetime, enabling efficient tracking of uses for royalty payments, usage reporting, and catalog management (International ISRC Registration Authority, 2021, p. 5).

Limitations for AI-Generated Music: The ISRC standard is focused narrowly on identifying a *recording* and carries no information about how that recording was created or who (or what) created it. In fact, the IFPI explicitly notes that *ISRC identifies sound recordings and music videos*, and *is not used to identify compositions, musical works, products, or performers*. This means that any details about the creative process or authorship – human or algorithmic – are outside ISRC's scope. The code itself encodes a country and registrant code, year of reference, and a designation number, but it does not include any metadata about the songwriter, producer, or the method of creation. Consequently, ISRC has no mechanism to indicate algorithmic authorship or AI involvement. It treats an AI-generated recording the same as any other recording for identification purposes. There is also no provision for documenting training data or model parameters in an ISRC; those details are simply not part of what an ISRC is designed to capture. In summary, ISRC provides a unique identifier for the recording, but it offers zero transparency about the recording's origin (e.g. whether it was created by a human artist or an AI system) or the creative process behind it.

ISWC (International Standard Musical Work Code)

The ISWC (ISO 15707) is a standard managed by CISAC for uniquely identifying musical works (i.e. the underlying compositions, as opposed to specific recordings). The ISWC system assigns each musical work a permanent code, which is used globally by composers, publishers, performing rights organizations, and others in the music value chain (*International Identifiers | CISAC*, n.d.). The ISWC helps standardize data for musical works and streamlines rights administration and royalty distribution on a worldwide basis. An ISWC identifies a musical work by linking it to its title and its credited creators (such as composers, lyricists, and arrangers). This allows different stakeholders to unambiguously refer to the same composition even if it is recorded or published in many forms.

Limitations for AI-Generated Music: By design, ISWC captures *who* wrote a piece of music (and what it's titled), but it was not designed to capture *how* the music was created. The standard assumes human authorship and doesn't provide a way to credit a non-human creator or an algorithm. In practice, if an AI system composes a piece of music, any ISWC registration would still require listing a creator name (often a human proxy or the owner of the AI) since the system has no concept of recognizing an algorithm or model as the "composer." Moreover, ISWC does not track any instance of how a work is used or produced. As CISAC's documentation states, ISWC *"does not… identify instances of use of the work in manifestations, such as publications, recordings or broadcasts"*. In other words, ISWC is blind to the work's instantiation and origin – it won't tell you if a song was generated by training a model on a dataset, or if it was written traditionally. There are no data fields in the ISWC system for recording training data provenance or AI model identifiers. The focus is strictly on the musical work and its (human) authors. Therefore, while ISWC is very effective for tracking ownership and royalties of compositions, it offers no built-in transparency about algorithmic composition processes or the lineage of creative material in an AI-generated work.

DDEX-ERN (Electronic Release Notification)

DDEX (Digital Data Exchange) is a consortium that develops standard message formats for the music industry. The Electronic Release Notification (ERN) standard is a family of XML message formats used for communicating detailed metadata about music releases from content owners (record labels or distributors) to digital service providers (DSPs) like Spotify, Apple Music, etc. ('Electronic Release Notification Message Suite', n.d.). An ERN message – particularly the core NewReleaseMessage – typically contains metadata about the release (album or single) and all the constituent resources (tracks, videos), including information such as titles, artist and contributor names, ISRCs for recordings, ISWCs for works, release dates, genres, and more. It also carries the terms and conditions under which the music can be made available (for example, territories, start/end dates, usage rights, and price tiers). The DDEX-ERN standard is quite comprehensive in scope: it allows multiple titles (e.g. different languages or abbreviations), localized metadata per territory, and even credits like producer, mixer, and engineer names to be included to enrich the release information (Isherwood et al., 2016, p. 22). By providing a common format, DDEX-ERN has greatly improved efficiency and accuracy in metadata exchange, reducing errors in royalty reporting and ensuring consistent data across platforms (*Metadata Standardization*, n.d.).

Limitations for AI-Generated Music: Despite its richness in traditional metadata, the ERN standard does not account for AI-specific provenance or authorship details. Its data model expects human-readable credits and identifiers that are standard in the industry (artist, songwriter, publisher, etc.). There are no fields in the ERN schema to declare that "this track was created by algorithm X" or that "it was trained on dataset Y." If a song is AI-generated, from the ERN perspective it will still be delivered with an ISRC, a title, an artist name (perhaps the name of the project or AI pseudonym), and potentially a composer name – but nothing in the ERN message would explicitly flag the track as AI-created or link it to the underlying model or training data. The omission is understandable, as DDEX standards were initially developed in the 2000s to address interoperability in digital music distribution, long before *generative AI in music* became a concern. Even as of recent versions, the ERN's focus remains on the *released product metadata* and licensing terms, not the creative process. In short, an ERN file can convey extensive information about a music release's commercial metadata and rights, but it provides no transparency about whether the content was generated by a machine learning model, nor any mechanism to include model parameters or training data lineage. (Notably, industry discussions are now emerging on how to extend metadata standards for AI content, but such extensions are still in development and not part of the established ERN specification.) (*Metadata Standardization*, n.d.)

C2PA (Coalition for Content Provenance and Authenticity)

The C2PA is a newer standard (spearheaded by a consortium including Adobe, Microsoft, BBC, and others) that addresses digital content provenance and authenticity across media types. Unlike the music-specific codes above, C2PA is a general framework to cryptographically bind metadata (so-called "content credentials") to an image, audio, video or document in a tamper-evident way (*C2PA and Content Credentials Explainer :: C2PA Specifications*, n.d.). The overarching goal is to help publishers and creators convey the origin and edit history of media to consumers, thereby combating misinformation and establishing trust in content. In C2PA's model, a piece of content can carry a *manifest* containing one or more assertions – statements about the content's provenance. These can include details such as *who* created it, *when and where* it was created, and *what tools or processes* were used in its creation or modification. Importantly, C2PA was designed with AI-manipulated media in mind: for example, it supports assertions about the *use of AI in content creation (i.e. how the content was authored)*. The specification even allows for indicating if the creator permits the output to be used for AI training in the future. All such assertions are digitally signed and can be verified to ensure they haven't been tampered with. C2PA provides a flexible infrastructure for content creators to voluntarily embed transparency information about the provenance of a media asset, including some AI-related context, as *metadata that travels with the content*.

Limitations for AI-Generated Music: C2PA is arguably the most relevant standard here for addressing AI transparency, yet it still has important limitations in the context of AI-generated music. First, adoption in the music ecosystem is not yet widespread – C2PA is an opt-in system, and if no Content Credentials are attached to a track, then none of this provenance information is available to the listener or platforms. Even when used, C2PA can record *that* a piece of audio was AI-generated and by which tool, but it does not inherently reveal the full *training data provenance* or *model parameters* behind that AI generation. For instance, a music file's C2PA manifest might include an assertion like "Generated by *XYZ Music AI* on 2025-09-10" and could optionally include the *prompt* or settings used for generation. It may also link to a content credential for the AI model itself, and list any source "ingredients" (input assets) that went into the model's output. This can improve transparency around the immediate provenance of the AI output (which model, which prompt, etc.). However, C2PA stops short of cataloguing the model's own history – it doesn't automatically tell you, for example, which 10,000 songs were in the training dataset of the model that produced the music, or what the model's hyperparameters or architecture are. Those deep details would only be available if the model provider chooses to publish them (potentially via a separate content credential for the model) and if the workflow links that to the music asset. In practice, C2PA's content credentials can encapsulate *some* AI-related metadata (authorship, tool names, usage rights), but they are not a full solution for AI lineage. The C2PA spec itself acknowledges it is "not a cure-all for misinformation" or a guarantee of truth – it provides a secure framework for recording claims about content, not verifying the underlying facts. While C2PA introduces a way to carry authorship and provenance information (including AI usage) with music files, it was not designed to, nor can it feasibly, embed the entire complexity of an AI model's training provenance or internal parameters. Significant gaps remain in using C2PA for exhaustive AI transparency: the standard can flag that a song is AI-generated and by whom, but it relies on voluntary disclosure and cannot automatically trace the full lineage of creative data behind AI music.

Each of these standards was created to solve specific identification and metadata needs in the content ecosystem, and none was originally intended to handle the nuances of AI-generated works. ISRC and ISWC were conceived in an era of human-created music and focus on identifying recordings and compositions (and their traditional rightsholders) – they contain no fields for *algorithmic creators*, *source datasets*, or *AI model IDs*. DDEX-ERN facilitates rich metadata exchange for music releases, but it, too, centers on conventional credits and rights; it has no provisions for encoding how a piece of music was created or the involvement of AI. C2PA brings the promise of content provenance tracking and can denote AI involvement at the content level, yet it is limited by voluntary implementation and does not inherently solve the problem of tracing an AI model's training data or internal workings. In the context of AI-generated music, these standards collectively fall short of providing transparency about data lineage and algorithmic authorship. New extensions or complementary frameworks will likely be needed to fill this gap, so that future music metadata can account for the role of AI in creation and ensure proper attribution and rights management in an increasingly AI-influenced music industry.

Methodology

Approach Overview

The development of the MS-AIS framework followed a multi-stage methodology that combined normative research with stakeholder engagement and practical testing. In summary, we first performed a normative gap analysis of existing metadata standards to pinpoint missing elements needed for AI-generated audio. Next, extensive stakeholder interviews were conducted with creators, music industry professionals, and archivists to gather requirements and validate that the proposed metadata would meet real-world needs. Finally, a pilot implementation was planned in collaboration with Iberian sound repositories to deploy the preliminary schema in practice. This iterative approach – from analysis, to user input, to field testing – ensured that the MS-AIS framework was both comprehensive in theory and grounded in practical applicability. Each of these stages is detailed below.

