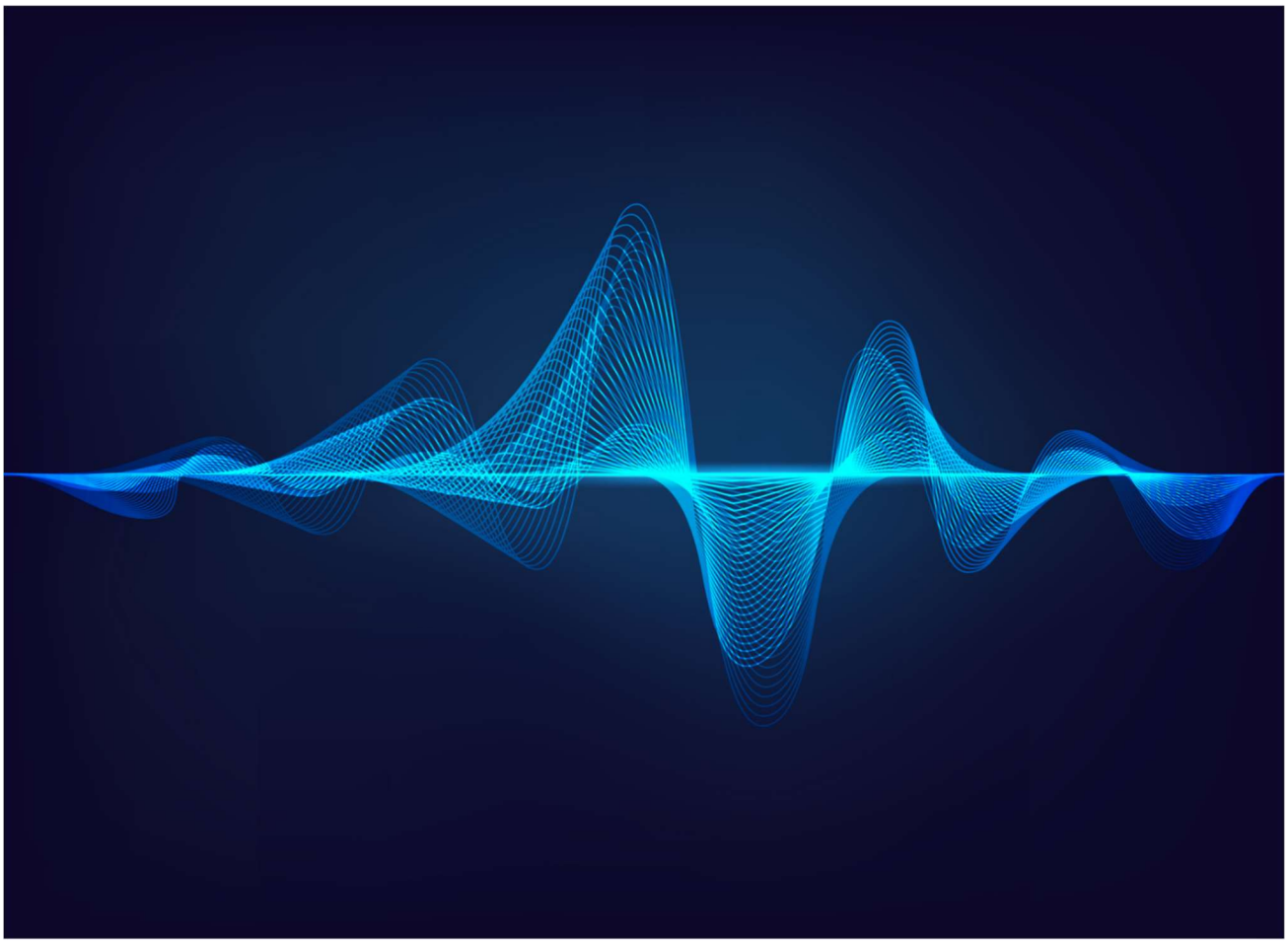


A Green Paper on AI, Data Governance, and Metadata Policies for Europe's Music Ecosystem

OpenMusE

An open, scalable data to-policy pipeline for European music ecosystems

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Project Factsheet

Acronym:	**OpenMusE**
Title:	**An open, scalable data to-policy pipeline for European music ecosystems**
Coordinator:	SINUS Markt- und Sozialforschung GmbH
Creator:	Daniel Antal, Reprex B.V.
Reference:	101095295
Type:	Research and Innovation Action
Program:	Horizon Europe
Start:	1st January 2023
Duration:	36 months
Website:	https://www.openmuse.eu/
Citation:	Antal, D. (2025). A Green Paper on AI, Data Governance, and Metadata Policies for Europe's Music Ecosystem. Open Music Observatory. https://doi.org/10.5281/zenodo.17767905
Consortium:	**SINUS Markt- und Sozialforschung GmbH**, Germany (SINUS), Coordinator **TURUN YLIOPISTO,** Finland (UTU) **UNIVERSITEIT VAN AMSTERDAM,** Netherlands (UVA) **Scuola Superiore di Studi Universitari e di Perfezionament,** Italy (SSSA) **EKONOMICKA UNIVERZITA V BRATISLAVE,** Slovakia (EUBA) **Reprex B.V.,** Netherlands (REPREX) **SYNYO GmbH,** Austria (SYNYO) **MUSIC INNOVATION HUB SPA IMPRESA SOCIALE,** Italy (MIH) **Slovenský ochranný zväz autorský pre práva k hudobným,** Slovakia (SOZA) **Aloaded AB,** Sweden (ALOADED) **Music Export Ukraine,** Ukraine (MEU) **Muzikos Eksporto Fondas,** Lithuania (MXF) **ARTISJUS MAGYAR SZERZOI JOGVEDO IRODA EGYESULET,** Hungary (ARTISJUS) **MUSICAUTOR SDRUZHENIE,** Bulgaria (MUSICAUTOR) **HEARDIS! GMBH,** Germany (HEARDIS)

A Green Paper on AI, Data Governance, and Metadata Policies for Europe's Music Ecosystem

Practical Steps Towards a Decentralised and Open European Music Observatory

Daniel Antal, CFA

2024-11-30

Table of contents

Introduction

There are musical works that are reinterpreted thousands of times across centuries. A symphony by Beethoven or a folk song from the Baltic coast can be heard again and again, each performance producing a new reading of something that never becomes "final." The same is true of sound recordings. Some perennial recordings are rediscovered after sixty years, remastered, and brought into circulation for new audiences.

Music assets, in other words, have an unusually long lifecycle. This is just as true of their documentation — the metadata that accompanies them from creation to archiving. Metadata does not freeze a work or recording in time. Instead, it evolves with it: from the moment of rights registration, through commercial distribution and playlisting, to preservation in a library or archive. Each new interpretation, remix, or reissue generates new metadata; and each new information system demands new connections and contexts.

i Why this Green Paper matters for music professionals?

- Streaming has centralised power in platforms, but left rights-holders with micro-royalties and huge admin burdens.
- Metadata mistakes mean lost revenue — each unlinked ISRC or ISWC is money left on the table.
- AI is already changing music — either it helps you fix documentation and get paid, or it floods the system with untracked works.
- Europe needs federated, cooperative solutions so independents, CMOs, and archives can compete on fairer terms.

There is rarely a single moment when music metadata can be considered complete. Metadata, like music itself, is open to reinterpretation. A name can be reconciled with an identifier; a work can be linked to a new performance; a recording can be embedded in new file formats. Each act of documentation adds layers of meaning and makes the music informative in a new environment.

This is not an invitation to reinvent the wheel. We can read Beethoven's early prints as well as Iris Szeghy's 21st-century scores because music notation — a standardised way of presenting the metadata of musical works — has remained remarkably stable for centuries. Notation shows that standardisation can endure, and that shared conventions make music legible across time, geography, and institutions.

The invention of the computer, and later the internet, introduced new ways to document and transmit music. These innovations brought powerful efficiencies: identifiers like the ISRC and ISWC, digital distribution pipelines, and networked catalogues have enabled the global circulation of music at unprecedented scale. But they also created new fragmentation. Standards proliferated, identifiers failed to interconnect, and workflows designed for one purpose often broke down in another. What was intended as progress sometimes left behind a mess of overlapping, incompatible, or incomplete metadata — a mess that now needs to be cleared up.

Note

This Green Paper is an early-stage policy document, prepared in line with **Open Policy Analysis** and the **Horizon Europe Data Management Guidelines**. It has been released early to allow consultation, incorporate stakeholder input, and provide a transparent development process. This Green Paper extends the analysis developed in the first OpenMusE policy brief on music metadata mainstreaming and EU law (Deliverable D5.6), and its findings are condensed into the second policy brief (Deliverable D5.7), which incorporates wider stakeholder consultations.[1]

Transparency note: Following the principles of **Open Policy Analysis**, all related deliverables and technical documentation are publicly accessible to foster broad engagement and ensure a clear audit trail. Supporting documents for each chapter of this Green Paper are referenced in similar boxes. The current version (and future White Paper drafts) is available at https://zenodo.org/records/17075796. Standardised folders, figures, and bibliographies are available at https://github.com/dataobservatory-eu/open-music-data-white-paper.

Please note that this document puts the Open Music Observatory, a prototype of a modern European Music Observatory developed by the OpenMusE consortium, which is being currently populated with economy, diversity, society, innovation data and has already three federated modules, can reviewed in the technical documentation (see versioned Zenodo DOIs), our viewn on the temporary landing page.

Funding acknowledgement: This project has received funding from the European Union's Horizon Europe programme under Grant Agreement No. 101095295. The views expressed are those of the authors only and do not necessarily reflect those of the European Commission or its agencies.[2]

Citation note: When citing this Green Paper, please use the latest **versioned DOI** available on Zenodo, and include the date of access if referring to material hosted on our GitHub repository.[3] This is an early version (0.9.0.)

Our document has been presented and discussed with industry specialists on the following forums:

- Big Data Value Association, Gaia-X: Dataweek[2] : Introducing a new European music dataspace[4]
- Echoes/ECCH:
- Hungarian stakeholders interested in replication of the Slovak pilot versions [5]
- CISAC: Protecting Creators' Rights in the AI Era: OpenMusE at the European Committee Meeting, Vilnius, 29-30 April [6].
- The Fair MusE - Prelude to a fairermusic industry Fair MusE project[7]
- IAMIC [8]: The *International Association of Music Information Centres* and several key members of the organisation.
- IAML: The *International Association of Music Libraries, Archives and Documentation Centers* and several national chapters and key members [9].

[3]The *Policy Brief 1: Music Metadata Mainstreaming and EU Law* (Senftleben et al. 2024) provides the legal and institutional framing for metadata mainstreaming in European copyright and data law. The present Green Paper builds on that foundation with a lifecycle- and sovereignty-oriented conceptual framework, tested in pilots such as the Slovak Comprehensive Music Database. Its key recommendations are further condensed in *OpenMusE Policy Brief 2: An Open, Scalable Data-to-Policy Pipeline for European Music Ecosystems* (Deliverable D5.7, 2025) (Open Music Europe Consortium 2025), which integrates broader stakeholder consultations (CISAC, IAMIC, IAML, FairMusE, Music360, ECCCH forums, among others) and translates them into policy actions for EU institutions.

[3]This document has been prepared by *Open Music Europe* (OpenMusE) project partners as an account of work carried out within the framework of this contract. Any dissemination of results must indicate that it reflects only the author's view and that the Commission Agency is not responsible for any use that may be made of the information it contains. Neither Project Coordinator, nor any signatory party of *Open Music Europe (OpenMusE) Project Consortium Agreement*, nor any person acting on behalf of any of them:

(a) makes any warranty or representation whatsoever, express or implied, (i) with respect to the use of any information, apparatus, method, process, or similar item disclosed in this document, including merchantability and fitness for a particular purpose, or (ii) that such use does not infringe on or interfere with privately owned rights, including any party's intellectual property, or (iii) that this document is suitable to any particular user's circumstance; or

(b) assumes responsibility for any damages or other liability whatsoever (including any consequential damages, even if advised of the possibility) resulting from your selection or use of this document or any information, apparatus, method, process, or similar item disclosed herein.

[3]Always use the latest versioned DOI when citing this Green Paper, available via Zenodo. If you rely on supporting material hosted in the GitHub repository, please add the date of access in your reference. The figures and charts can be found on FigShare and may be reused separately, citing their DOI and, for context, the Green Paper that contains them.

[4]Jun 5, 2024, Dataweek[2] , Leuven, Belgium.

[5]*Federation possibilities of the Slovak music data sharing space in Hungary* (Antal 2024a)

[6]Protecting Creators' Rights in the AI Era: OpenMusE at the European Committee Meeting, our presentation (Mikš 2025)

[7]We received useful feedback for this Green Ppaer from the project and see further synergies in presenting our policy findings together. https://fairmuse.eu/about/

[8]We presented and discussed these ideas at the *International Association of Music Information Centres* on the General Assembly and Annual Conference 2024 on November 21, 2024, at Music Austria, Vienna. See the presentation and its poster format (Antal 2024d).

[9]We presented and discussed these ideas at the *International Association of Music Libraries, Archives and Documentation Centers* on the General Assembly and Annual Conference 7th and 9th of July 2025 in Salzburg, Austria. See the presentation and its poster format (Antal 2025a, 2025b).

- Polifonia: In October 2023 Polifonia invited a few stakeholders - Podiumkunst.net, the Open Music Observatory, Uni Firenze, IC Fonseca School, Joséphine Simonnot/PRISM, Maria Luisa Onida/D'Istruzione Superiore Leonardo Da Vinci, Carnegie Hall Archive, Municipality of Bologna - for a work session, which gave us a great opportunity to strengthen the metadata framework of our policy recommendations and infrastructure planning.
- Music Futures: the AHRC Creative Industries Cluster project MusicFutures in the United Kingdom.
- Slovak national stakeholders interested in cultural data.[10]
- Wikimedia community and developers[11].
- European music industry stakeholders on *LineCheck 2025* [12]

The CITF's *First Project Report* (Ministry of Education and Culture, Finland, 2025) validates and extends the policy logic of this Green Paper. CITF formulates cross-sectoral requirements for trustworthy, machine-readable copyright infrastructures in the AI era — focusing on identifiers, rights management information, provenance, and federated governance. OpenMusE provides a concrete domain implementation of these ideas within the European music ecosystem, showing how interoperable identifiers, FAIR principles, and data spaces can work in practice for a highly fragmented cultural sector.

Because of this complementarity, we have aligned all major sections of the Green Paper with CITF's three-layer model. The Chapter 2 reflects CITF's foundational layer (authoritative identifiers and repairable rights metadata); the Chapter 3 corresponds to the semantic and technical layers (federated registries, mappings, and interoperability profiles); and the Chapter 4 translates CITF's AI-era requirements into music-specific governance questions. This deliberate alignment is meant to make both documents usable in parallel: CITF as the horizontal framework for copyright data in Europe, and this Green Paper as a domain-specific blueprint for music that can be reused, extended, or replicated in other cultural sectors[13].

[10] Based on a memorandum of understanding with a broad range of public and private stakeholders, (Ministerstvo kultúry SR and Open Music Europe 2023) we developed a model for renewing statistical production for better cultural and music statistics (Antal 2023).

[11] Our work was presented in the Technology session of the *Wikimedia CEE Meeting 2024* in Istanbul, and the *Wikimedia CEE Meeting 2025 in Thessaloniki*, and the *Wikidata Conf 2025* online; we have built relationships with various national chapters and the Wikidata and Abstract Wikipedia teams, and joined the Wikidata Ontology Cleanup Task Force and the Wikidata Mereology Task Force to help the coordiantion of our open source technology, data curation and dissemination efforts. (Antal 2024b, 2025c; Antal, Pigozne, and Federico 2025).

[12] *Open Access Music Dataspaces – Open Music Observatory* presented on LineCheck 2025 (Mikš and Antal 2025)

[13] We are planning to give feedback to the (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzu 2025) on 19 November, and we asked the authors of the report to comment on our Green Paper, too.

Glossary

Music terms

`audio recording`: fixation of sounds (ISO 2019a)

`creator`: in the context of this policy paper, we use the broad term for the `arranger`, `author`, `composer`, `lyricist`; for individual definitions see ISWC standard (ISO 2022)

`DSP` or `digital streaming platform`: Digital service providers (DSPs), or Digital Streaming Platforms are companies or organisations that provide access to services online. DSPs can provide access to music downloads, like Apple's iTunes Store, or access to streaming music like Spotify, or even provide satellite-delivered content such as SiriusXM in the USA.

`expression`: intellectual or artistic realisation of one and only one work
Note: may take the form of a notation , sound, image, object, movement or text (ISO 2017b)

`manifestation`: physical embodiment of an expression (ISO 2017b)

`movement`: A principal division of a musical work. (ISO 2022)

`music video recording`: fixation of sounds synchronized with pictures or moving pictures where (a) the fixed sounds are wholly or substantially a musical performance or (b) the *recording* is intended for viewing in association with a recording of a musical performance. This definition includes music videos and concert recordings, together with music-related interviews and documentaries, but does not extend to genera! audiovisual material, even if it includes music.(ISO 2019a)

`musical work`: composed of a combination of sounds, with or without accompanying text (ISO 2022)

`original title`: A title given to the work by its creator(s) shown in its original language. (ISO 2022)

`formal title`: A standardized title in which the elements are arranged in a pre-determined order, such as titles created for classical works. (ISO 2022)

`rights management (organisations)`: the function of managing the rights on behalf of rights owners. It can be companies whose sole purpose is to ensure that content that has been licensed has delivered royalties that are identified and accounted for. The role can be taken by collective management organisations or by private companies on behalf of songwriters, composers, performers, music publishers, or record labels.

`original version`: The first established form of a work. (ISO 2022)

`performer`: The performer of a musical work; in case of a sound recording, the performer whose performance is fixed in the recording. They may be entitled to neighbouring rights in sound recordings. In some contexts, the `performer` is part of the broadly defined `creator` group.

`producer`: The person or legal entity that produces the recorded fixation of the sound recording. They are entitled to neighbouring or sound recording copyrights. In some contexts, the `producer` is part of the broadly defined `creator` group.

`track`: single recording on a sound carrier (ISO 2017b)

`work`: distinct, abstract creation of the mind whose existence is revealed through one or more expressions (e.g. a performance) or manifestations (e.g. an object) (ISO 2022)

Data terms

`conceptualisation`: an abstract, simplified view of some selected part of the world, containing the objects, concepts, and other entities that are presumed of interest for some particular purpose and the relationships between them.

`data`: reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing Note 1 to entry: Data can be processed by humans or by automatic means.[SOURCE: ISO/IEC 2382:2015, 2121272] (ISO 2019b)

`database`: collection of `data` organized according to a conceptual structure describing the characteristics of these data and the relationships among their corresponding entities, supporting one or more application areas. [SOURCE: ISO/IEC 2382:2015, 2121413] (ISO 2019b)

`data set` or `dataset`: identifiable collection of `data` available for access or download in one or more formats [SOURCE: Adapted from ISO 19115-2:2009, 4.7] *Beware: various conceptual and information models use different dataset definitions.* (ISO 2019b)

`data model`: description of the organization of data in a manner that reflects an information structure [SOURCE:ISO 28258:2013, definition 3.9] (ISO 2017b); or pattern of structuring data in a database according to the formal descriptions in its information system and according to the requirements of the database management system to be applied (ISO 2023b)

`big data`: extensive datasets – primarily in the data characteristics of `volume`, `variety`, `velocity`, and/or `variability`. – that require a scalable technology for efficient storage, manipulation, management, and analysis. note : Big data is commonly used in many different ways, for example as the name of the scalable technology used to handle big

data extensive datasets. (ISO 2019b)

`data portability`: Ability to easily transfer data from one system to another without being required to re-enter data.

`data science`: extraction of actionable knowledge from `data` through a process of discovery, or hypothesis and hypothesis testing (ISO 2019b)

`file`: named set of records treated as a unit [SOURCE:ISO/IEC 2382:2015, 04.07.10] (ISO 2023b)

`identifier authority`: trusted organisation responsible for issuing or validating identifiers for agents or assets. Includes registration agencies, national libraries, CMOs, or similar bodies. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 85)

`interoperable identifier mappings`: explicit, machine-interpretable relationships describing how different identifier systems relate to each other, including equivalence or preferred authority. CITF requires clear mappings to ensure that agent and asset identifiers remain usable across systems. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 86)

`knowledge base` or `K-base`: database that contains inference rules and information about human experience and expertise in a domain. 1: In self-improving systems, the knowledge base additionally contains information resulting from the solution of previously encountered problems. The terms `knowledge base` and `K-base` are standardized by ISO/IEC [ISO/IEC 2382-1:1993]. (ISO 2023b)

`knowledge graph`: a `knowledge representation` that uses a graph-structured data model to represent and operate on data. (ISO 2023b)

`metadata`: data that define and describe other data (ISO 2023a); we use the more functional definition "a statement about a potentially informative object." (Pomerantz 2015);metadata is `data` about data or data elements, possibly including their data descriptions, and data about data ownership, access paths, access rights and data volatility (ISO 2023b).

`semantic interoperability layer`: shared framework enabling the consistent interpretation of meaning across systems without requiring a single unified ontology. Based on the CITF concept of a global semantic layer for rights and metadata. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 33)

AI & Systems Terms

`access level`: level of authority required from an entity to access a protected resource (ISO 2023b)

`AI` or `artificial intelligence` is, according to the EU definition (harmonised with the OECD) is a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. The technical definition "an engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives." ISO/IEC 22989:2022 Artificial intelligence concepts and terminology ISO/IEC 42001:2023 – AI Management Systems

`AI-generated content flag`: metadata indicator specifying whether part or all of a work has been generated by an artificial intelligence system. Required to preserve provenance and strengthen content trust. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 20)

`AI-related reservation`: machine-readable expression signalling restrictions on the use of content for AI training, inference, or generative reuse. Required to support lawful AI development in cultural heritage and creative sectors. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 34)

`algorithm`: finite ordered set of well-defined rules for the solution of a problem [SOURCE: ISO/IEC 2382-1:1993] (ISO 2023b)

`identifier`: data string or pointer that establishes the identity of an item, organization or person alone or in combination with other elements [SOURCE:ISO 8459:2009, definition 2.27, modified] (ISO 2017b)

`interoperability`: Ability of two or more systems or applications to exchange information and to mutually use the information that has been exchanged. [SOURCE:ISO/IEC 19941:2017] (ISO 2017a)

`cloud computing`: paradigm for enabling network access to a scalable and elastic pool of shareable physical or virtual resources with self-service provisioning and administration on-demand. (ISO 2019b)

`cluster`: <distributed data processing> set of functional units under common control [SOURCE:ISO/IEC 2382:2015, 2120586] (ISO 2023b)

`document`: named, structured unit of text and possibly images that can be stored, edited, retrieved, and exchanged among systems or users as a separate unit (ISO 2023b)

`expert system`: knowledge-based system that provides for solving problems in a particular domain or application area by drawing inferences from a knowledge base developed from human expertise
Note 1: The term "expert system" is sometimes used synonymously with "knowledge-based system", but should be taken to emphasize expert knowledge.
Note 2. Some expert systems are able to improve their knowledge base and develop new inference rules based on their experience with previous problems. (ISO 2023b) Expert systems fall under the definition of the AI Act.

`organization`: unique framework of authority within which a person or persons act, or are designated to act, towards some purpose [SOURCE:ISO/IEC 6523-1:1998, definition 3.1] (ISO 2017b)

`persistent identifier` or `PID`: unique identifier that ensures permanent access for a digital object by providing access to it independently of its physical location or current ownership [SOURCE:ISO 24619:2011, definition 3.2.4] (ISO 2017b)

`ontology`: formal, explicit specification of a shared conceptualization.
Note to entry: An ontology typically includes definitions of concepts and specified relationships between them, set out in a formal way so that a machine can use them for reasoning. [SOURCE:ISO 25964-2:2013, definition 3.57] (ISO 2017b)

`thesaurus`: controlled vocabulary and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms.
Note 1 to entry: The purpose of a thesaurus is to guide both the indexer and the searcher to select the same preferred term or combination of preferred terms to represent a given subject. For this reason a thesaurus is optimized for human navigability and terminological coverage of a domain. [SOURCE:ISO 25964-1:2011, definition 2.62] (ISO 2017b)

`machine-readable opt-out declaration`: metadata field used to reserve rights under Article 4 of the DSM Directive for text-and-data mining. CITF identifies this as one of the most urgent requirements for AI transparency and compliance. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 14)

`NERD`: named-entity recognition and disambiguation is a natural language processing technique that aims to resolve the ambiguity that arises from named entities in text.