Normative Gap Analysis

Our first step was a formal review of current metadata standards and practices in the audio and music domain to identify gaps related to AI-generated content. We surveyed widely-used metadata schemas (e.g., ID3 tags, Dublin Core, Broadcast WAVE/EBU Core) and emerging guidelines for AI content labeling. This normative gap analysis revealed that while conventional metadata covers basic descriptive data (title, artist, ISRC code, etc.), it fails to capture the *creative process* or origin of a track (AI Music detection team, 2025). In other words, existing standards do not indicate *how* a piece of audio was produced – a track can appear entirely legitimate in metadata yet be fully machine-generated (possibly even trained on copyrighted material) with no disclosure. Without metadata to verify AI involvement, platforms and rights managers risk misidentifying AI-generated music as human-made, leading to potential copyright or attribution violations. This analysis underscored a clear gap: additional metadata is needed to represent the unique context and provenance of AI-created audio.

Through the gap analysis, we identified several critical metadata elements missing from current standards that are required for AI-generated audio. In particular, the framework determined the need for:

- **AI Generation Flag:** a field to explicitly mark whether the content was AI-generated or involved synthetic processes (a tag denoting "AI-generated" content).
- **Generative Tool Details:** metadata capturing which AI model or algorithm was used to create the audio (including model name/version or training data, where applicable), to provide transparency into the creation process.
- **Provenance and Rights Information:** data documenting the content's provenance and any intellectual property considerations – for example, indicators if the AI's training material included licensed audio, consent from rights holders, or other usage rights. This could also include cryptographic signatures or watermarks to certify authenticity and detect tampering.

These missing elements formed the basis of the MS-AIS schema extension. They align with recommendations in emerging AI transparency frameworks that call for embedding key "transparency" metadata and IP identifiers with AI-generated media (*TransparentMeta*, n.d.). In essence, the gap analysis provided a checklist of minimal data points (beyond traditional metadata) needed to responsibly describe AI audio content in compliance with evolving norms and regulations.

Metadata Category	ISRC	ISWC	DDEX-ERN	C2PA	Dublin Core	ID3	EBUCore
Authorship & Attribution *(e.g. creator identity, contributor roles)*	**Partial:** Designed primarily as an identifier, ISRC requires maintaining a **Main Artist** name with each code, but does not embed composer or detailed contributor data.	**Full:** The ISWC (work code) registration mandates listing *all* composers, authors, and arrangers with their roles (via IPI codes), ensuring comprehensive authorship attribution.	**Full:** DDEX ERN supports rich contributor metadata – a *DisplayArtist* composite with roles (MainArtist, FeaturedArtist, Composer, etc.) is provided, allowing full credit to creators and performers.	**Full:** C2PA content credentials can include authenticated author identity assertions, enabling explicit attribution of the creator in a cryptographically verifiable manner.	**Full:** Dublin Core's core elements include **Creator** (primary author) and **Contributor** for additional credits, facilitating basic authorship metadata in any content description.	**Partial:** ID3 tags offer standard frames for artist and author (e.g. **TPE1** for lead performer, **TCOM** for composer), covering primary attribution. However, role information is limited and less granular (no distinct fields for each role beyond generic text).	**Full:** EBUCore defines extensive creator/contributor fields – *Creator* for primary intellectual author and *Contributor* for others – with the ability to specify roles (author, performer, producer, etc.), providing robust support for attribution.
Provenance & Technical Lineage *(e.g. source materials, derivation history, editing/process steps)*	**No:** The ISRC standard does **not** capture content provenance or derivation. It identifies recordings uniquely, but has no mechanism to record if a track was derived or remixed from others (no lineage metadata).	**Partial:** ISWC can indicate if a musical work is a *version* or arrangement of another work, linking to the original composition. This covers intellectual lineage (song version history), but *technical* production lineage (e.g. audio editing or generation process) is not addressed.	**Partial:** DDEX ERN focuses on release metadata and rights. It links recordings to underlying works and allows some context (e.g. indicating remixer in contributor role), but it does not natively trace the step-by-step production history or source audio lineage. Provenance beyond basic relationships is outside ERN's scope.	**Supported/Optional (assertion-based)::** Provenance is a core focus of C2PA. Manifests can record **ingredients** (source assets) and tools used, building a verifiable history of how the audio was generated or edited. Each edit or AI generation step can be logged, providing a chain-of-custody for the content. However, disclosure is optional.	**Partial:** Dublin Core includes a general **Source** element to reference an original resource from which the current item is derived, and **Relation** qualifiers (e.g. *IsVersionOf*, *HasPart*) for basic lineage relationships. These allow one-level provenance links, but not detailed multi-step process metadata.	**No:** ID3 has no dedicated fields for capturing provenance or derivation history. There is no standard tag to denote an audio file's origin or prior versions (aside from a generic "source webpage" URL in v2.x, which is rarely used for lineage). In practice, any AI lineage or edit history would not be recorded in ID3.	**Partial:** EBUCore supports content relations to express lineage, using predefined relationship tags like **isVersionOf**, **hasVersion**, **isPartOf**, etc. to link media assets. This covers structural or version relationships (e.g. an excerpt or variant), but the schema does not inherently log detailed technical production history unless extended or combined with other metadata (e.g. no step-by-step edit log by default).
AI-Specific Metadata *(e.g. AI model name/version, generation prompt, training data reference)*	**No:** ISRC predates AI content and provides no fields for AI model or generative parameters. It simply identifies the recording and carries no information about how it was created (human or AI).	**No:** ISWC is limited to musical work identification and human authorship metadata. It has no capacity to note AI generation details (such as algorithm or prompt) for a composition.	**No:** Present DDEX standards do not define tags for AI model names, prompts, or dataset IDs. There is currently **no** support in ERN for AI generation specifics – these frameworks are only beginning to consider AI metadata needs. (Future DDEX updates are being explored to handle AI-generated content, but no official fields exist yet.)	**Supported/Optional (assertion-based):** C2PA includes AI provenance. It can capture the **model** used (via asset type assertions including model name/version), and prompt text. C2PA can carry AI assertions and tool traces, but disclosure is optional and adoption uneven; it does not ensure training-data lineage by itself	**No:** Dublin Core has no specialized elements for AI model or prompt information. While one could theoretically put an AI model name in a **Description** field, there is no standard or optional field in DC dedicated to AI-specific provenance.	**No:** ID3 tags do not cover AI generative details. There are no standard ID3 frames for storing the name of an AI model, the prompt used, or any training data reference. Such data would have to be embedded manually in comment fields, which is not standardized (hence effectively unsupported).	**No:** The EBUCore schema (as of current versions) has **no AI-specific metadata fields**. It does not natively include properties for recording the generative model or prompt. Any AI metadata would require non-standard extensions or external linking, as the core focuses on traditional media descriptors.
Legal & Licensing Information *(e.g. copyright, usage rights, licensing terms, rights holder)*	**Partial:** ISRC itself carries minimal rights info. It requires the **publication date (P-date)** of the recording (used for copyright term calculation) but	**Partial:** An ISWC identifies a work for royalty tracking and is used by Collectives in **licenses and usage reporting**, but the code's metadata does not include	**Full:** DDEX (ERN) provides extensive support for rights and licensing data. The standard communicates the **terms and conditions** under which a release may be	**Partial:** C2PA can include usage and licensing assertions, though its emphasis is on provenance. For example, an AI-generated asset's manifest may carry a **"do not train" usage restriction** (a data-mining rights assertion). It	**Full:** Dublin Core offers a general **Rights** element to specify a rights statement or license for the resource. Qualified Dublin Core adds **RightsHolder** to name	**Partial:** ID3 tagging supports basic copyright and licensing information. There is a **TCOP (Copyright)** frame for a text notice (often including year and owner)	**Full:** EBUCore has comprehensive legal metadata support. It includes fields for **rights management** information, the **rightsHolder** (entity owning or managing the rights), usage **constraints**

Metadata Category	ISRC	ISWC	DDEX-ERN	C2PA	Dublin Core	ID3	EBUCore
	does **not** embed details on rights holders or license terms. Licensing is handled via external registries and not encoded in the ISRC code or its basic metadata.	license terms or conditions. (It relies on publishers/CMOs to apply rights information; ISWC itself just links to the work's creators.)	used, including territorial availability, usage types, and other deal information. It also can convey rights holder identifiers and roles, and is often paired with detailed rights claim messages. In summary, licensing and usage rights metadata are integral to DDEX's design.	can also encapsulate copyright info or artist identity for rights (as shown in Adobe's Content Credentials usage). However, C2PA is not a licensing framework per se; it provides a vehicle to declare rights, but those are optional assertions rather than a fixed schema for licenses.	the owner of rights. This flexibility allows inclusion of copyright notices, Creative Commons license URLs, or any relevant legal text as part of the metadata.	and a **WCOP (Copyright/Legal Information URL)** frame for a link to a license or rights webpage. These allow an MP3 to carry a copyright statement and, for example, a Creative Commons license URL. Still, the detail is limited (no structured rights schema beyond free-text and URLs).	**or restrictions** (exploitation terms), **copyright notice** statements, temporal and geographic coverage of rights, a clearance flag (whether rights are cleared), and even contact info for rights administrators. This level of detail makes EBUCore well-equipped to represent licensing and rights data alongside the content description.

Stakeholder Interviews

To validate and refine these requirements, we engaged directly with the people who would create, manage, or preserve AI-generated audio. We conducted 59 stakeholder interviews in total, using an exploratory, semi-structured format. This allowed us to cover predefined questions about metadata needs while also letting participants raise additional insights. The interviewees spanned three key groups: music creators (artists and producers using and not AI tools), music industry professionals (such as record label and distribution executives), and archive and library specialists who manage sound collections. All interviews were conducted under confidentiality (we do not disclose specific individuals or institutions), encouraging participants to speak freely about their experiences and requirements.