`provenance assertion`: claim describing the origin, authorship, creation circumstances, or authenticity of a digital asset. CITF recommends that provenance assertions remain intact throughout distribution and reuse. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 101)

`taxonomy`: scheme of categories and subcategories that can be used to sort and otherwise organize itemized knowledge or information [SOURCE:ISO 25964-2:2013, definition 3.83 modified] (ISO 2017b)

Data protection terms

`audit trail`: chronological record documenting actions taken on a digital artefact or metadata object, required for traceability and accountability. CITF stresses its role in rights and AI governance. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 101)

`DPIA`: Data Protection Impact Assessment (DPIA) is a process used to identify and minimize the risks associated with processing personal data.

`DPO`: the Data Protection Officer (DPO) is an individual designated by an organization to oversee its compliance with data protection laws, such as the GDPR. They act as a point of contact for data subjects and supervisory authorities, and they advise on and monitor data protection practices within the organization.

`GDPR`: The General Data Protection Regulation (GDPR) is a legal framework made by the European Union that sets guidelines for the collection and processing of personal information from individuals who live in and outside of the European Union.

Data curation and collection terms

`aggregation`: acquisition of sensitive information by collecting and correlating information of lesser sensitivity (ISO 2023b)

`collection`: gathering of items assembled on the basis of some common characteristic, for some purpose, or as the result of some process (ISO 2017b)

`holdings`: totality of documents in the custody of an information and documentation organization (ISO 2017b)

`digital collection`: collection formed by a collection process on existing data and data sets where the collected data is in digital form (ISO 2017b)

`library collection`: all documents provided by a library for its users(ISO 2017b)

`anthology`: document consisting of a collection of full documents or of extracts, usually of literary works (ISO 2017b)

`exhibition`: curated display of objects on a clear concept and communicating a message [SOURCE:ISO 18461:2016, definition 2.4.6 modified] (ISO 2017b)

`curator`: person responsible for overseeing a collection or exhibition (ISO 2017b)

`data curation`: managed process, throughout the data lifecycle, by which data/data collections are cleansed, documented, standardized, formatted and interrelated (ISO 2017b)

`register`: an official list or record of names or items; it aims to be a complete list of the objects in a specific group of objects or population, for example, all copyright-protected musical works in a country, or all legal person enterprises in another country;

a document, usually a volume, in which data are entered in a formal manner by a statutory authority Note 1 to entry: In modern usage, usually a database. (ISO 2017b)

`registration`: act of giving an entity a unique identifier on its entry into a system (ISO 2017b) a set of rules, operations, and procedures for inclusion of an item in a registry (ISO 2023a)

`registrant`: organization or person that has either registered an authentication protocol or registered the adoption of an authentication protocol [SOURCE: ISO/IEC 24727-6:2010, definition 3.4] (ISO 2017b); an entity wishing to assign an ISRC to an applicable recording (ISO 2019a); a `party` that requests an ISNI from the Registration Authority (ISNI 3.2 (ISO 2012, p15))

`party`: natural person or legal person, whether or not incorporated, or a group of either (ISO 2012)

`usage logging`: recording of access, reproduction, or other interactions with an asset to support proportional remuneration, compliance monitoring, and transparency obligations. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 102)

Rights management terms

`rights-management information`: machine-readable metadata describing the ownership, licensing terms, permitted uses, and exceptions related to an asset. CITF requires RMI to be trustworthy, interoperable, and protected against alteration. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 84)

`rights expression`: structured, machine-readable statement describing copyright status, rights holders, exceptions, limitations, and licensing conditions of an asset, supporting automated interpretation across jurisdictions. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 84)

`usage conditions`: machine-readable description of permitted or restricted uses of an asset under copyright, contract, or statutory provisions. Required for automated rights-aware systems. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 84)

Statistical terms

`administrative records`: data generated by a non-statistical source, usually a public body, the main aim of which is not the provision of statistics.

`code list`: predefined list from which some statistical coded concepts take their values (ISO 2013)

`data pipeline`: a method in which raw data is ingested from various data sources and then ported to data store.

`FAIR` or `FAIR Guiding Principles for scientific data management and stewardship`: guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets, emphasising machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention.)

`indicator`: the representation of statistical data for a specified time, place or any other relevant characteristic, corrected for at least one dimension (usually size) so as to allow for meaningful comparison.

`microdata`: non-aggregated observations or measurements of characteristics of individual units, without direct identifier.

`observation unit`: an identifiable entity about which data can be obtained, it is also often called a *statistical unit* or *data subject* in case of a natural person.

`Open Policy Analysis Guidelines`: a set of information management rules to make policy analysis more transparent.

`personal data`: any information relating to an identified or identifiable natural person.

`pseudonymisation`: processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information.

`survey`: a systematic examination and record of a physical or social area and its features so as to construct a map, plan, or description. In social sciences it usually refers to a well-structured questionnaire and answers given to its items by a target population.

`statistics`: quantitative and qualitative, aggregated and representative information characterising a collective phenomenon in a considered population.

`visualisations`: schematic charts, drawings, photographs, and their collages will as still image files that help to explain the relationship between information carriers, data points, or processes.

Registers, authorities, standards and identifiers

`agent identifier`: persistent identifier assigned to an author, performer, contributor, or other agent. CITF requires agent identifiers to be standardised, trustworthy, and interoperable. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 85–86)

`asset identifier`: persistent identifier used for musical works, sound recordings, editions, audiovisual items, or other cultural objects. CITF requires asset identifiers to be resolvable

and interoperable across systems. (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, 86)

`IČO`: The organisation identification number (IČO) is an identifier assigned to all types of legal entities, entrepreneurs and public authorities by the Statistical Office of the Slovak Republic. The Czech Republic's organisation identifier is also called IČO. (→ `agent identifier`)

`OpenCorporates`: a public corporation database which sources data from national business registries. (→ `agent identifier`)

`ISNI`: an ISO certified global standard number for identifying the millions of contributors to creative works and those active in their distribution. (→ `agent identifier`)

`VIAF`: The Virtual International Authority File (VIAF) is an international service that consolidates multiple name authority files into a single database. Their primary goal is to enhance the efficiency and usability of library authority files by linking and merging widely used authority records and making them accessible online.

`VIAF ID`: The VIAF (Virtual International Authority File) combines multiple name authority files into a single OCLC-hosted name authority service. (→ `agent identifier`)

`ISRC`: The International Standard Recording Code (ISRC) is a standard identifying code that can be used to identify sound recordings and music video recordings so that each such recording can be referred to uniquely and unambiguously. (→ `asset identifier`)

`ISWC`: The purpose in creating an ISWC for musical works is to enable more efficient administration of rights to those works on a worldwide basis. The ISWC provides an efficient means of identifying musical works in computer databases and related documentation and for the exchange of information between rights societies, publishers, record companies and other interested parties on an international level.

`ISBN`: the International Standard Book Number is an identification system for the publishing industry and its supply chains. (→ `asset identifier`)

`ISMN`: The International standard music number (ISMN) was developed by, and for, the music publishing sector as a separate system to complement the International standard book number (ISBN). The existence of the ISMN as a separate identifier system makes it possible to identify printed and notated music as a distinct category of publication within the global supply chain and to develop trade directories and similar services for the specialized market for music publications. (→ `asset identifier`)

`ISCC`: The International Standard Content Code (ISCC) is an identifier for numerous types of digital assets. (→ `asset identifier`)

`DOI`: The Digital Object Identifier is a standardised unique number given to many (but not all) articles, papers and books, by some publishers, to identify a particular publication.

`ORCID`: the Open Researcher and Contributor ID is a unique, persistent identifier free of charge to researchers. (→ `agent identifier`)

`URI`: A Uniform Resource Identifier (URI) is a string of characters used to identify a resource on the internet. This resource can be either abstract or physical, such as a website, an email address, or a file. URIs are essential for enabling interactions with resources over a network using specific protocols.

`DDI`: The `Data Documentation Initiative` is originating for the world of social sciences data archives and more and more in use in statistical organisations for the documentation of microdata.

`Wikibase`: Wikibase is a software system that help the collaborative management of knowledge in a central repository. It was originally developed for the management of Wikidata, but it is available now for the creation of private, or public-private partnership knowledge graphs. It is developed by Wikimedia Deutschland.

`SDMX`: Statistical Data and Metadata eXchange (SDMX), is an international initiative that aims at standardising and modernising ("industrialising") the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries.

`CIDOC-CRM`: The conceptual model of CIDOC, the standard conceptualisation of collection management systems in heritage organisations.

`RiC`: *Records in Context* is a new conceptual model that replaces the four most important international archiving standards.

`DCTERMS` or `DCMI`: the Dublin Core Metadata Terms is a vocabulary of metadata terms developed and maintained by the Dublin Core Metadata Initiative (DCMI). These terms are used to describe various aspects of digital resources, such as web pages, documents, and other online content. They provide a standardized way to assign metadata to resources, making them easier to discover, manage, and exchange.

`RDFS`: the Resource Description Framework Schema is an extension of the Resource Description Framework (RDF) that provides a vocabulary for describing classes and properties of resources within an RDF graph.

`EDM`: the Europeana Data Model is a framework for collecting, connecting, and enriching cultural heritage metadata. It's designed to facilitate the sharing and reuse of cultural heritage information by providing a standardized way to represent and link data.

`PROV-O`: the Provenance ontology is a formal ontology developed by W3C to represent and interchange provenance information.

`MARC`: MAchine-Readable Cataloging, is a standard digital format used by libraries to represent and exchange bibliographic information.

`DCAT`: an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web.

Organisations

AEPO-ARTIS: Organisation representing European artists-performers. Regroups most of the European CMO representing performers.

ALOADED: is a company which distributes and exploits recordings.

CISAC: The International Confederation of Societies of Authors and Composers is an international non-governmental, not-for-profit organisation that aims to protect the rights and promote the interests of creators worldwide.

CNM (former CNV): the Centre National de la Musique is a public organisation managing a tax on concert tickets

EMO: The European Music Observatory (EMO) is envisioned as a hub for collecting and analysing data on the music sector across Europe. Its primary aim is to address the current gaps and inconsistencies in music data collection, which have been a significant challenge for the sector.

Europeana: a digital platform provided by the European Union that aggregates digitized cultural heritage from institutions across Europe.

GESAC: The European Grouping of Societies of Authors and Composers (GESAC) comprises of 32 European authors' societies in music, audiovisual, visual arts, literature and drama.

IAML: International Association of Music Libraries, Archives and Documentation Centres

IAMIC: International Association of Music Centres, an international network of organisations that collectively and collaboratively provides information and promotes the music of their countries or regions.

ICMP: the global trade body representing the music publishing industry worldwide.

SCAPR: International association for the development of the practical cooperation between performers' collective management organisations (CMOs)

SOZA: SOZA (Slovenský ochranný zväz autorský pre práva k hudobným dielam, Slovak Performing and Mechanical Rights Society) is a legal entity, non-profit civic association of authors and publishers of musical works, association of natural persons and legal entities.

Hudobné Centrum: Music Centre Slovakia is a music organisation with a mission to promote Slovak contemporaly music.

Other abbreviations

`CEEMID`: the Central European Music Industry Databases is a multi-country project that was a predecessor of Reprex's Digital Music Observatory

`DSP`: Digital service providers (DSPs), or Digital Streaming Platforms are companies or organisations that provide access to services online.

`EIF`: The European Interoperability Framework (EIF) is a set of recommendations and guidelines that aims to facilitate communication and collaboration between public administrations, businesses, and citizens within the European Union and across national borders.

`ECCCH`: The European Collaborative Cloud for Cultural Heritage is a European Union initiative for a digital infrastructure that will connect cultural heritage institutions and professionals across the EU.

`EOSC`: The European Open Science Cloud (EOSC) aims to create a trusted, open, and multidisciplinary environment for researchers and innovators in Europe.

`PPP`: A Public-Private Partnership (PPP) is a collaborative arrangement between government entities and private sector companies aimed at financing, designing, implementing, and operating projects or services traditionally provided by the public sector.

`RDM`: Research Data Management refers to the suite of practices, policies, and processes used to handle data throughout the lifecycle of a research project.

`W3C`: The World Wide Web Consortium (W3C) is an international community that develops standards for the World Wide Web. Their mission is to lead the Web to its full potential by creating technical specifications and guidelines that are designed to be open and royalty-free. These standards include HTML, CSS, and other web technologies, which ensure that web content is accessible across different browsers and devices.

Our glossary is harmonised with relevant music-sector specific standards and with the

- ISO Information technology *Vocabulary* (ISO 2023b); *Cloud computing — Taxonomy based data handling for cloud services* (ISO 2020); *Cloud computing — Interoperability and portability* (ISO 2017a); *Metadata registries (MDR) — 1. Framework* (ISO 2023a) standards and the *Information and documentation — Foundation and vocabulary* (ISO 2017b) standard.
- ISO Information technology *Artificial intelligence — Concepts and terminology* (ISO/IEC 2022) and *Artificial intelligence — Management system* and (ISO/IEC 2023) standard's vocabulary.

1 Policy context and problem map

The European music ecosystem has undergone disruptive transformations in recent decades. In the 2010s, the arrival of agentic AI in streaming platforms radically reconfigured distribution and consumption. These systems centralised global sales, expanding the commercially available repertoire in a typical EU country from roughly 100,000 titles to over 100 million titles competing for attention. At the same time, the average transaction value collapsed from around €18 (in current prices) to less than €0.005. This shock hollowed out much of the traditional infrastructure — record stores, radios, and music television — and shifted value capture toward data-driven platforms able to control access through recommender algorithms.

In the 2020s, the rise of generative AI further exacerbates this situation. Large-scale models can mass-produce new compositions and recordings, often imitating or plagiarising patterns of human creators. This inflates supply, undermines the position of professional authors and performers, and aggravates existing problems of remuneration and discoverability.[1]

EU-level studies and policy frameworks have recognised these dynamics and increasingly frame them as systemic challenges. The *Feasibility Study for the Establishment of a European Music Observatory* diagnosed the fragmented, scarce, and poorly harmonised nature of music data collection across Member States, calling it the fundamental reason for an EU-level observatory. The *Music Ecosystem 2025 study* reframes the sector as an interconnected ecosystem, where platformisation, market consolidation, and emerging technologies like AI interact with broader societal challenges such as precarity, gender inequality, and sustainability. The European Parliament, in its *Resolution on cultural diversity and the conditions for authors in the European music streaming market*, echoed these concerns with explicit calls for reform.[2]

[1] *Music Ecosystem 2025: Study on the Music Ecosystem* (Music Moves Europe 2024); it frames the sector as an adaptive, networked ecosystem, highlights AI's ability to disrupt on pp. 6–7, and mentions it as an opportunity particularly on p. 23. *Feasibility Study for the Establishment of a European Music Observatory* (Commission et al. 2020); stresses the fragmented, scarce, and poorly harmonised nature of music data (pp. 9–10), the need for cooperation with rights organisations, statistical agencies, and industry stakeholders (p. 61), and introduces CEEMID as a best practice (pp. 147–148). CEEMID emerged from Budapest, Bratislava, and Zagreb as an early effort to address data poverty in Eastern EU Member States.

[2] *European Parliament Resolution on cultural diversity and the conditions for authors in the European music streaming market* (European Parliament 2024); it recognises streaming as the dominant global revenue source while leaving many authors with very low income (recitals F–H), stresses accurate metadata allocation at the time of creation using identifiers ISWC, ISRC, ISNI, IPI, and IPN (recital R, and 9.), highlights the lack of quality data to properly identify authors, performers, and rights holders (recital L), and warns that AI-generated tracks are flooding streaming platforms, aggravating discoverability and remuneration imbalances (recital O).

A third major contribution to this landscape is the 2025 *CITF First Project Report*, coordinated by the National Libraries of Finland and Latvia. CITF identifies open identifiers, machine-readable rights metadata, and national libraries as core components of a future copyright infrastructure. It introduces a three-layer model (foundational, semantic, technical) and provides lifecycle analysis of protected works in the AI era. Its findings complement the *Music Ecosystem 2025* and *EMO feasibility* studies by foregrounding the role of cultural heritage institutions and the need for trustworthy, interoperable copyright registries[3].

Our policy brief positions itself within this landscape. It aims to support and extend the Music Moves Europe framework by highlighting six crucial dimensions:

1. **Practical solutions**, grounded in dialogue between research and industry, and inspired by concrete experiences with open, federated data-sharing approaches.
2. **Potential pitfalls** where well-meaning initiatives may clash with legacy systems, existing business practices, or contradictions in legislation.
3. **Legal and operational conflicts**, such as the tension between GDPR's data protection regime and the Berne Convention's requirement of author attribution.
4. **Cooperation and workflow sharing**, recognising that no single actor can bear the full burden of metadata documentation.
5. **Technology**, including automation, entity recognition, reconciliation, and persistent identifiers.
6. **AI adaptation and cooperative infrastructures**, since most stakeholders cannot attract or retain scarce AI expertise.

By foregrounding these issues, the brief complements the calls of the *Music Ecosystem 2025 study* and the *European Music Observatory feasibility study*, while remaining attentive to the practical challenges of implementation across Europe's diverse music and cultural landscapes.

1.1 Three structural pressures

Three structural pressures frame today's metadata challenges:

1. **Extreme efficiency pressure.** Music is now monetised in micro-transactions worth a fraction of a cent. Each metadata mistake means lost royalties, while big-tech platforms enjoy economies of scale that self-releasing artists, small labels, and national CMOs cannot match. National libraries in many countries already maintain massive copyright-protected collections and identifier systems, which could be cross-utilised with CMOs.

[3] *Interoperable, trustworthy, and machine-readable copyright data in the AI era. Report of the CITF First Project* (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025)

2. **AI-driven disruption.** Agentic AI in streaming platforms has already displaced much of the traditional retail and promotion infrastructure. Both pre-deployment and post-deployment of AI affect reproduction, distribution, and attribution rights. Generative AI risks flooding platforms with derivative works and further destabilising discoverability and revenues. Yet AI tools could also support documentation and reconciliation — if governance frameworks can enable them.

3. **Governance and incentive conflicts.** Identifiers such as ISWC, ISRC, ISNI, and IPN are essential for attribution and royalty distribution, but are maintained under costly, largely private regimes. Public policy increasingly demands more open metadata, but sustaining investment in these registers remains a challenge. Opt-out rights in AI training and the need for harmonised opt-out registries[4], further complicate governance and incentive structures.

These pressures mean that improving metadata is not only a matter of technical interoperability. It is also a question of economic sustainability, legal coherence, and cultural policy.

1.2 National and European pilots as anchors

From the outset, we draw on concrete pilots that illustrate both the problems and possible solutions. Two of them — the **Slovak Comprehensive Music Database (SKCMDb)** and **Unlabel** — will recur throughout this paper as reference points. Together, they anchor the three thematic chapters: **curation** (Chapter 2), **observatory** (Chapter 3), and **AI** (Chapter 4).

1.2.1 The Slovak Comprehensive Music Database (SKCMDb)

SKCMDb is our national pilot for federated metadata governance. It links together data from collective management (SOZA), national and city libraries, and archives, while ensuring that works can also be discovered in the digital environments where people actually listen: Spotify, YouTube, Apple Classical, and others.

A further layer reconciles this metadata with the Slovak Statistical Office via a **Satellite Business Register**, so that cultural production is visible in official economic data.

The **SKCMDb** is anchored in the **Memorandum of Understanding** signed between collective management organisations (SOZA, SLOVGRAM), cultural institutions (Hudobné centrum, Slovak National Library, Hudobný fond), and Reprex. SKCMDb's strategy of combining copyright data (SOZA), neighbouring rights (SLOVGRAM), and national library authority control directly reflects CITF's observation that national libraries must be integrated into copyright infrastructure, not treated as purely heritage institutions.

[4] As emphasised by CITF on pp. 17, 23–24 (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, p17, pp. 23-25).

This MoU formalises a federated governance model where:

- **Attribution** (names of authors, performers, composers) is preserved as legally mandatory under copyright law.
- **Privacy** is safeguarded by layered access: public data (names, works, identifiers) circulate broadly, while sensitive data (e.g., addresses, birth dates) remain restricted.
- **Interoperability** is achieved by aligning with VIAF, ISNI, ISWC, ISRC, and Europeana.

As such, the Memorandum provides the *legal and institutional foundation* for SKCMDb, turning a technical pilot into a national dataspace aligned with the EU Data Strategy.

Our pilot also

The SKCMDb in action

The chart illustrates the biography and works of Slovak composer **Iris Szeghy** as an example:

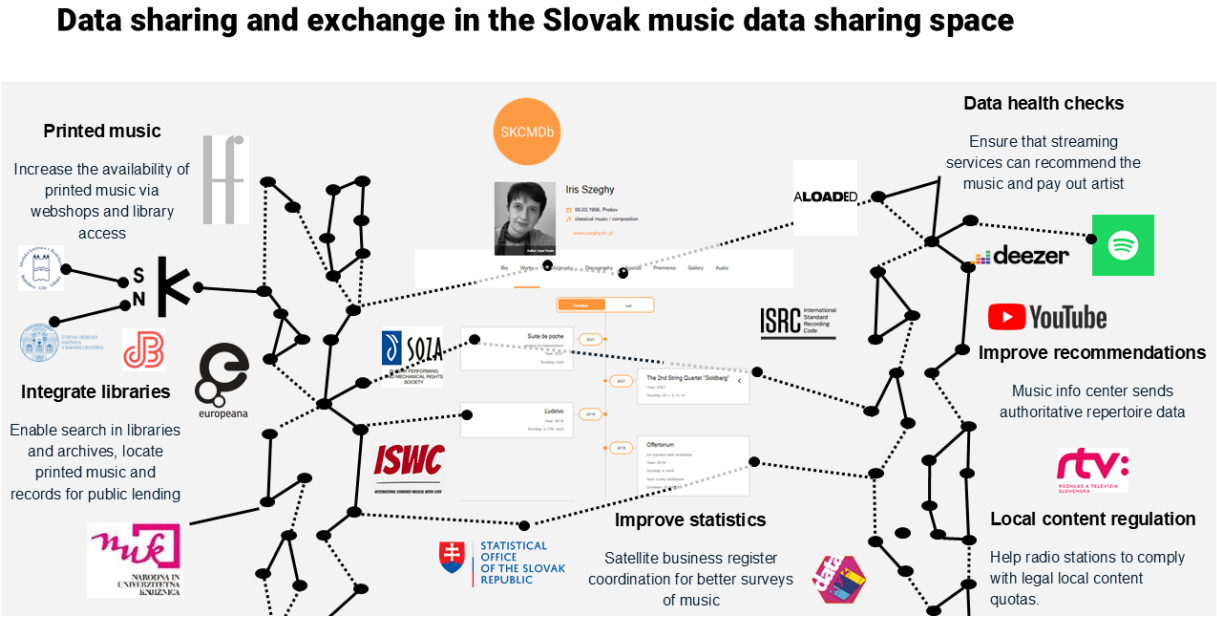


Figure 1.1: A slide taken from: SKCMDb: Interoperability of Music Libraries and Archives with Public and Private Music Services (presentation at the IAML 2025 conference in Salzburg) <https://zenodo.org/records/16634558>

- **Left side:** reconciliation of her works across SOZA, the Slovak National Library, the Bratislava City Library, and archives.
- **Right side:** linking to listening platforms (Spotify, YouTube, Apple Classical).
- **Bottom:** reconciliation with the Slovak Statistical Office via the Satellite Business Register.

SKCMDb thus acts as a bridge between cultural memory institutions, rights management, digital distribution, and public policy.

SKCMDb provides a pragmatic response to fragmentation and duplication. It anchors the discussion of **preventive metadata strategies** in Chapter 2.

This challenge is not unique to Slovakia. A recent Horizon Europe policy brief has highlighted how **inadequate metadata infrastructures** and **fragmented European initiatives** risk leaving the field open to dominance by extra-European players (for example, the US Mechanical Licensing Collective).[5]

1.2.2 Unlabel

If SKCMDb focuses on building preventive infrastructures, **Unlabel** demonstrates how to repair the past. It is a collaborative pipeline connecting archives, libraries, collective rights organisations, and distributors to bring under-documented repertoires into the global digital supply chain.