The purpose of these interviews was twofold: requirement gathering and practical validation. First, we asked creators and industry professionals what information they deemed important when labeling AI-generated music – for instance, how they would want AI involvement to be credited or disclosed, and what data would help in rights management and attribution. Likewise, archive professionals were asked how they would preserve information about an audio file's origin and authenticity for future users. These discussions confirmed the importance of the metadata elements identified in the gap analysis (e.g. clearly flagging AI-generated works, documenting the AI tool used). Moreover, stakeholders helped prioritize which metadata elements were truly essential versus nice-to-have. This was crucial in keeping the MS-AIS schema as minimal as possible while still covering all practical needs. For example, creators emphasized the need for an "AI Creator" credit field, whereas archivists stressed long-term provenance metadata. The semi-structured format also surfaced real-world scenarios (such as managing an AI-generated remix in a music catalog, or preserving a synthetic speech recording in an archive) that informed how the metadata schema should be designed.

By the end of this phase, the interview feedback had validated the initial framework and also prompted minor adjustments. The MS-AIS schema was refined to ensure each metadata element was both meaningful and feasible to capture in real-world workflows. This stakeholder-driven validation gave us confidence that the framework would address actual user requirements and industry constraints, not just theoretical ideals.

Stakeholder selection and representativeness

We adopted a stratified purposive sampling strategy (with maximum-variation quotas) to ensure coverage across three primary stakeholder groups—creators (artists/producers), independent labels/distributors, and archives/memory institutions—while balancing gender, age, education, professional role, country of provenance, musical style, and AI exposure. This approach is standard in qualitative inquiry when the goal is to capture the breadth of positions and practice contexts rather than to generalize statistically.

Strata and quotas. We set *a priori* quotas for each group and for key diversity axes:

- **Group balance.** Creators were intentionally the largest stratum (reflecting their plurality in the field), complemented by labels/distributors and archives as decision-makers and stewards of rights and provenance.

- **Gender balance.** We targeted near parity across women and men, with explicit room for non-binary/other self-descriptions.
- **Age distribution.** Four brackets (18–29, 30–44, 45–59, 60+) to capture career stage and technology adoption differences.
- **Education.** Vocational/secondary, BA/Conservatory, and MA/PhD to reflect varied routes into music creation/management/preservation.
- **Country of provenance.** Emphasis on Spain and Portugal (given the project's Iberian anchoring), with a complementary Other EU stratum to avoid regional bias.
- **Musical style.** Coverage of Pop/Urban, Electronic/Experimental, Classical/Contemporary, Jazz/World/Traditional, and Sound art/Podcast/AV to reflect distinct production and cataloguing practices.
- **AI exposure.** Two dimensions were captured separately: usage frequency (None/Curious, Occasional, Regular, Advanced/Prototyper) and knowledge level (Basic, Intermediate, Advanced, Expert).

Recruitment and inclusion. Candidates were identified via professional associations, conservatories, artist residencies, independent label networks, archivist forums, and snowball sampling to reach under-represented profiles. Inclusion criteria required direct experience (past 24 months) with creating, distributing, or preserving digital audio and informed familiarity (even if basic) with AI-assisted workflows or their implications. Interviews were semi-structured, enabling comparison across strata while keeping space for emergent themes. All participation was confidential; institutional names are not disclosed.

Composition of stakeholders (N = 59)

Group (N)	Gender W/M/NB	Age 18–29/30–44/45–59/60+	Education Voc/BA/MA+	Country ES/PT/Other EU	Music Pop/Elect/Clas/Jazz/SoundArt	AI Usage None/Occ/Reg/Adv	AI Knowledge Basic/Interm/Adv/Expert
Creators (Artists/Producers) (30)	15/13/2	10/12/7/1	6/18/6	12/6/12	9/8/6/5/2	3/10/12/5	8/12/8/2
Independent Labels/Distributors (14)	7/6/1	2/7/4/1	2/8/4	5/3/6	5/3/2/2/2	4/5/4/1	6/5/3/0
Archives & Memory Institutions (15)	8/7/0	3/8/3/1	2/1/12	5/3/7	2/1/4/4/4	3/5/4/3	4/7/2/2
Total (59)	**30/26/3**	**15/27/14/3**	**10/27/22**	**22/12/25**	**16/12/12/11/8**	**10/20/20/9**	**18/24/13/4**

How to read the table.

- *Gender W/M/NB* = Women/Men/Non-binary (or self-described other).
- *Education Voc/BA/MA+* = Vocational or Secondary / BA or Conservatory / MA or PhD.
- *Country ES/PT/Other EU* = Spain / Portugal / other European Union countries.
- *Music Pop/Elect/Clas/Jazz/SoundArt* = Pop-Urban / Electronic-Experimental / Classical-Contemporary / Jazz-World-Traditional / Sound-art-Podcast-AV.

- *AI Usage None/Occ/Reg/Adv* = None or Curious / Occasional / Regular / Advanced or Prototyper.
- *AI Knowledge* levels reflect self-assessment corroborated during the interview warm-up.

Brief rationale for quotas

- **Creators (n=30)** are the most numerous and stylistically diverse; higher quotas maximize variance in AI practices (from prompt-based generation to hybrid studio workflows).
- **Labels (n=14)** bring rights, metadata exchange (e.g., DDEX), and catalog-risk perspectives; we ensured presence of small catalogue owners handling or expecting AI submissions.
- **Archives (n=15)** steward provenance and long-term preservation; the higher *MA/PhD* share reflects typical training in GLAM institutions.

Thematic saturation

We adopted a stratified purposive sampling design with maximum-variation quotas and conducted analysis iteratively alongside data collection to monitor saturation. Consistent with qualitative methodology, we distinguished code saturation (the point at which no new thematic codes emerge) from meaning saturation (when further interviews yield no additional nuance, depth, or dimensions to existing codes) and used both as stopping heuristics. Because prior work shows that code saturation can occur with relatively few interviews in homogeneous samples (often within the first dozen) but increases with sample heterogeneity, our cross-strata design (creators, labels/distributors, and archives; varied ages, genders, musical styles, countries, and AI exposure) justified a larger target ($N = 59$). We also followed the information power principle—sample adequacy depends on study aim specificity, sample specificity, theoretical anchoring, interview quality, and analytic strategy—which further supports our achieved size given the breadth of stakeholder perspectives and the framework-development objective. Taken together, these criteria provided a defensible basis for claiming adequate thematic coverage across strata while minimizing redundant collection.

Pilot Implementation

As a final methodological step, we planned a pilot implementation of the MS-AIS metadata schema in a real archive setting. The pilot was designed to deploy the preliminary schema in one or more Iberian sound repositories (audio archives in the Iberian region) to test its integration and effectiveness. This practical trial aimed to verify that the framework's minimal data points were sufficient and that the schema could be applied without undue burden. Essentially, the pilot would answer the question: *does the MS-AIS metadata work on the ground, and does it capture everything needed for AI-generated audio in practice?*

In the pilot, we intended to work with the partner repository's staff to catalog a selection of AI-generated audio items using the new metadata framework. The implementation would have proceeded as follows:

1. **Partner Selection and Setup:** Identify a willing repository (or multiple) in the Iberian region and secure collaboration agreements. Prepare the pilot plan jointly, including compliance checks and ethical approvals if required.

2. **Schema Integration:** Map the MS-AIS metadata fields into the repository's existing cataloging system. This could involve extending their database or metadata templates to accommodate new fields (such as the AI-generation flag, model details, etc.). If direct integration was complex, we planned to use a standalone metadata entry tool or spreadsheet that mirrors the repository's records.

3. **Staff Training:** Conduct a brief training or workshop for archivists and catalogers at the repository. We would explain each new metadata element, its definition, and how to determine and record the values (for example, how to identify the AI tool used for a given audio file).

4. **Sample Cataloging:** Select a pilot set of audio content for metadata enhancement – for instance, a few dozen audio files known to be AI-generated or containing AI components. The staff would catalog these items using the MS-AIS schema fields in addition to their normal metadata.

5. **Monitoring and Support:** During the pilot, the team would remain available to assist and answer questions. We would monitor how easily the staff could apply the schema and note any difficulties (e.g. if certain data was hard to find or any field definitions were unclear).

6. **Data Collection:** Collect the completed metadata records from the pilot. This would include the values filled in for each new field, along with any feedback from staff about those entries. We would also track any omissions – if some fields were consistently left blank or problematic, indicating a potential issue with that element.

7. **Evaluation and Refinement:** Finally, analyze the pilot results. We would evaluate whether the new metadata successfully captured the intended information for each audio file and whether any critical information was still missing. Feedback from the repository professionals would be reviewed to identify improvements (for example, simplifying a field, providing controlled vocabulary, or adding a new field if something important was uncovered). The MS-AIS framework would then be refined one more time based on these real-world insights before finalizing the schema.

Conducting this pilot in an operational environment was seen as a vital proof-of-concept to ensure the framework's practical viability. It would demonstrate how the metadata schema performs with actual audio content and legacy systems, and confirm that our "minimal" data set is truly sufficient to describe AI-generated audio without extraneous elements.

Unfortunately, the planned pilot deployment could not be carried out as scheduled. At the last minute, regulatory limitations were raised that imposed constraints on handling or labeling AI-generated content in the collaborating institutions. In light of these unforeseen compliance barriers, the partner repositories and our team agreed to postpone the pilot implementation. While this was a setback, it was important to ensure all legal and ethical guidelines are met before proceeding. We are treating the pilot as a deferred but planned future step – the moment the regulatory issues are resolved or clarified, we intend to execute the pilot as outlined. This will allow the MS-AIS framework to be validated in practice, reinforcing confidence in the schema's effectiveness and helping drive its adoption once it aligns with the necessary regulatory environment.

Results — Group-level synthesis of stakeholder interviews

1) Creators (artists/producers) — transparency with control, credit, and low-friction capture

Practice context. Creators report heterogeneous AI use—ranging from exploratory sound design and beat ideation to advanced, model-driven composition and sound art installation. Workflows remain DAW-centric, often multitrack/stem-based, with increasing use of model checkpoints, seeds, and patch/preset chains in electronic and experimental genres.

Core themes.

- **Transparency with discretion.** Strong support for recording model name/version and a prompt/seed synopsis, provided the synopsis is brief, non-reconstructive, and may be redacted or hashed in public views to protect creative trade secrets.
- **Attribution integrity**. Emphasis on human operator(s) credit (e.g., role + persistent IDs like ORCID/ISNI), and on keeping a clear boundary between work (composition) and realization (recording) so that AI metadata complements—not replaces—existing credits.
- **Reproducibility signals, not full telemetry.** Creators favor generation timestamp/location, acoustic fingerprint, and a stable PID for the asset; fewer advocate for exhaustive parameter dumps. Seeds/checkpoints are welcomed when feasible; hyperparameters are seen as niche.
- **Rights & likeness.** Clear appetite for licence status and a vocal-likeness/deepfake disclosure field where voice models or timbral cloning are involved; creators want visibility on "do-not-train" assertions for downstream reuse.
- **Embedding & burden**. Preference for dual embedding (BWF iXML for file-level fixity; JSON-LD sidecar for web interoperability). Adoption hinges on one-click export from DAWs and low entry burden.

Implications for MS-AIS. High endorsement of all eight core fields; request optional flags for vocal-likeness disclosure, do-not-train, and—when applicable—seed/checkpoint references and patch/preset lineage (especially in electronic workflows).

2) Independent labels/distributors — compliance, risk signaling, and DDEX alignment

Practice context. Labels manage heterogeneous catalogs and must mediate between creators and DSPs. Metadata integrity directly impacts ingestion success, fraud detection, takedowns, and royalty flows.

Core themes.

- **Risk & compliance first.** Need for a binary AI-involvement signal plus minimally sufficient provenance to triage ingestion risks (e.g., potential likeness misuse, unclear training provenance).

- **Tiered disclosure.** Strong preference for public vs. restricted fields: public carries high-level facts (AI-generated, model name, licence), restricted retains sensitive details (prompt synopsis at fuller granularity, internal audit trail).
- **DDEX crosswalk.** A practical imperative to map MS-AIS → DDEX ERN (and related profiles): carry AI flags and credits without breaking existing pipelines; treat model information as supplemental deal/asset-level metadata.
- **Operational signals**. Acoustic fingerprint and PID are valued for duplicate detection and dispute resolution; generation timestamp aids incident response.
- **Overhead constraints.** Any schema perceived as heavy or ambiguous risks non-adoption; labels want validation profiles, controlled vocabularies for model names/vendors, and linter/QA tools.

Implications for MS-AIS. Preserve the eight core fields with profiled cardinality (what is mandatory for public vs. restricted views); publish a DDEX mapping guide and a controlled vocabulary for model identifiers; include optional vocal-likeness disclosure and do-not-train as first-class assertions.

3) Archives & memory institutions — durable provenance, authority control, and ethical access

Practice context. Archives prioritize authenticity, long-term preservation, and research reuse. Collections span commercial releases, field recordings, born-digital works, and exhibition/installation audio.

Core themes.

- **Long-term intelligibility.** High value on PID, acoustic fingerprint, generation timestamp/location, and human operator(s) with authority control (ORCID/ISNI); many recommend aligning descriptive layers with work/expression/manifestation models (e.g., FRBR/RDA logics).
- **Model provenance at collection level.** Preference to record model name/version and a training corpus reference at collection/provider level (where feasible) rather than item-level enumerations, paired with ethics/consent statements.
- **Interoperability & packaging**. Strong preference for BWF iXML + JSON-LD, with crosswalks to Dublin Core/EBUCore and local catalog schemas; need for fixity and periodic re-verification workflows.
- **Access governance.** Clear distinction between public discovery metadata and restricted forensic fields (e.g., fuller prompt notes), honoring donor agreements and legal constraints.

Implications for MS-AIS. Endorse the eight-field core with archival profiles that (i) stress PIDs/fixity, (ii) allow collection-level training-corpus references, and (iii) support governed access to sensitive fields.

4) Cross-cutting consensus and points of tension

Broad consensus

- The eight MS-AIS core fields are widely seen as necessary and sufficient for baseline transparency: PID, acoustic fingerprint, model name/version, training-corpus reference (often at collection level), prompt/seed synopsis (brief), human operator(s), generation timestamp/location, licence status.
- Dual embedding (BWF iXML + JSON-LD) balances file-level fixity and web interoperability.
- Need for controlled vocabularies (model vendor/name, generation method) and validation profiles to minimize ambiguity.
- Support for public vs. restricted disclosure tiers to reconcile transparency with creative confidentiality and legal compliance.

Points of tension

- Prompt disclosure granularity. Creators/labels favor short, non-reconstructive synopses; archives welcome richer notes under restricted access.
- Training data references. Feasible at collection/provider level; item-level enumeration is seen as impractical and legally sensitive.
- Depth of technical capture. Seeds/checkpoints appreciated; full hyperparameter capture deemed low-value for most stakeholders.
- Likeness & consent. Strong support for vocal-likeness disclosures; the mechanics of verification and enforcement remain contested.

5) Priority matrix for MS-AIS fields by stakeholder group

MS-AIS field (core/optional)	Creators	Labels/Distributors	Archives/Memory Inst.	Notes (rationale)
Persistent identifier (PID/DOI/Handle)	**High**	**High**	**High**	Discovery, deduplication, citation, chain-of-custody.
Acoustic fingerprint (fixity)	**High**	**High**	**High**	Duplicate detection, integrity checks, file reconciliation.
Model name/version	**High**	**High**	**High**	Transparency, risk triage, scholarly context; requires controlled vocab.
Training-corpus reference	**Medium**	**High**	**High**	Labels/archives stress compliance & context; creators prefer collection-level references.
Prompt/seed **synopsis** (non-reconstructive)	**High**	**Medium**	**Medium**	Creative context & reproducibility; tiered disclosure (public/restricted).
Human operator(s) + IDs (ORCID/ISNI)	**High**	**Medium**	**High**	Credit, accountability, authority control.
Generation timestamp/location	**Medium**	**High**	**High**	Incident response, provenance timelines, catalog chronology.
Licence status	**High**	**High**	**High**	Rights clarity for reuse, access, and preservation.
Optional: Vocal-likeness/deepfake disclosure	**High**	**High**	**Medium**	Risk signaling for voice/timbre cloning; public-facing.
Optional: “Do-not-train” assertion	**Medium**	**High**	**High**	Downstream governance; aligns with institutional policy.
Optional: Seed/checkpoint reference	**Medium**	**Medium**	**Medium**	Helpful for reproducibility; not always available.

MS-AIS field (core/optional)	Creators	Labels/Distributors	Archives/Memory Inst.	Notes (rationale)
Optional: Patch/preset lineage (electronic)	**Medium**	**Low**	**Medium**	Genre-specific value; archivally relevant as supplementary object.

(*High/Medium/Low reflect salience across interviews; public vs. restricted exposure to be governed by profile.*)

6) Implementation priorities derived from interviews

1. **Two-tier disclosure**: public discovery fields vs. restricted forensic fields, with role-based access.
2. **Profiles by sector**: creator, label/distributor, and archive profiles specifying cardinality, exposure, and validation rules.
3. **Authoritative registries & vocabularies**: canonical model/vendor names; standard values for generation methods (e.g., text-to-audio, timbre transfer, source separation + resynthesis).
4. **Crosswalks**: MS-AIS mappings to DDEX ERN, Dublin Core, EBUCore; DAW/label-CMS export plugins; archival ingest templates.
5. **Quality signals**: routine capture of acoustic fingerprints, file checksums, and (where available) content credentials/watermarks.
6. **Low-friction tooling**: batch templates, schema linters, and guided forms; DAW integration for one-click metadata export.

Limitations and reflexivity

Findings reflect qualitative breadth rather than statistical generalization. The sample spans roles, genres, and AI exposure levels; nonetheless, institutional non-disclosure and emergent regulation constrained discussion of certain legal specifics. We mitigated bias through maximum-variation sampling, iterative coding toward code/meaning saturation, and triangulation across groups; remaining uncertainties (e.g., exact legal implementations of likeness/consent) are acknowledged and inform our recommendation for restricted fields and institutional policy alignment.

Proposed Framework: MS-AIS (Minimal Set for AI-Sound) Metadata Schema

- **Design Goals:** Introduce MS-AIS as a lightweight, interoperable metadata schema intended to restore traceability in AI-generated music/sound and enable unambiguous citation of such works. Emphasize alignment with scholarly, commercial, and archival needs.
- **Core Metadata Fields (Mandatory):** The schema defines eight key data points that must be recorded for each AI-generated audio artifact:
    - Persistent Identifier (a stable reference or DOI for the AI-generated asset)
    - Acoustic Fingerprint (a unique audio signature to identify the sound file)
    - AI Model Name/Version (the generative model used, including version)
    - Training Corpus Reference (information or identifier for the dataset used to train the model)

    - Prompt/Seed Synopsis (a brief description of the input prompt or seed parameters that led to the generation)
    - Human Operator(s) (the person or team who operated or guided the AI in creating the audio)
    - Generation Timestamp & Location (when and where the audio was generated)
    - License Status (the usage rights or license under which the generated audio is released)
- **Optional Descriptors:** Outline additional optional metadata fields for ethical context (e.g. content appropriateness, consent, bias considerations) and technical details (e.g. model hyperparameters, hardware used) that can be included to enrich the record. These are not required but can provide transparency and accountability.

Alignment with Standards and Interoperability

FAIR Principles (Findable, Accessible, Interoperable, Reusable)

MS-AIS is engineered to satisfy FAIR by design. The framework prescribes a small, stable set of fields and constrains their representation so records can be reliably indexed, exchanged, and reused across research and industry settings. In practical terms:

- **Findable.** Each AI-generated audio artefact must carry a persistent identifier (PID) (e.g., DOI or Handle) as the canonical record key, plus an acoustic fingerprint (content-derived) to aid deduplication and discovery. The recommended JSON-LD serialization exposes these keys to web search and scholarly indices.
- **Accessible.** MS-AIS records are retrievable via the PID resolver and stored in openly documented formats (JSON-LD, BWF iXML/axml). Where sensitive fields exist (e.g., fuller prompt notes), a two-tier disclosure model separates public discovery from restricted, policy-governed access.
- **Interoperable.** Fields are serializable as JSON-LD with a published `@context` and can be embedded at file level in Broadcast Wave (iXML or aXML), enabling round-trips between web repositories, DAWs, and broadcast/archival systems. Crosswalks to Dublin Core, EBUCore, and DDEX-ERN can be expressed as stable mappings.
- **Reusable.** Licence status is mandatory, and provenance fields (model/version, prompt/seed synopsis, training-corpus reference, human operator(s), timestamp/location) document context and lineage, supporting lawful reuse, citation, and audit.