A striking example is the case of **Hilda Griva**, a bilingual Livonian–Estonian artist active in the interwar Finno-Ugric revival. Her recordings were rediscovered in the Latvian Archives of Folklore but lacked the metadata required for circulation. Through Unlabel, we translated and enriched her records, reconciled them with international authorities, and extended them with **DDEX catalogue transfer metadata**, enabling release via Spotify, YouTube, and Apple Music.

Our multi-layer model (DDEX, DCTERMS, RiC patterns, and rights metadata) aligns with CITF's three-layer structure: DCTERMS in the foundational layer, RiC and DDEX conceptual mappings in the semantic layer, and DDEX catalogue-transfer formats in the technical layer.

> **Note**
>
> **Infobox: Unlabel and Hilda Griva**
>
> - Metadata repair began with archival records in the Latvian Archives of Folklore.
> - Records were translated, enriched, and reconciled with Wikidata, MusicBrainz, and VIAF.
> - DDEX-compliant catalogue transfer metadata enabled digital distribution.
> - The enriched catalogue allowed Hilda Griva's recordings to be released and discovered globally.

[5] See *Policy Brief 1: Music Metadata Mainstreaming and EU Law* (Senftleben et al. 2024) (Deliverable D5.6, OpenMusE project). That brief emphasises that without a European metadata infrastructure, EU repertoires may remain underexploited and culturally invisible, while foreign platforms consolidate hegemony. The present Green Paper extends on this line of argument by focusing on lifecycle-based interoperability and federated observatories as safeguards for European sovereignty.The Policy Brief 1 Annex references the Slovak Listen Local / SKCMDb project as a national pilot, underlining its relevance for EU-level policy design. The *Green Paper* complements this by situating the MoU as a replicable governance framework for federated metadata spaces.

Unlabel demonstrates how public heritage institutions and private distributors can cooperate through shared standards. It anchors both the **curative AI approaches** in Chapter 4 and the **observatory perspective** in Chapter 3.

1.3 Quest for efficiency

Technological progress, digitisation, automation, and now AI have transformed the music industry more dramatically than most sectors. After the collapse of the CD era under peer-to-peer piracy, a newly configured recording industry emerged around global platforms. Traditional retail and wholesale jobs largely disappeared, replaced by streaming platforms such as YouTube, Apple Music, and Spotify.

This shift coincided with a structural **devaluation of music**. The licensed streaming model never recovered the real revenues of the pre-collapse recording market, and from this diminished base, platforms take a significant share. Where a CD sale once brought around €10–18 in today's terms, the unit of account in streaming is a fraction of a cent — typically $0.003–0.005 per play.

To replace the economic weight of a single album sale, a rightsholder must now process and account for roughly 4,000 successful streams. This is not merely an economic shift, but an **administrative revolution**. The documentation efficiency needed to handle millions of micro-transactions profitably is far higher than in the pre-streaming era.

Streaming platforms are genuine big-data companies. Alphabet's YouTube, Apple, and Spotify operate at a scale where billions of transactions and hundreds of millions of assets can be managed by autonomous agents and recommender engines. But the typical rightsholder — a self-releasing artist, an independent label, or even a national collective rights agency — works at a scale where each metadata mistake means lost royalties, and where IT or documentation specialists are often absent altogether. This asymmetry is so stark that even major CMOs rely on shared infrastructures like the digital services of "Mint" to manage repertoire at scale.

Music, then, is now sold in **extremely low-value transactions mediated by autonomous agents**. This reality enforces a very strong pressure on the entire ecosystem to improve data interoperability and metadata quality.

As the CITF report emphasises, AI introduces legal lifecycle pressures: both the training and deployment of AI systems may trigger reproduction, distribution, adaptation, and communication-to-the-public rights. This amplifies the economic consequences of metadata fragmentation: missing or inconsistent identifiers now propagate across algorithmic pipelines as well as financial ones[6].

By contrast, in most industries administrative overhead is modest:

- Retail/distribution: ~2–5% of net sales

[6] See (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, p31).

- Manufacturing: ~3–7%
- Professional services: 10–15% (because administration blurs into the product)
- OECD/EU cross-industry averages: 3–8% of turnover

In "normal" industries, then, €50 of administrative cost is justified on €1000 of revenue. By comparison, in the recorded music industry, achieving that same 5% efficiency requires delivering faultlessly some 200,000 streaming transactions. This is a **very tall order** for a sector dominated by micro-enterprises and small independents without dedicated IT or metadata teams.

The pressure for efficiency is not only present on the production side of the music business. In the **non-profit sector**, digitisation has profoundly transformed the workflows of archives, libraries, and heritage institutions as well. Streaming has reduced demand for physical collections, forcing libraries to reframe their role around digitisation, knowledge organisation, and community functions rather than lending CDs or scores. New spaces like creative studios and digital repositories are expected, but funding is limited, so efficiency is critical. At the same time, the vast amount of born-digital assets — and now the endless output of generative AI systems — creates a puzzle for archives that remains unsolved today.[7]

Metadata as provenance

In today's music ecosystem, almost every asset is born digital. A modern composer's score is produced in notation software; a performer's recording originates as a digital file; even printing, distribution, and promotion leave their own digital traces. From the very start, each musical work and each recording comes with a dense **digital fingerprint**.

As these works move through their lifecycle — composition, registration, performance, recording, distribution, preservation — they accumulate **provenance statements**: *"X composed this," "Y registered that," "Z archived this file."* Taken together, these traces form a chain of knowledge about the history of the work. Unlike in earlier centuries, this history is now almost continuously captured, though often fragmented or messy — the "shadows" that Karabinos has described.

[7] See for example the Katona József Library's adaptive strategies (Virág 2024). Archives, on the other hand, face a problem that instead of receiving records on paper, they are becoming gigantic data silos in the age of born-digital documents. They are being transformed into data through digitisation and born-digital records, face volumes too large for manual processing. This pressures traditional archival concepts such as provenance, original order, fixity, and authenticity (Colavizza et al. 2022).

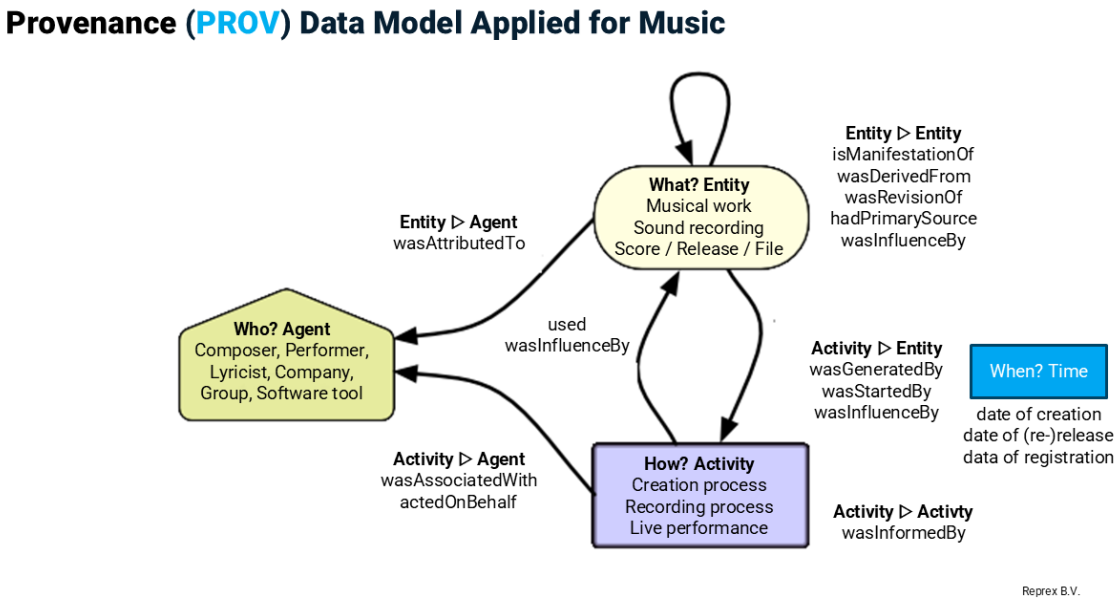


Figure 1.2: The PROV model helps us describe the lifecycle of music: who did what, when, and with what. A composer, performer, or software tool (agent) engages in an activity such as composing or recording, which results in a musical work or a sound recording (entity). Capturing these links over time makes provenance transparent, ensures correct attribution, and supports trustworthy data exchange across the music sector. Reuse: DOI: 10.6084/m9.figshare.30073210

Metadata is "data about data." But in practice, what counts as data or metadata is relative: a duration may be descriptive for one actor, identifying for another, and algorithmic input for a third. This distributed record of provenance resembles a **chain of statements**, some verifiable, some contradictory, some lost in the shadows. The challenge is not to build a single immutable blockchain, but to make the distributed record reliable, reusable, and interoperable.

As shown in Section 1.2, pilots like **SKCMDb** and **Unlabel** provide two complementary responses: preventive governance of metadata at creation (Chapter 2), and curative repair of legacy repertoires (Chapter 4; Chapter 3).

1.4 Potential solutions

The challenges described above call for coordinated responses that combine technical, organisational, regulatory, and governance measures. This policy brief develops them in detail across three thematic chapters — **curation** (Chapter 2), the **observatory** (Chapter 3), and **AI** (Chapter 4). Here we present an integrated overview of the solution families.

1. **Reducing redundancy and improving efficiency.**
   Shared registries and federated pipelines ensure that data is captured once and reused many times. The Slovak Comprehensive Music Database (SKCMDb, Chapter 2) demonstrates how libraries, rights societies, and archives can align their catalogues while retaining institutional autonomy.

2. **Reconciling attribution and privacy.**
   Metadata must balance GDPR requirements with author attribution duties under copyright law. Identifier pilots such as **PRS Nexus** and **Teosto ISNI** show preventive strategies at the point of creation, while SKCMDb offers curative repair of legacy repertoires.

3. **Pragmatic metadata alignment.**
   Instead of one universal ontology, modular and pattern-based approaches allow interoperability across domains. Initiatives such as **Polifonia**, **MusicBase**, and the **Unlabel** pipeline provide practical bridges between archival, library, and distribution metadata (Chapter 2, Chapter 3).

4. **Cross-sector observatories and data spaces.**
   The **Open Music Observatory** (Chapter 3) applies the European Interoperability Framework and 8-Star FAIR model to connect rights societies, libraries, archives, and statistical offices. Data sharing spaces provide governance, semantic, and technical layers that make public and private infrastructures interoperable.

5. **Curative strategies with AI.**
   Many repertoires remain invisible due to incomplete or inconsistent documentation. **Curative AI** (Chapter 4) can support enrichment, translation, duplicate detection, and plagiarism monitoring, extending the principles of Unlabel to broader repertoires.

6. **Bridges to public infrastructures.**
   Europe already invests in the **European Open Science Cloud (EOSC)**, the **European Collaborative Cloud for Cultural Heritage (ECCCH)**, and **Europeana**. These infrastructures should be aligned with the music sector to support both cultural preservation and competitive participation in digital markets.

7. **Shared AI services.**
   Micro-enterprises, NGOs, and CMOs cannot build in-house AI capacity. Cooperative AI utilities — reconciliation-as-a-service, metadata repair pipelines, watchlists for duplicates — can be pooled under shared governance (Chapter 4).

This integrated roadmap frames the more detailed analysis and recommendations in the chapters that follow.

CITF

Figure 1.3: You can download this report at <https://julkaisut.valtioneuvosto.fi/server/api/core/bitst e93a-48fb-92e8-93c61a999e48/content>

The CITF report defines a structured approach to future copyright infrastructures through three layers (foundational, semantic, technical). The solution pathways proposed in this Green Paper can be mapped onto these layers: identifier governance corresponds to the foundational layer; pragmatic ontology patterns map to the semantic layer; and federated pipelines align with the technical layer.