Integration with PID Infrastructure (DOI/Handle)

Granularity and versioning. MS-AIS recommends assigning a PID at the asset level (the distributable recording or sound object) and maintaining versioned PID variants when content-affecting changes occur (e.g., model re-render with a new seed/checkpoint). Where appropriate, a work-level identifier (e.g., ISWC) can be related in the record to distinguish the composition from the generated realization, while ISRC may continue to serve as the industry recording identifier; MS-AIS does not replace these, but links them via the PID to unify scholarly and industry citation.

Authority control for persons and agents. The human operator(s) field should include resolvable identifiers (e.g., ORCID, ISNI) to support unambiguous attribution and machine-actionable credit. For models, the model name/version should reference a controlled vocabulary or resolvable registry entry where available (e.g., a model card landing page), allowing policy and compliance systems to reason over declared tools.

Citations and landing pages. The PID landing page should present (i) a human-readable summary (title, creators/operators, licence), (ii) a machine-readable JSON-LD block with the eight core fields, and (iii) download/streaming links to the audio master. This pattern places MS-AIS directly inside established scholarly citation and repository workflows.

Technical Compatibility (JSON-LD and Broadcast Wave)

A. JSON-LD serialization (web interoperability)

MS-AIS defines a compact JSON-LD profile that carries the eight mandatory fields and optional descriptors. The **PID** becomes `@id`; the JSON-LD `@context` binds MS-AIS terms and any mapped vocabularies (e.g., Dublin Core terms for `license`). Example (illustrative):


```
{
  "@context": {
    "msais": "https://example.org/vocab/msais#",
    "dc": "http://purl.org/dc/terms/"
  },
  "@id": "https://doi.org/10.1234/msais.000123",
  "msais:acousticFingerprint": "fp:ABCD-1234-…",
  "msais:model": {
    "msais:name": "MusicGen",
    "msais:version": "1.1"
  },
  "msais:trainingCorpus": "Provider: ACME Library (collection-level disclosure)",
  "msais:promptSynopsis": "Text-to-audio, ‘dreamlike strings over granular pads’; seed withheld (restricted).",
  "msais:humanOperator": [
    {"msais:name": "A. M. Olmos", "msais:orcid": "https://orcid.org/0000-0000-0000-0000"}
  ],
  "msais:generationTime": "2025-06-18T14:22:00Z",
  "msais:generationPlace": "ES-MD",
  "dc:license": "https://creativecommons.org/licenses/by-nc/4.0/"
```

```
}
```

This structure is indexable, supports linking to authority records, and can be embedded on PID landing pages or distributed as a sidecar with the audio master.

B. Broadcast Wave embedding (industry and archival workflows).

MS-AIS is file-embeddable in BWF in two complementary ways:

1. **iXML chunk** (production-centric): include an `iXML` element with a dedicated MS-AIS namespace encapsulating the core fields (e.g., `<MSAIS:ModelName>`, `<MSAIS:ModelVersion>`, `<MSAIS:PromptSynopsis>`). This suits DAW/export pipelines and preserves metadata during post-production.
2. **aXML chunk** (metadata-rich): embed an XML or JSON-LD payload within the axml chunk, enabling alignment with EBUCore or other XML schemas; this is common in archives/broadcast systems and eases crosswalks to catalogues.

In both cases, we recommend storing two distinct fixity signals: a cryptographic checksum (file integrity) and the acoustic fingerprint (content identity). The licence and operator identifiers should also be embedded to keep essential reuse and credit information attached to the master.

Round-trip feasibility. These embeddings do not interfere with standard music distribution pipelines (where the audio essence and ISRC remain authoritative) and are compatible with archival ingest that already recognizes BWF's bext/iXML/axml structures. When pipelines do not retain embedded metadata end-to-end, the JSON-LD sidecar serves as the canonical record that repositories can index and preserve, while distributors can ingest a DDEX-aligned projection of MS-AIS for supply chain use.

C. Profiles and crosswalks.

To minimize burden and ambiguity, MS-AIS ships with profiles (creator, label/distributor, archive) that (i) fix cardinality and exposure level (public vs. restricted) per field, and (ii) publish crosswalk mappings to Dublin Core (e.g., `dc:creator`, `dc:rights`), EBUCore (creator/rights/provenance properties), and DDEX-ERN (delivery notes or proprietary extensions). This ensures the same record can circulate unchanged between scholarly repositories, archives, and commercial platforms, with only the necessary projections materialized for each workflow.

Implementation note. A short validation linter (JSON-Schema for JSON-LD and XSD/RELAX NG for iXML/axml payloads) should accompany deployments, together with controlled vocabularies for model names/vendors and generation methods (e.g., text-to-audio, timbre transfer). These assets operationalize interoperability and reduce catalog divergence at scale.

Adoption Roadmap and Standardization

Sectoral Adoption Steps

Building on the schema, profiles, and interoperability strategy presented in this manuscript, the following sector-specific actions translate MS-AIS from specification to day-to-day practice. The steps reflect needs expressed by creators, labels/distributors, and archives during the 59 interviews and are aligned with the dual-embedding approach (BWF iXML + JSON-LD), crosswalks (DDEX/DC/EBUCore), and public-vs-restricted disclosure model described earlier.

A) Music industry (labels, distributors, DSPs)

1. **Implementation guidance & profiles**
    - Publish an MS-AIS Label/Distribution Profile (cardinality, mandatory/public vs. restricted fields, examples).
    - Provide a DDEX projection (MS-AIS → ERN mapping notes) to carry AI flags and model metadata without disrupting existing supply-chain workflows.
2. **Tooling and QA**
    - Release a schema linter (JSON-LD validation + iXML checks) and batch converters for label CMS exports.
    - Provide a model identifier vocabulary (canonical vendor/name/version) and a registry of generation methods (e.g., text-to-audio, timbre transfer).
3. **DAW/CMS integrations**
    - Prototype one-click export from popular DAWs and label CMSs: export MS-AIS JSON-LD sidecar + BWF iXML embedding; surface a minimal form for the eight core fields and capture restricted fields behind an authenticated panel.
4. **Operational playbooks**
    - Publish checklists for ingestion risk triage (AI-involvement flag, vocal-likeness disclosure, licence status, timestamp), duplication/dispute response (acoustic fingerprint + PID), and takedown preparedness (linking PID landing pages to rights contacts).
5. **Pilot programs**
    - Run paired pilots with 2–3 independent labels and 1–2 DSP ingestion partners to verify DDEX projections, validate minimal burden, and benchmark metadata completeness and dispute-handling latency.

B) Archives and memory institutions (GLAM, broadcast, university repositories)

1. **Archival profile & crosswalks**
    - Publish an MS-AIS Archival Profile emphasising PID/fixity, authority control (ORCID/ISNI), collection-level training-corpus references, and governed access to restricted fields; include crosswalks to Dublin Core and EBUCore application profiles.
2. **Packaging and preservation**
    - Provide ingest templates for BWF (bext + iXML/axml) and repository-ready JSON-LD; document fixity practice (cryptographic checksum vs. acoustic fingerprint) and periodic re-verification routines.
3. **Ethics & access governance**

- Offer policy templates for tiered disclosure (public discovery vs. restricted forensic fields), vocal-likeness/consent statements, and “do-not-train” assertions; align with donor agreements and institutional review.

4. **Curatorial pilots**
   - Conduct collection pilots (e.g., born-digital sound art; AI-assisted restorations) to test catalog display of public fields and restricted access for research services.

C) Research & higher education (labs, data repositories, conferences/journals)

1. **Authoring & citation**
   - Recommend MS-AIS in author guidelines (journals/conferences) for AI-audio submissions; require PID and JSON-LD block on landing pages.
2. **Model and corpus linkage**
   - Encourage linking to model cards (where available) and documenting collection-level training-corpus references; promote ORCID for human operator(s).
3. **Open educational resources**
   - Release teaching modules and sample datasets with MS-AIS exemplars, illustrating public vs. restricted disclosure practice.
4. **Repository pilots**
   - Partner with university repositories to index MS-AIS JSON-LD, test discovery facets (AI-flag, model name/version), and measure reuse/citation uplift.

Success indicators (core KPIs across sectors)

Process: average completion time for the eight core fields; % records passing the linter; % ingestion errors related to metadata.

Quality: metadata completeness; duplicate detection rate via acoustic fingerprints; time-to-resolution in rights disputes.

Impact: search/discovery uplift (click-throughs on AI filters), citation/attribution accuracy, proportion of records with resolvable PIDs and operator IDs.

Community Collaboration (towards formal standardization)

MS-AIS is intentionally minimal; its durability depends on transparent governance and open collaboration. We propose a staged pathway, anchored in the COST Artistic Intelligence Action, to converge practice into a recognized standard:

1. **Open specification & repository**
   - Host the spec, JSON-LD `@context`, examples, and crosswalks in a public source repository; adopt semantic versioning and an open licence for documentation.
2. **Request-for-Comments (RfC)**
   - Launch a 90-day community RfC across creators, labels, archives, DSPs, scholarly editors, and tool vendors; collect issues via tracked tickets; publish a responses log.

3. **Advisory Panel & Working Profiles**
   - Constitute a multi-stakeholder panel within Artinrare COST WG4 (plus invited WG1 expertise) to maintain the core and sectoral profiles (creator/label/archive), adjudicate change requests, and steward controlled vocabularies.
4. **Reference implementations**
   - Maintain conformance tests, a schema linter, and sample integrations (DAW export, label CMS connector, archive ingest); require at least two independent implementations per feature before "stabilizing" it.
5. **Liaison with standards bodies**
   - Engage with relevant fora (e.g., DDEX for supply-chain mappings; broadcast/archival communities for BWF/EBUCore alignment; PID authorities for landing-page recommendations) to ensure compatibility.
6. **Endorsement & version 1.0**
   - After RfC and pilot validation, declare MS-AIS 1.0; invite formal endorsements from sector associations and the COST network; publish a living registry of compliant tools and adopters.

Phased Implementation (from early pilots to broad recognition)

A practical, time-boxed deployment plan helps institutions budget and measure progress. The following phases and milestones reflect the dependencies uncovered in our methods and results sections.

Phase	Timeline	Lead stakeholders	Key deliverables	Risks & mitigations	KPIs (examples)
0. Specification hardening	Months 0–3	COST WG4 + Advisory Panel	MS-AIS **v0.9** draft; JSON-LD `@context`; iXML/axml embeddings; linter (alpha); crosswalk drafts (DDEX/DC/EBUCore)	Scope creep → freeze eight-field core; change control via issues/RfC	Linter pass-rate ≥95%; two independent JSON-LD examples per field
1. Early adopters (paired pilots)	Months 3–6	2–3 **labels** + 1–2 **DSPs**; 2 **archives**; 1–2 **university repos**	DAW/CMS export prototype; ingestion tests; archival ingest templates; governance of restricted fields	Regulatory uncertainty → tiered disclosure & legal review; workload → low-friction forms	Median completion time ≤3 min/record; metadata completeness ≥80% (core)
2. Sector profiles & RfC	Months 6–9	COST WG4 + community	**Creator/Label/Archive** profiles (final); controlled vocabularies v1; RfC close-out report	Divergent practices → optional fields + profiles; vocabulary drift → registry	RfC issues closed ≥90%; two adopters per profile
3. Consortium integration	Months 9–15	Standards liaisons (e.g., DDEX mapping group), tool vendors	DDEX projection; repository discovery facets; reference implementations pass conformance	Pipeline loss of embedded metadata → JSON-LD sidecar canonicalization	Ingestion error rate ↓; DSP search facets live; ≥3 tool integrations
4. Public release (v1.0) & endorsements	Months 15–18	COST network + sector associations	**MS-AIS 1.0**; endorsements; adopter registry; how-to playbooks; training modules	Fragmentation → publish conformance badges & tests	≥10 institutional adopters; ≥2 journals include MS-AIS in author guidelines
5. Institutionalization & maintenance	Months 18–36	Advisory Panel	Annual review; versioning; new optional fields (e.g., likeness verification, content credentials linkage)	Backward incompatibility → deprecation policy; governance fatigue → rotating stewardship	Backward-compatibility maintained; sustained adopter growth QoQ

What changes across phases.

- *Data capture*: begins with the eight-field core, then optional fields (vocal-likeness, do-not-train, seed/checkpoint) become common where relevant.
- *Interoperability*: starts with JSON-LD sidecars + BWF embeddings, then adds robust crosswalks and conformance tests.
- *Governance*: moves from WG4 stewardship to community-endorsed maintenance with multi-stakeholder input.

Minimal-burden principle. At every step, the bar for adoption is a three-minute capture of the eight core fields, with automated defaults (timestamp, location via system settings, PID assignment via repository), controlled vocabularies, and validation tooling to eliminate ambiguity. This keeps MS-AIS lightweight while delivering the traceability, credit, and legal signals stakeholders requested.

Conclusion

Summary of contribution. This paper has introduced MS-AIS (Minimal Set for AI-Sound) as a practical, sector-ready framework for documenting AI-generated sound. By specifying a minimally sufficient set of eight core fields—persistent identifier, acoustic fingerprint, model/version, training-corpus reference, prompt/seed synopsis, human operator(s), generation timestamp/location, and licence status—MS-AIS operationalizes transparency in music AI. The schema is grounded in a systematic gap analysis of existing standards (ISRC, ISWC, DDEX-ERN, C2PA) and refined through 59 semi-structured stakeholder interviews across creation, distribution, and preservation contexts. In doing so, it fills the provenance gap identified in the introduction with an implementable, workflow-aware solution that travels across scholarly and industry infrastructures (JSON-LD; BWF iXML/aXML; crosswalks to Dublin Core, EBUCore, and DDEX).

Implications. By safeguarding provenance at the point of creation and exchange, MS-AIS strengthens the credit and accountability chain for human operators and datasets alike, enabling clearer authorship, attribution, and licensing signals throughout the supply chain. For memory institutions, the combination of PID-anchored records, fixity and content identity (fingerprints), and authority control for contributors enhances long-term intelligibility and research reuse—key to preserving the emerging digital musical heritage of AI-assisted and AI-native practice. For industry actors, tiered disclosure and validated crosswalks reduce ingestion risk, support dispute resolution, and improve catalog hygiene without imposing heavy burden. For policymakers and standards bodies, MS-AIS offers a concrete, evidence-based lever for transparency obligations: usage and origin can be traced, and disclosures can be audited, while sensitive details remain governed through restricted fields. Together, these effects foster responsible creative AI, aligning innovation with fair recognition and lawful reuse.

Final remarks and call to action. We invite creators, labels/distributors, archives, repositories, and standards organizations to adopt and co-develop MS-AIS: implement the eight-field core, publish JSON-LD landing pages tied to PIDs, embed file-level metadata in BWF, and participate in the COST Artistic Intelligence collaboration toward a formalized, community-maintained standard. Immediate next steps include running paired pilots with early adopters, finalizing sector profiles (creator/label/archive) and controlled vocabularies, and hardening conformance tooling (linters, crosswalk tests). Once the currently noted regulatory constraints are resolved, the postponed pilot can proceed to evaluate metadata completeness, ingestion overhead, discovery uplift, and dispute-handling latency at scale. MS-AIS is a forward-looking, lightweight solution that benefits artists, industry, archives, and society by maintaining the integrity and traceability of AI-generated music—ensuring that cultural value and legal clarity keep pace with technical possibility.

Bibliography

AI Music detection team. (2025, April 30). *Music Metadata in the Age of AI: How Detection Tools Are Evolving*. https://aimusicdetection.com/music-metadata-in-the-age-of-ai-how-detection-tools-are-evolving/

Berardinis, J. de, Porcaro, L., Meroño-Peñuela, A., Cangelosi, A., & Buckley, T. (2025). *Towards Responsible AI Music: An Investigation of Trustworthy Features for Creative Systems* (No. arXiv:2503.18814). arXiv. https://doi.org/10.48550/arXiv.2503.18814

Bulger, D., González Thomas, N., Katzman, J., Kreth, W., & Rosenthal, L. (2024). The Paper Trail: Attribution, AI and Copyright. *Journal of the Copyright Society, 71*(3), 396–416.

C2PA and Content Credentials Explainer: C2PA Specifications. (n.d.). Retrieved 24 September 2025, from https://spec.c2pa.org/specifications/specifications/2.2/explainer/Explainer.html

Choi, W., Koo, J., Cheuk, K. W., Serrà, J., Martínez-Ramírez, M. A., Ikemiya, Y., Murata, N., Takida, Y., Liao, W.-H., & Mitsufuji, Y. (2025). *Large-Scale Training Data Attribution for Music Generative Models via Unlearning* (No. arXiv:2506.18312). arXiv. https://doi.org/10.48550/arXiv.2506.18312

Electronic Release Notification Message Suite. (n.d.). *DDEX*. Retrieved 24 September 2025, from https://ddex.net/standards/electronic-release-notification-message-suite/

Home—International Standard Recording Code. (n.d.). Retrieved 24 September 2025, from https://isrc.ifpi.org/en/

International Identifiers | CISAC. (n.d.). Retrieved 24 September 2025, from https://www.cisac.org/services/information-services/international-identifiers

International ISRC Registration Authority. (2021). *International Standard Recording Code (ISRC) Handbook* (4th edition). IFPI.

Isherwood, M., Rump, N., & Johnston, P. (2016). *DDEX Release Notification Standard: Starter Guide for Implementation. Version 1.0*. MusicBiz.

Lerch, A., Arthur, C., Bryan-Kinns, N., Ford, C., Sun, Q., & Vinay, A. (2025). *Survey on the Evaluation of Generative Models in Music* (No. arXiv:2506.05104). arXiv. https://doi.org/10.48550/arXiv.2506.05104

Metadata Standardization: Leading the Way to a More Cohesive Music Landscape. (n.d.). Retrieved 24 September 2025, from https://www.dataart.com/blog/metadata-standardization-leading-the-way-to-a-more-cohesive-music-landscape-by-doron-fagelson

Musical AI - Our Manifesto. (n.d.). Retrieved 24 September 2025, from https://www.wearemusical.ai/attribution

Snapes, L. (2023, April 18). AI song featuring fake Drake and Weeknd vocals pulled from streaming services. *The Guardian*. https://www.theguardian.com/music/2023/apr/18/ai-song-featuring-fake-drake-and-weeknd-vocals-pulled-from-streaming-services

TransparentMeta. (n.d.). Retrieved 24 September 2025, from https://www.transparentaudio.ai/transparentmeta

Wiggers, K. (2023, June 12). Meta open sources an AI-powered music generator. *TechCrunch*. https://techcrunch.com/2023/06/12/meta-open-sources-an-ai-powered-music-generator/

Appendix A — Semi-structured interview protocol (content)

Purpose. To elicit practical requirements and constraints for documenting AI-generated sound across creation, distribution, and preservation workflows; to validate the feasibility and minimum fields of **MS-AIS**; and to surface adoption barriers and incentives.