2 Fixing Music Data at the Source

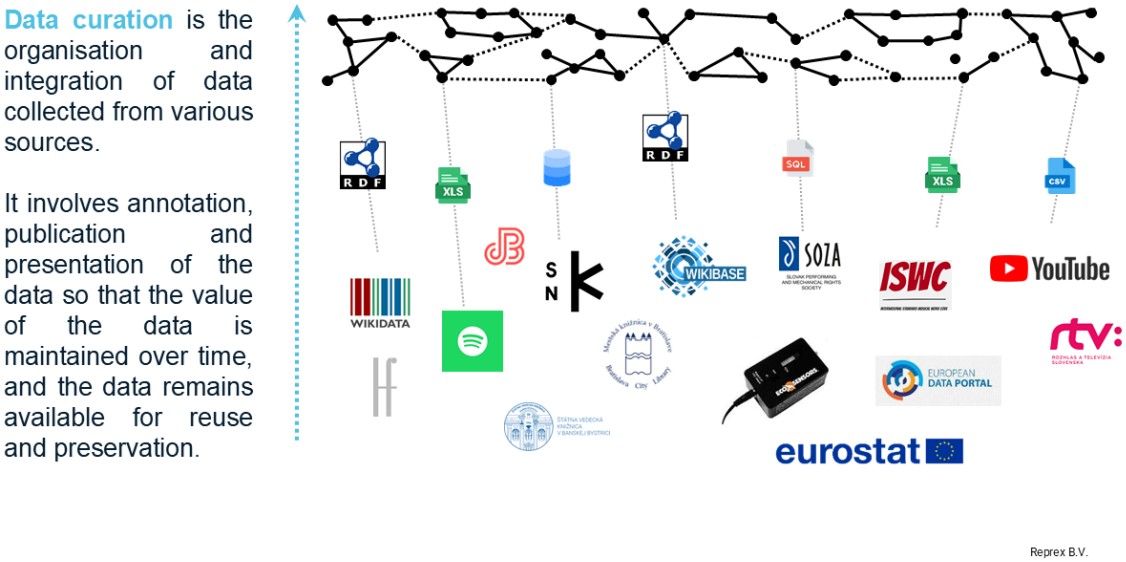


Figure 2.1: Curating data from multiple sources ensures that music information stays accurate, visible, and reusable over time. DOI: 10.6084/m9.figshare.30073888.v1 (click on image to reuse)

2.1 Discussion

2.1.1 Structural fragmentation of data and value flows

In the music ecosystem, data is not simply decentralised by design but *structurally scattered*. Rights metadata is maintained by hundreds of collective management organisations and publishers, while recordings and distribution data are spread across labels, distributors, and global platforms. Libraries and archives manage their own authority files, often linked only imperfectly to international standards such as ISNI, VIAF, or ISBN. Independent projects and community-driven infrastructures, such as *Wikidata* and Wikibase, add yet another layer of documentation. CITF identifies similar fragmentation across the wider copyright infrastructure. It notes that rights metadata, identifiers, and RMI are

distributed across many actors with differing mandates and data models, and that inconsistent identifier governance contributes to systemic opacity and recurring reconciliation costs [Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula (2025), pp10–18; p23].

This fragmentation is not an anomaly but the normal condition of the sector: tens of thousands of micro-enterprises and NGOs in Europe each manage slivers of data about works, recordings, or performances. As the *Feasibility Study for a European Music Observatory* underlined, *"the fragmented, scarce and poorly harmonised nature of the data collection landscape in the field of music has led to calls … for a European Music Observatory"* (Commission et al. 2020, p9). Likewise, the *Music Ecosystem 2025* study frames the sector as an ecosystem, where knowledge and value are distributed across many small actors, each with partial perspectives (Music Moves Europe 2024, pp6–7).

The institutional anchoring of a future *European Music Observatory is indeed a critical question.* In our own feasibility planning we reviewed approximately 80 functional and discontinued data observatories, understood here as permanent institutions for ongoing data collection and dissemination. The majority in Europe were initiated by the European Commission and maintained under various public–private partnership (PPP) formats, rather than as heavy agencies or autonomous bodies.

In this sense, Europeana offers a useful analogy: it coordinates metadata and access across hundreds of institutions without requiring the scale or mandate of entities such as the EUIPO or the European Audiovisual Observatory. In our interim report deliverable we suggested a similar creation path like that of the Europeana Foundation and its various layers of stakeholders (Antal 2024c).

From this perspective, we believe the Observatory should follow the lighter, federated PPP model: anchored by the Commission to ensure continuity and legitimacy, but implemented through a distributed network of partners across the public, private, and research domains. This strikes a balance between stability and flexibility, while staying true to the cooperative, federated spirit that underpins our proposal. The CITF report reaches a similar conclusion: copyright data cannot be centralised at European scale and must instead be organised through layered, federated arrangements where national libraries, rights organisations, and cultural institutions each retain their roles while interoperating through open standards[1].

Recognising this scattered landscape is essential. It explains why reconciliation overheads are high, why identifier coverage is incomplete, and why "capture once, reuse many" pipelines are necessary. It also provides the foundation for the next chapter: explaining why *attempts at centralisation are futile* in such an ecosystem, and why sustainable solutions must build on federation and interoperability.

Yet fragmentation is not only institutional — it is also economic. Classic value-chain analyses describe three main income streams — live performance, publishing, and recordings

[1] See (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, pp13–18).

— that still structure industry practice.[2] Digital distribution has blurred these categories without unifying the underlying infrastructures. Each handover in the lifecycle — authoring, performing, recording, distributing, streaming — generates both a financial flow and a data event. Business flows are continuous, but metadata flows are siloed. ISWCs do not connect seamlessly with ISRCs; ISRCs are rarely linked to ISNIs or VIAF authority files. The result is redundancy, inconsistency, and costly reconciliation work.

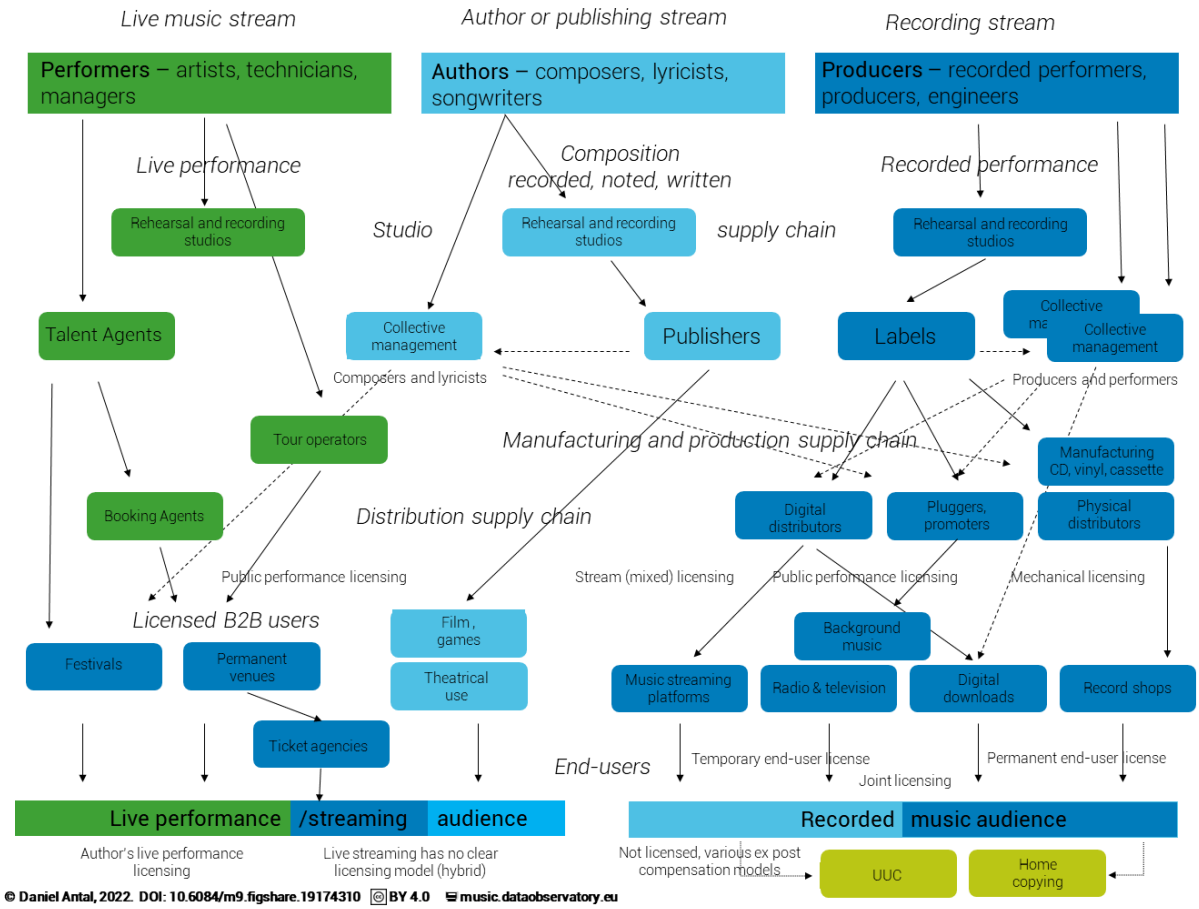


Figure 2.2: Adoption of the value chain model of the European music ecosystem in the CEEMID report

To address these challenges, we have adopted the value chain model of the European music ecosystem[3]. This approach is especially useful for designing data collection that measures

[2]This value-chain framing originates in Hull's *The Music Business and Recording Industry* (Hull et al. 2011) and Leurdijk et al.'s *Statistical, Ecosystem and Competitiveness Analysis of the Media and Content Industries* (Adnra Leurdijk and Ottilie 2012), and was adapted in subsequent CEEMID reports Antal (2020). The CEEMID work was recognised as a best practice in the *Feasibility Study for a European Music Observatory* (Commission et al. 2020), which highlighted its role in linking fragmented data sources into a coherent economic analysis.

[3]For the standard American/global analytical breakup of the music industry is described in (Hull et al. 2011), its European adaptation in (Andra Leurdijk and Ottilie 2012); our more detailed Central European breakup and the figure can is described in (Antal 2020, 2021), for the reuse of the figure please refer to (Antal 2022).

cash flows, gross value added, and zero-price uses of music. It highlights both typical price points (e.g. averages or medians) and the interlocking metadata flows that accompany transactions. For policymakers, the model provides a way to trace how consumption — such as a consumer buying a recording through a shop, distributor, and label — translates into revenues for performers and composers. For data governance, it illustrates why capturing the metadata trail of cash flows is essential not only for valuation and cultural statistics but also for building an audit trail for fair remuneration. In the context of this policy brief, the value chain perspective therefore complements the current ecosystem analysis by clarifying which agents must be accounted for in conceptual models of data interoperability.

2.1.2 Cost barriers in documentation and claims

For small publishers, labels, and self-publishing artists, the economics of documentation create a vicious circle. Most European repertoire is released by micro-enterprises that cannot afford dedicated staff for accounting or metadata. They save costs by using spreadsheets or freelance accountants, but this is efficient only in total terms — on a per-unit basis, the costs of documentation and claims are very high. Poor metadata then leads to poor discoverability on platforms, which in turn depresses revenues and leaves even less money for proper documentation.

Capital investments (CAPEX) present the same dilemma. Enterprise IT systems or royalty accounting platforms may be cost-effective for catalogues with millions of assets, but are unsustainable for catalogues of a few thousand. As a result, many small actors are locked into obsolete systems that are costly to maintain but too expensive to replace.

This structural imbalance means that metadata costs are proportionally higher for small entities than for large ones. Without a way to share infrastructure or reduce per-unit costs, small rightsholders remain stuck: they cannot spend more on documentation and claims than their total royalty income allows, yet under-documentation ensures that much of their income is never collected.

These cost barriers are not isolated bookkeeping problems — they are structural features of music data curation. How a data sharing space can provide scale effects and relieve these constraints is discussed in Section 3.2.2.

2.1.3 Why one grand collection model will not work

Every actor in music — a library, an archive, a label, or a rights society — has its own way of defining what counts as music, what is a sound recording, how to collect such things, and what belongs in a "collection." These logics are shaped by their missions, legal obligations, and incentives. A library may collect under a national deposit law, a collective management organisation must register what its members submit, and a distributor includes whatever its clients release. None of these logics are wrong, but they are different. This is why attempts to force everything into one universal collection model have failed.