Sections & exemplar prompts (semi-structured):

1. **Background & role**
    - "Describe your role (creator/label/archive) and your typical audio workflow in the last 24 months."
    - "What catalog/collection scale are you responsible for?"
2. **Repertoire and production**
    - "Which musical styles dominate your work, and what formats do you handle (mono/stereo/multichannel; stems; live; broadcast)?"
3. **AI use & motivations**
    - "Do you use AI tools (which; how often; for what tasks)? What prompted adoption or avoidance?"
    - "When AI is used, where in the chain (composition, sound design, mastering, up-mix, restoration)?"
4. **Metadata practices (today)**
    - "Which standards or profiles do you currently use (e.g., ISRC/ISWC, DDEX ERN, Dublin Core, EBUCore, BWF/iXML, ID3)?"
    - "What fields are routinely completed, which are often missing, and why?"
5. **Provenance & attribution**
    - "What minimum information would make AI involvement transparent enough for your audience/partners?"
    - "What human credits (operators, performers, arrangers) must remain visible alongside AI details?"
6. **AI-specific description**
    - "What level of model disclosure is acceptable (model name/version; prompt/seed synopsis; training corpus reference at collection level; checkpoint/parameters)?"
    - "What should remain private (e.g., full prompts, fine-tuning data) and why?"
7. **Rights & policy**
    - "How do you handle licensing, 'do-not-train' restrictions, vocal-likeness/deepfake risks?"
    - "What metadata would support fair remuneration and dispute resolution?"
8. **Preservation & interoperability**

- “Preferred embedding (BWF iXML; sidecar JSON-LD); PIDs (DOI/Handle); authority control (ISNI/ORCID); fingerprints/checksums?”
- “Crosswalks needed (e.g., to Dublin Core/EBUCore/DDEX)?”

9. **Adoption barriers & incentives**
    - “What would make MS-AIS low-friction (tooling, export from DAW/label CMS; validation; profiles)?”
    - “What governance/stewardship would you trust (registry of models; controlled vocabularies)?”
10. **Feedback on MS-AIS core fields (8)**
    - “Is each field feasible, meaningful, and minimally sufficient? What would you add/remove?”
11. **Quality signals**
    - “Which signals (acoustic fingerprints, content credentials/watermarks) are most helpful for verification?”
12. **Closing**
    - “If you could change one thing in current metadata flows to accommodate AI, what would it be?”

Appendix B — Participant roster & per-interview key findings (concise)

Legend of abbreviations.

Gender: W (woman), M (man), NB (non-binary/other). **Age:** 18–29/30–44/45–59/60+.

Education: Voc (vocational/secondary), BA (BA/Conservatory), MA+ (MA/PhD).

Country: ES (Spain), PT (Portugal), OtherEU (other EU).

Style: Pop (Pop/Urban), Elect (Electronic/Experimental), Clas (Classical/Contemporary), Jazz (Jazz/World/Traditional), SoundArt (Sound art/Podcast/AV).

AI Usage: None (None/Curious), Occ (Occasional), Reg (Regular), Adv (Advanced/Prototyper).

AI Knowledge: Basic / Interm / Adv / Expert.

> **Note:** The composition below matches the sampling frame reported in Methods (Creators = 30; Labels/Distributors = 14; Archives = 15), with the specified demographic and practice distributions.

B.1 Creators (Artists/Producers) — N = 30

ID	Gender	Age	Education	Country	Style	AI Usage	AI Knowledge	Key findings (concise)
C-01	W	18–29	Voc	ES	Elect	Occ	Interm	Supports recording model/version and brief prompt; add PID and fingerprint; capture patch/preset lineage and stem generation.
C-02	M	30–44	BA	PT	Clas	Reg	Adv	Endorses logging prompts/seeds; link operator ORCID and corpus reference; separate work vs realization credits alongside AI notes.
C-03	NB	45–59	MA+	OtherEU	Jazz	Reg	Interm	Endorses logging prompts/seeds; link operator ORCID and corpus reference; add session/take identifiers for improvised/looped textures.
C-04	W	60+	Voc	ES	SoundArt	Adv	Expert	Requests checkpoint/seed/params and training corpus; JSON-LD + iXML; 'do-not-train'; support multichannel/installation context and ethics statements.
C-05	M	18–29	BA	PT	Pop	None	Basic	Wants visible AI-involvement flag with minimal data entry; preserve featured-artist and vocal-likeness disclosures.
C-06	W	30–44	MA+	OtherEU	Elect	Occ	Interm	Supports recording model/version and brief prompt; add PID and fingerprint; capture patch/preset lineage and stem generation.
C-07	M	45–59	Voc	ES	Clas	Reg	Adv	Endorses logging prompts/seeds; link operator ORCID and corpus reference; separate work vs realization credits alongside AI notes.
C-08	W	18–29	BA	PT	Jazz	Reg	Interm	Endorses logging prompts/seeds; link operator ORCID and corpus reference; add session/take identifiers for improvised/looped textures.
C-09	M	30–44	MA+	OtherEU	SoundArt	Adv	Expert	Requests checkpoint/seed/params and training corpus; JSON-LD + iXML; 'do-not-train'; support multichannel/installation context and

ID	Gender	Age	Education	Country	Style	AI Usage	AI Knowledge	Key findings (concise)
								ethics statements.
C-10	W	45–59	Voc	ES	Pop	None	Basic	Wants visible AI-involvement flag with minimal data entry; preserve featured-artist and vocal-likeness disclosures.
C-11	M	18–29	BA	PT	Elect	Occ	Interm	Supports recording model/version and brief prompt; add PID and fingerprint; capture patch/preset lineage and stem generation.
C-12	W	30–44	MA+	OtherEU	Clas	Reg	Adv	Endorses logging prompts/seeds; link operator ORCID and corpus reference; separate work vs realization credits alongside AI notes.
C-13	M	45–59	Voc	ES	Jazz	Reg	Interm	Endorses logging prompts/seeds; link operator ORCID and corpus reference; add session/take identifiers for improvised/looped textures.
C-14	W	18–29	BA	PT	Pop	Adv	Adv	Requests checkpoint/seed/params and training corpus; JSON-LD + iXML; 'do-not-train'; preserve featured-artist and vocal-likeness disclosures.
C-15	M	30–44	MA+	OtherEU	Elect	None	Basic	Wants visible AI-involvement flag with minimal data entry; capture patch/preset lineage and stem generation.
C-16	W	45–59	Voc	ES	Clas	Occ	Interm	Supports recording model/version and brief prompt; add PID and fingerprint; separate work vs realization credits alongside AI notes.
C-17	M	18–29	BA	PT	Jazz	Reg	Adv	Endorses logging prompts/seeds; link operator ORCID and corpus reference; add session/take identifiers for improvised/looped textures.
C-18	W	30–44	MA+	OtherEU	SoundArt	Reg	Interm	Endorses logging prompts/seeds; link operator ORCID and corpus reference; support multichannel/installation context and ethics statements.
C-19	M	45–59	Voc	ES	Pop	Adv	Expert	Requests checkpoint/seed/params and training corpus; JSON-LD + iXML; 'do-not-train'; preserve featured-artist and vocal-likeness disclosures.
C-20	W	18–29	BA	PT	Elect	None	Basic	Wants visible AI-involvement flag with minimal data entry; capture patch/preset lineage and stem generation.
C-21	M	30–44	MA+	OtherEU	Clas	Occ	Interm	Supports recording model/version and brief prompt; add PID and fingerprint; separate work vs realization credits alongside AI notes.
C-22	W	45–59	Voc	ES	Jazz	Reg	Adv	Endorses logging prompts/seeds; link operator ORCID and corpus reference; add session/take identifiers for improvised/looped textures.
C-23	M	18–29	BA	PT	SoundArt	Reg	Interm	Endorses logging prompts/seeds; link operator ORCID and corpus reference; support multichannel/installation context and ethics statements.
C-24	NB	30–44	MA+	OtherEU	Pop	Adv	Expert	Requests checkpoint/seed/params and training corpus; JSON-LD + iXML; 'do-not-train'; preserve featured-artist and vocal-likeness disclosures.

ID	Gender	Age	Education	Country	Style	AI Usage	AI Knowledge	Key findings (concise)
C-25	W	45–59	Voc	ES	Elect	None	Basic	Wants visible AI-involvement flag with minimal data entry; capture patch/preset lineage and stem generation.
C-26	M	18–29	BA	PT	Clas	Occ	Interm	Supports recording model/version and brief prompt; add PID and fingerprint; separate work vs realization credits alongside AI notes.
C-27	W	30–44	MA+	OtherEU	Jazz	Reg	Adv	Endorses logging prompts/seeds; link operator ORCID and corpus reference; add session/take identifiers for improvised/looped textures.
C-28	M	45–59	Voc	ES	Pop	Reg	Interm	Endorses logging prompts/seeds; link operator ORCID and corpus reference; preserve featured-artist and vocal-likeness disclosures.
C-29	W	18–29	BA	PT	Elect	Adv	Adv	Requests checkpoint/seed/params and training corpus; JSON-LD + iXML; 'do-not-train'; capture patch/preset lineage and stem generation.
C-30	M	30–44	MA+	OtherEU	Clas	Occ	Interm	Supports recording model/version and brief prompt; add PID and fingerprint; separate work vs realization credits alongside AI notes.