In abstract terms, there is no single "conceptualisation" of the world that can fit a rights management organisation, a library, and a music archive equally well. On a very abstract level, the same lesson was drawn in mathematics and philosophy: Gödel showed that no formal system can capture all truths within itself, and Quine argued that reference is always relative to a conceptual scheme. In computer and information science, we know this as the impossibility of a universal ontology that could serve all databases.[4] These limits are well understood, but recognising them is not an excuse for inaction. It means we should work pragmatically: accept that multiple logics exist, and focus on making them interoperable where possible.

i Why collections differ in databases

- **Libraries** collect under *legal deposit rules*: every book or score published in a country must be included, regardless of popularity.
- **Archives** follow *provenance*: they keep what an organisation or individual produced, not necessarily what is "important."
- **Collective management organisations (CMOs)** must register *only what their members submit* — the collection reflects contracts and repertoire, not cultural completeness.
- **Distributors** take what their clients release: the "collection" is shaped by market demand and contracts.

Each of these logics is valid, but none can be reduced to the others. This is why a single "grand ontology" for all collections is not achievable. The pragmatic task is to connect them through lightweight, modular patterns that allow data to flow across boundaries while respecting institutional differences.

2.1.4 Legacy metadata

The European Parliament has emphasised that accurate and standardised metadata is essential for ensuring fair remuneration and proper attribution in the music streaming market. It calls for identifiers such as ISWC, ISRC, ISNI, IPI, and IPN to be allocated

[4] As information science shows, a *collection* is not a mathematical set but a socially and institutionally constructed grouping, shaped by curatorial or organisational logics. Attempts to create one "mega-ontology" for music metadata have consistently failed, because the sector is too heterogeneous — collective management organisations, libraries, archives, platforms, and distributors operate under different standards and governance models. At a more philosophical level, Quine reminds us that any ontology is relative to its conceptual scheme, and there is no absolute description of the world that can serve all purposes equally ((Quine 1968)). Gödel's incompleteness results, likewise, show the inherent limits of formal systems, underscoring why computer science and database theory recognise that no single universal ontology can capture all possible cases. The CITF report likewise rejects any attempt to impose a single universal schema. Instead, it argues for a semantic interoperability layer that allows heterogeneous systems to exchange meaning without erasing institutional differences[^citf-patterns].

at the moment of creation, and warns that the flood of AI-generated tracks will worsen discoverability and revenue imbalances if metadata remains incomplete or inconsistent.[5]

In practice, achieving this goal has proven very difficult. The registers that underpin music metadata are privately governed, require continuous investment, and cannot simply be rebuilt from scratch. Hundreds of millions of assets are already circulating, and billions of transactions are handled annually on the basis of this legacy infrastructure. Even the term *metadata* is ambiguous: in libraries and IT it means descriptive information (title, genre, provenance), but in the music industry it usually refers narrowly to administrative identifiers that drive royalty distribution. This gap in terminology adds to confusion and misplaced expectations.

i Forward-looking identifier pilots: PRS Nexus and Teosto ISNI

Two recent initiatives show how the industry is moving towards *better identifier coverage at source*:

- **PRS for Music – Nexus.** A new portal linking works (ISWC) and recordings (ISRC) at the moment of release. It already covers nearly 3 million works and offers APIs for rights-holders and DSPs (PRS for Music 2023; World Intellectual Property Organization (WIPO) 2023). By embedding ISWC allocation into distribution workflows, Nexus aims to accelerate royalty payments and reduce reconciliation delays that often last months or years.
- **Teosto – ISNI for authors.** The Finnish CMO Teosto now assigns ISNIs to its members, giving authors and composers persistent identifiers that interlink with VIAF, ORCID, and Wikidata (Teosto 2024). This connects music rights data with library and research infrastructures and strengthens international interoperability.

These projects simplify metadata at the point of creation and release, aligning with persistent identifier strategies in the research sector (Cruz and Tatum 2021). But they mainly address *future repertoire*. The much larger challenge lies in the **hundreds of millions of legacy assets** already circulating without complete identifier links — a problem that requires complementary solutions, discussed later in this chapter.

Together, ISRC (recordings), ISWC (works), and ISMN (printed music) form the backbone of music identification. In theory they provide global coverage, but in practice they remain fragmented: many recordings never receive identifiers, links between identifiers are often missing, and uptake is uneven across registries. This fragility makes the European Parliament's ambitions difficult to realise without new layers of interoperability, observability, and shared responsibility. The sheer growth in repertoire makes this gap impossible to close with manual workflows: by 2024, more music was released in a single day than in the entire year of 1989 (Abing 2024)[6]. This scale of legacy under-documentation cannot

[5] European Parliament resolution of 17 January 2024 on cultural diversity and the conditions for authors in the European music streaming market, recital 32 (European Parliament 2024).

[6] The *International Standard Recording Code* (ISRC) was introduced in 1986 as a 12-character identi-

realistically be resolved with manual workflows alone — it points directly to the need for **curative AI** approaches, which we return to in Section 4.2.

Although metadata repair is indispensable, metadata is never neutral. Without corrected identifiers, reconciled names, and enriched annotations, works remain invisible in royalty and discovery systems. However, just as heritage data spaces show how repairing metadata can restore visibility while also reinforcing institutional logics, in music ecosystems the same repair practices can unexpectedly increase exposure to generative AI. By making works more legible to agentic applications, enriched metadata improves attribution but also sharpens the ability of AI systems to imitate and substitute. This paradox is most acute for small-scale repertoires and independent artists, whose economic position mirrors the epistemic vulnerability of minority heritage collections.

CITF emphasises this duality. It notes that richer attribution and provenance metadata are essential for rights enforcement, yet these same signals can enhance the capacity of AI systems to generate derivative content. This makes trustworthy provenance, audit trails, and transparent RMI all the more important in AI-era infrastructures[7].

i Case Study: Metadata Repair — Heritage and Repertoire

Repairing heritage metadata

- In the *Finno-Ugric Data Sharing Space* we worked with the *Latvian Archive of Folklore* and regional museums to repair and enrich metadata around Livonian, Latvian, and Hungarian folk music.
- In Hungary, together with the *House of Music*, we began repairing the lost documentation of recordings suppressed under Communist censorship. Here, repair is not only a matter of accuracy but also of restitution: without corrected metadata, these works remain locked behind outdated copyright classifications long after the state label monopoly has ended.
- Original records in both contexts were shallow, monolingual, and shaped by institutional or censored taxonomies. By reconciling names, places, languages, and cultural terms, we enabled works to be rediscovered across Wikidata and Wikipedia.

fier for recordings (ISO 3901) and is managed operationally by IFPI (ISO 2019a; International ISRC Registration Authority 2021). Persistent problems include retroactive assignment, inconsistent embedding, and weak interoperability with ISWC (Paskin 2006, p4). The *International Standard Musical Work Code* (ISWC) identifies compositions and lyrics (ISO 15707), managed by CISAC through the ISWC Agency (ISO 2022). Challenges include duplicate codes, mismatches with ISRC, and uneven adoption by CMOs (Paskin 2006, p7). The *International Standard Music Number* (ISMN, ISO 10957) identifies printed music publications (ISO 2021). It provides a bridge between bibliographic and rights-management practices, but remains underused in digital workflows. CITF also highlights the fragility of legacy rights metadata, noting that many identifiers lack persistent links, that national and sectoral registries follow incompatible governance models, and that historical gaps in RMI complicate both attribution and AI-related provenance. It stresses the need for repairable metadata chains capable of supporting lifecycle analysis and AI-era compliance [Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula (2025), pp12–18; pp28–33].

[7] See [Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula (2025), pp20; p31; pp101–102].

- These are extreme cases of damaged metadata (through censorship, Soviet-type copyright, or minority language non-standardisation). Yet similar problems affect the long tail of European music heritage and today's independent or self-releasing artists.
- As our forthcoming academic paper shows, this is not a neutral "clean-up": choices about vocabularies and identifiers determine what communities can see of themselves. Repair here means **cultural repair** — restoring epistemic visibility to communities, legal heirs, and cultural stewards.

Repairing repertoire metadata

- Through the *Unlabel* prototype, we apply similar practices to contemporary self-released music: enriching works with ISRC/ISWC codes, multilingual annotations, and library-standard metadata.
- This makes previously "invisible" tracks legible to streaming platforms and collection societies, improving discoverability and royalty flows.
- Again, repair is not neutral: the way identifiers and categories are assigned shapes how artists' works are found, monetised, or sidelined.

The paradox

- These cases illustrate that **metadata is never neutral**. Repair empowers artists and communities, but it also encodes assumptions and makes works more legible to agentic applications.
- In heritage, institutional schemas may flatten local epistemologies; in the market, generative AI may exploit enriched metadata to imitate and substitute — a problem we will discuss in Chapter 4.
- In both contexts, **metadata repair empowers and exposes** — visibility and risk are two sides of the same process, which makes metadata governance a **policy concern**, not a purely technical one.

Our approach

- Our solution is to use decentralised systems like Wikidata and Wikibase together with strong ontological patterns.
- Heavy-weight ontologies take up to a decade to develop, may introduce new biases through the non-neutral nature of metadata, and by the time they are created, they may not address new challenges — for example, providing guardrails against negative outcomes of agentic or generative AI.
- As with the infrastructure in Chapter 3, we aim for decentralisation already at the metadata-definition level. An **Open Music Observatory** will allow metadata to be managed through flexible, open processes that create definitions and establish equivalences to existing standards.

2.1.5 Named-entity resolution, attribution, and privacy

Attribution is not optional in music: the names of authors, performers, and producers are structurally necessary for copyright, royalties, and cultural record-keeping. Yet under GDPR, these names count as personal data, creating a contradiction at the very foundations of metadata curation. What is mandatory under copyright law becomes a liability under data protection law. In practice, private actors face repeated balancing tests,

inconsistent interpretations, and the risk of complaints even when attribution is legally required.

The CITF report explicitly identifies this contradiction. It notes that names and attribution data constitute rights-management information protected under Article 7 of the InfoSoc Directive, yet they are also personal data under GDPR. CITF therefore calls for trustworthy, machine-readable RMI governance that distinguishes public-interest attribution data from restricted personal information and supports layered access models[8].

This contradiction drives up costs and discourages investment in better metadata. Small publishers and self-releasing artists already face disproportionately high OPEX (documentation, bookkeeping) and CAPEX (IT systems). Without affordable, legally secure ways to resolve named entities, their works perform badly on platforms and royalties are lost.

Policy communities in Europe recognise these issues. The **Big Data Value Association (BDVA)** has long argued that trust frameworks and governance pillars are essential for data sharing, while the **Federation Working Group** stresses that federation — not centralisation — is the only realistic model for connecting Europe's fragmented data ecosystems (Big Data Value Association 2019; BDVA/DAIRO 2023; BDVA/DAIRO Federation Working Group 2023). These principles apply equally in music. But given the sector's extreme fragmentation and micro-enterprise structure, implementing them here is especially difficult.

How these structural problems can be addressed at systemic level is the subject of Section 3.1.3, where we show how data sharing spaces provide a way forward.

[8] See [Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula (2025), pp12–18; p84].