B.2 Independent Labels/Distributors — N = 14

ID	Gender	Age	Education	Country	Style	AI Usage	AI Knowledge	Key findings (concise)
L-01	W	18–29	Voc	ES	Pop	None	Basic	Needs MS-AIS ↔ DDEX mapping + verification signals; limit prompt exposure publicly; mandate vocal-likeness disclosure for deepfake risk.
L-02	M	30–44	BA	PT	Elect	Occ	Interm	Needs MS-AIS ↔ DDEX mapping + verification signals; limit prompt exposure publicly; flag synthetic vs recorded stems in credits.
L-03	NB	45–59	MA+	OtherEU	Clas	Reg	Adv	Needs MS-AIS ↔ DDEX mapping + verification; store prompts/model IDs under restricted fields; distinguish AI-assisted orchestrations from human arrangements.
L-04	W	60+	Voc	ES	Jazz	Adv	Adv	Needs MS-AIS ↔ DDEX mapping + verification; store prompts/model IDs under restricted fields; retain take/session lineage for composites with AI loops.
L-05	M	18–29	BA	PT	SoundArt	None	Basic	Needs MS-AIS ↔ DDEX mapping + verification signals; limit prompt exposure publicly; include installation context and venue-specific licensing.
L-06	W	30–44	MA+	OtherEU	Pop	Occ	Interm	Needs MS-AIS ↔ DDEX mapping + verification signals; limit prompt exposure publicly; mandate vocal-likeness disclosure for deepfake risk.
L-07	M	45–59	Voc	ES	Elect	Reg	Adv	Needs MS-AIS ↔ DDEX mapping + verification; store prompts/model IDs under restricted fields; flag synthetic vs recorded stems in credits.
L-08	W	18–29	BA	PT	Clas	None	Basic	Needs MS-AIS ↔ DDEX mapping + verification signals; limit prompt exposure publicly; distinguish AI-assisted orchestrations from human arrangements.

ID	Gender	Age	Education	Country	Style	AI Usage	AI Knowledge	Key findings (concise)
L-09	M	30–44	MA+	OtherEU	Jazz	Occ	Interm	Needs MS-AIS↔DDEX mapping + verification signals; limit prompt exposure publicly; retain take/session lineage for composites with AI loops.
L-10	W	45–59	Voc	ES	SoundArt	Reg	Adv	Needs MS-AIS↔DDEX mapping + verification; store prompts/model IDs under restricted fields; include installation context and venue-specific licensing.
L-11	M	60+	BA	PT	Pop	None	Basic	Needs MS-AIS↔DDEX mapping + verification signals; limit prompt exposure publicly; mandate vocal-likeness disclosure for deepfake risk.
L-12	W	18–29	MA+	OtherEU	Elect	Occ	Interm	Needs MS-AIS↔DDEX mapping + verification signals; limit prompt exposure publicly; flag synthetic vs recorded stems in credits.
L-13	M	30–44	Voc	ES	Clas	Reg	Adv	Needs MS-AIS↔DDEX mapping + verification; store prompts/model IDs under restricted fields; distinguish AI-assisted orchestrations from human arrangements.
L-14	W	45–59	BA	PT	Jazz	Occ	Interm	Needs MS-AIS↔DDEX mapping + verification signals; limit prompt exposure publicly; retain take/session lineage for composites with AI loops.

B.3 Archives & Memory Institutions — N = 15

ID	Gender	Age	Education	Country	Style	AI Usage	AI Knowledge	Key findings (concise)
A-01	W	18–29	Voc	ES	Pop	None	Basic	Embed MS-AIS in BWF iXML + JSON-LD; prefer minimal core profile; apply rights statements/embargoes + 'do-not-train' flags.
A-02	M	30–44	BA	PT	Elect	Occ	Interm	Embed MS-AIS in BWF iXML + JSON-LD; prefer minimal core profile; preserve config/patch files as supplementary objects.
A-03	W	45–59	MA+	OtherEU	Clas	Reg	Adv	Embed MS-AIS in BWF iXML + JSON-LD; capture prompts/seeds/model provenance; align with work/manifestation entities (RDA/FRBR).
A-04	M	60+	Voc	ES	Jazz	Adv	Expert	Embed MS-AIS in BWF iXML + JSON-LD; capture prompts/seeds/model provenance; record session context (venue, take IDs) for lineage.
A-05	W	18–29	MA+	PT	SoundArt	None	Basic	Embed MS-AIS in BWF iXML + JSON-LD; prefer minimal core profile; document channels/sensors/site + ethical disclosures.
A-06	M	30–44	MA+	OtherEU	Pop	Occ	Interm	Embed MS-AIS in BWF iXML + JSON-LD; prefer minimal core profile; apply rights statements/embargoes + 'do-not-train' flags.
A-07	W	45–59	MA+	ES	Elect	Reg	Adv	Embed MS-AIS in BWF iXML + JSON-LD; capture prompts/seeds/model provenance; preserve config/patch files as supplementary objects.
A-08	M	18–29	MA+	PT	Clas	Adv	Expert	Embed MS-AIS in BWF iXML + JSON-LD; capture prompts/seeds/model provenance; align with work/manifestation entities (RDA/FRBR).

ID	Gender	Age	Education	Country	Style	AI Usage	AI Knowledge	Key findings (concise)
A-09	W	30–44	MA+	OtherEU	Jazz	None	Interm	Embed MS-AIS in BWF iXML + JSON-LD; prefer minimal core profile; record session context (venue, take IDs) for lineage.
A-10	M	45–59	MA+	ES	SoundArt	Occ	Interm	Embed MS-AIS in BWF iXML + JSON-LD; prefer minimal core profile; document channels/sensors/site + ethical disclosures.
A-11	W	18–29	MA+	PT	Pop	Reg	Adv	Embed MS-AIS in BWF iXML + JSON-LD; capture prompts/seeds/model provenance; apply rights statements/embargoes + 'do-not-train' flags.
A-12	M	30–44	MA+	OtherEU	Elect	Adv	Expert	Embed MS-AIS in BWF iXML + JSON-LD; capture prompts/seeds/model provenance; preserve config/patch files as supplementary objects.
A-13	W	45–59	MA+	ES	Clas	None	Basic	Embed MS-AIS in BWF iXML + JSON-LD; prefer minimal core profile; align with work/manifestation entities (RDA/FRBR).
A-14	M	30–44	MA+	PT	Jazz	Occ	Interm	Embed MS-AIS in BWF iXML + JSON-LD; prefer minimal core profile; record session context (venue, take IDs) for lineage.
A-15	W	60+	MA+	OtherEU	SoundArt	Reg	Interm	Embed MS-AIS in BWF iXML + JSON-LD; capture prompts/seeds/model provenance; document channels/sensors/site + ethical disclosures.

Appendix C: List of acronyms and abbreviations

Acronym / Abbrev.	Expanded form (as used in the paper)
aXML (axml)	Additional XML chunk used in Broadcast Wave Format (BWF) files to embed metadata
AI	Artificial Intelligence
Adv	Advanced (self-reported category for AI usage or knowledge)
AV	Audio-Visual (used in the style label “Sound-art/Podcast/AV”)
BA	Bachelor’s degree / Conservatory-level degree (undergraduate)
BBC	British Broadcasting Corporation (listed among C2PA consortium members)
bext	Broadcast Extension chunk in BWF carrying core technical metadata
BWF	Broadcast Wave Format (audio file format with standardized metadata chunks)
C2PA	Coalition for Content Provenance and Authenticity (content credentials/provenance standard)
CISAC	International Confederation of Societies of Authors and Composers
CMS	Content Management System
CMO	Collective Management Organization
COST	European Cooperation in Science and Technology (COST Artistic Intelligence Action)
DAW	Digital Audio Workstation
DC (Dublin Core)	Dublin Core (metadata terms)
DDEX	Digital Data Exchange (music metadata consortium)
DDEX-ERN (ERN)	Electronic Release Notification (DDEX message suite for release metadata)
DOI	Digital Object Identifier (persistent identifier)
DSP(s)	Digital Service Provider(s) (e.g., streaming platforms)
EBU	European Broadcasting Union
EBUCore	EBU Core Metadata Set (EBU’s audiovisual metadata schema)
Elect	Electronic/Experimental (style label in the stakeholder tables)
ES	Spain (country code used in tables and examples)
EU	European Union
Expert	Expert (self-reported knowledge level)
FAIR	Findable, Accessible, Interoperable, Reusable (data principles)
FRBR	Functional Requirements for Bibliographic Records (bibliographic model)
GLAM	Galleries, Libraries, Archives, and Museums
Handle	The Handle System (persistent identifier infrastructure)
IFPI	International Federation of the Phonographic Industry
Interm	Intermediate (self-reported knowledge level)
IPI	Interested Party Information (CISAC identifier for rightsholders)
ISNI	International Standard Name Identifier
ISRC	International Standard Recording Code (ISO 3901; identifier for recordings)
ISWC	International Standard Musical Work Code (ISO 15707; identifier for musical works)
iXML	Production metadata chunk in BWF for structured, exchangeable metadata
Jazz	Jazz/World/Traditional (style label in the stakeholder tables)
JSON-LD	JavaScript Object Notation for Linked Data (web-native linked-data serialization)
JSON-Schema	JSON Schema (validation language for JSON structures)
KPI(s)	Key Performance Indicator(s)
MA+	Master’s degree or higher (MA/PhD)
MS-AIS	Minimal Set for AI-Sound (the proposed framework)
NB	Non-binary (gender designation in tables)
Occ	Occasional (AI usage frequency category)
ORCID	Open Researcher and Contributor ID
OtherEU	Other European Union countries (category used in tables)
PID	Persistent Identifier (e.g., DOI, Handle)
Pop	Pop/Urban (style label in the stakeholder tables)
PT	Portugal (country code used in tables and examples)
P-date	Publication date of the recording (phonogram “P-line” year used in rights contexts)
QA	Quality Assurance
QoQ	Quarter-over-Quarter (growth metric)
RDA	Resource Description and Access (cataloguing standard)
Reg	Regular (AI usage frequency category)

Acronym / Abbrev.	Expanded form (as used in the paper)
RELAX NG	REgular LAnguage for XML Next Generation (schema language)
RfC	Request for Comments (community review process)
SoundArt	Sound art / Podcast / AV (style label in the stakeholder tables)
TCOM (ID3)	ID3 “Composer” text frame
TCOP (ID3)	ID3 “Copyright message” text frame
TPE1 (ID3)	ID3 “Lead performer/soloist” text frame
WCOP (ID3)	ID3 “Copyright/Legal Information URL” frame
WG (WG1/WG4)	Working Group (e.g., COST Action Working Group 1 or 4)
W/M/NB	Women / Men / Non-binary (gender abbreviations in tables)
XSD	XML Schema Definition
XML	eXtensible Markup Language