2.2 Policy proposals

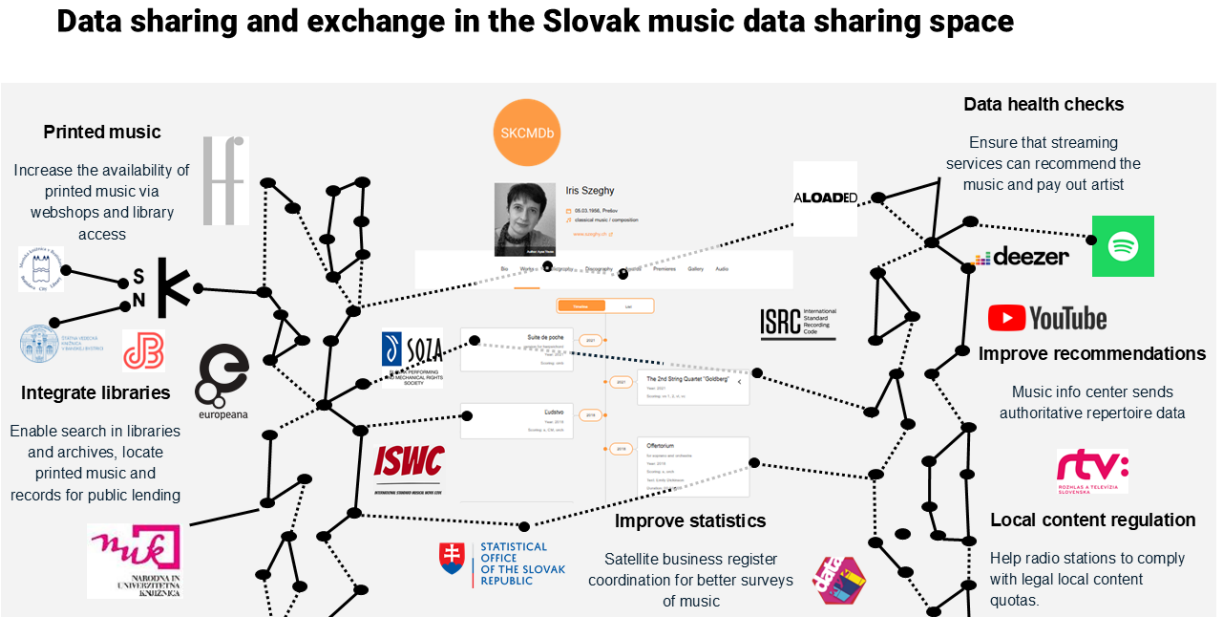


Figure 2.3: Explanation

2.2.1 Reducing redundancy

The European Parliament has rightly highlighted that fragmented and unreliable metadata remains a major obstacle in the music sector. We agree with this diagnosis, but stress that the root cause lies partly in the need for backward compatibility with hundreds of millions of legacy assets, and in the costly redundancy of today's practices: the same information must be repeatedly entered into separate systems such as ISNI, ISWC, ISRC, VIAF, or local authority files. This duplication creates errors, increases costs, and discourages accurate registration.

CITF frames this redundancy as a foundational problem of copyright infrastructure. Its proposed foundational layer centres on open, authoritative identifiers for agents and assets, issued by trusted institutions and supported by interoperable mappings[9].

Our policy solution is to support **redundancy-free registration** by aligning the workflows of those who already maintain authoritative data. Instead of duplicating efforts, registration steps can be coordinated once and reused many times. We demonstrate this approach with our *Open Music Registers* pilot: a federated infrastructure that interconnects persistent identifiers (ISWC, ISRC, ISNI, VIAF) and, where relevant, links them

[9]Aligning workflows around these identifiers directly addresses the structural issues described in the report [Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula (2025), pp23–28; pp85–87].

to business and statistical identifiers (e.g. OpenCorporates, NACE, ISCO). This allows music creators and organisations to benefit from smoother workflows, while downstream users gain more reliable data for royalty distribution, cultural visibility, and AI-driven discovery.

The *Open Music Registers* deliberately avoid centralisation. Each registrar — collective management organisations, libraries, archives, or statistical offices — retains ownership of its data but contributes to a shared semantic framework.[10] By connecting rather than merging registers, redundancy is reduced while subsidiarity, accountability, and trust are safeguarded across public and private actors. This distributed model directly answers European Parliament's call for metadata systems that are reliable, inclusive, and supportive of creators.[11]

2.2.2 Reconciling attribution and privacy

The problem of reconciling copyright attribution with GDPR obligations cannot be solved by ignoring either side: both are binding legal requirements. Our approach, tested in the *Slovak Comprehensive Music Database* (SkCMDb), shows that progress is possible through layered governance and careful balancing. Academic institutions and libraries, with their cultural and research mandates, can lawfully handle personal data under derogations for public-interest processing. Collective management organisations (CMOs) and private actors, by contrast, must rely on legitimate interest tests, supported by transparent documentation, notification to rightsholders, and opt-out mechanisms where possible.

i Interoperability is a means, not a goal

Our Slovak pilot, the **Slovak Comprehensive Music Database (SKCMDb)**, links libraries, rights management, streaming services, and the statistical office.
This is not "interoperability for its own sake." Ontologies and crosswalks are valuable only insofar as they enable **better services**:

- **For audiences**: making music findable and accessible across cultural and commercial platforms.
- **For rightsholders**: ensuring that attribution, identifiers, and royalty flows are correct.

[10] Technically, this corresponds to a *provenance-oriented modelling* approach such as the W3C **PROV-O** standard (W3C 2013b, 2013a), which connects actors, activities, and entities in chains of attribution ("a composer authors a work, a performer interprets it, a producer records it…"). These chains can be expressed in the layered terms of the *European Interoperability Framework* (EIF), ensuring legal, organisational, semantic, and technical interoperability (Commission and Digital Services 2017).

[11] The *Data Spaces Support Centre (DSSC)* Blueprint v2.0 underlines that identifiers and rulebooks are the foundation of any common European data space (Data Spaces Support Centre 2025b). In the music sector, however, attribution identifiers themselves are caught in the GDPR contradiction (see Section 2.1.5), which underscores the importance of redundancy-free but legally robust registration practices.

- **For policymakers**: providing reliable data to support cultural policy and to measure the music economy.

In short, interoperability at the data level is the condition for **usable services** at the societal level.
The Slovak Memorandum of Understanding shows how attribution and data protection can be balanced in practice.
- **Names of authors, performers, and producers** are treated as public-interest information necessary for copyright and royalty flows, justified under legitimate interest.
- **Sensitive fields** (e.g., addresses, nationality, pseudonyms) are excluded from public layers and restricted to controlled-access tiers.
- **Governance** is distributed across CMOs, libraries, and archives, ensuring subsidiarity and trust.
This layered compliance model demonstrates that copyright attribution and GDPR obligations can coexist — and offers a template for other Member States and for the European-level Open Music Observatory. These conclusions are consistent with the CITF report, which recommends separating public-interest attribution data from sensitive fields and managing both through tiered access and provenance-tracked RMI.

Balancing tests play a central role: each dataset is audited, divided into *public* and *non-public* categories, and then assessed again for personal vs. non-personal data. Public information such as names of authors, performers, and work titles—already widely available in catalogues and concert programmes—can justifiably be shared under legitimate interest, especially when linked to rights management purposes. Sensitive data (e.g. addresses, nationality, pseudonyms) require stricter access tiers and are only made available to selected stakeholders under contractual safeguards.

This layered compliance model does not eliminate GDPR challenges, but it creates a robust defence: it demonstrates that the legitimate interest in accurate attribution and royalty distribution outweighs the minimal risks of publishing already public information. In practice, this means rights metadata can circulate across the ecosystem while privacy-sensitive data are contained. Building such workflows into federated observatories and data spaces allows the music sector to comply with data protection rules without undermining attribution, and provides a model for European-scale solutions.

More broadly, these governance practices are supported by existing provisions in EU copyright and data legislation that already give metadata a central role. **Rights Management Information (RMI)** is explicitly protected under Article 7 of the *InfoSoc Directive* (2001/29/EC), making the removal or alteration of attribution data unlawful. The *CRM Directive* (2014/26/EU) obliges collective management organisations to maintain accurate and transparent repertoire and membership data. Under the *CDSM Directive* (2019/790/EU), Article 17(4)(b) requires platforms to act expeditiously on notices where metadata enable rightholders to identify and claim their works, while Article 4(3) uses metadata as the operational basis for text and data mining opt-outs. Beyond copyright, the *Data Governance Act* (2022/868), the *Data Act* (2023/2854), and the *Open Data Di-*

rective (2019/1024) provide the horizontal framework for treating music metadata as part of Europe's emerging common data spaces.[12]

2.2.3 Pragmatic metadata alignment

Attempts to build one comprehensive, harmonised schema for music metadata have repeatedly failed. The sector is too diverse: collective management organisations, libraries, archives, distributors, and platforms all operate with different standards and governance models. Trying to impose a single "grand schema" has proven brittle, costly, and unrealistic.

A more workable solution is *modular alignment*. Instead of a single heavy ontology, small reusable building blocks can be combined to describe recurring patterns — for example, how people, works, recordings, and performances are related. This approach allows interoperability to grow step by step, without forcing any actor to abandon its systems.[13]

It also helps to separate two complementary tasks. On the one hand, we need **conceptual scaffolding** that lets different databases describe similar structures in comparable ways. On the other, we need **identifier reconciliation** to make sure that the same person, work, or recording can be linked across different registers. Neither of these tasks is sufficient on its own: they must work together if metadata is to remain reliable at scale.[14]

Other domains show how this can be done. Research infrastructures have reconciled **ORCID** with **VIAF** authority files, and libraries have mapped **DataCite** metadata to **Dublin Core**. Both examples show how two different standards can be aligned systematically while keeping their distinct scopes.[15]

[12] See *Policy Brief 1: Music Metadata Mainstreaming and EU Law* (Senftleben et al. 2024) (Deliverable D5.6, OpenMusE project). That brief analyses how these instruments can be mobilised to improve the reliability and circulation of music metadata. The present Green Paper complements this by showing how federated observatories and interoperability strategies can operationalise these obligations in practice.

[13] On ontology design patterns and modular approaches, see (Gangemi 2005; Blomqvist, Hammar, and Presutti 2016; Carriero et al. 2021). The **Polifonia project** applied these methods at European scale (Berardinis et al. 2023), aligning with MusicBrainz and the **ChoCo** knowledge graph (Albanese et al. 2023). While Polifonia did not focus on rights metadata, it provides a strong foundation for connecting musicological knowledge with industry identifiers.

[14] This distinction between ontology modelling and identifier reconciliation clarifies why both layers are necessary. Ontology patterns provide conceptual scaffolding (e.g. work–recording–performance), while identifier reconciliation ensures that an author in ISNI is the same as a VIAF authority record or a performer in MusicBrainz.

[15] For ORCID–VIAF reconciliation via OpenRefine, see (OpenRefine Community 2021; Jegan et al. 2023). For systematic mappings between DataCite and Dublin Core, see (DataCite 2021).

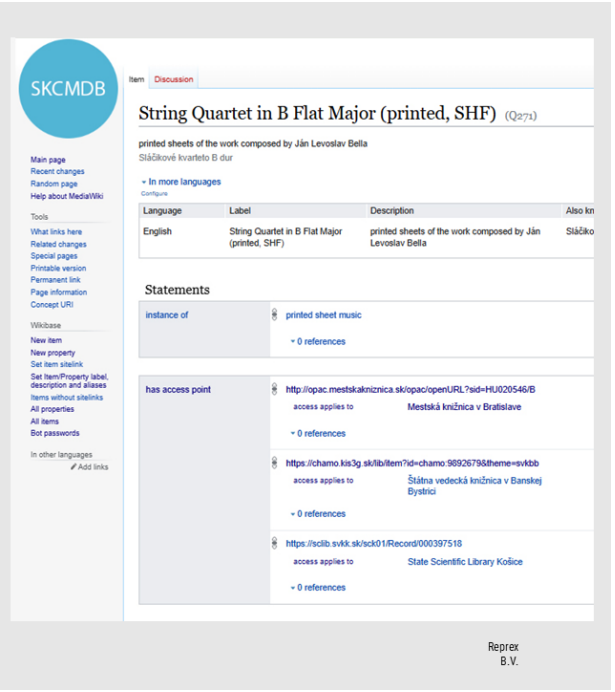


Figure 2.4: Pragmatic metadata alignment relies on modular patterns, not "giga-schemas." The example shown here from our Wikibase pilot encodes roles, events, and provenance using reusable ontology design patterns. This allowed identifiers from rights management (ISWC, ISRC) to be reconciled with library authorities (ISNI, VIAF), proving that interoperability can be achieved incrementally without forcing any actor to abandon its systems. DOI: [10.6084/m9.figshare.30075379.v1](https://doi.org/10.6084/m9.figshare.30075379.v1)

Music metadata needs the same periodic reconciliation. Rights identifiers such as ISRC, ISWC, and ISMN were designed separately and drift apart if not actively maintained. The same applies to personal and organisational identifiers such as ISNI, VIAF, and IPI. Without active cross-checking, records fragment, causing duplication and inconsistency.[16]

In our pilots, this modular alignment has already been tested. The Slovak Comprehensive Music Database reconciled rights identifiers with library authorities without schema unification. *MusicBase* used Wikibase to encode roles, events, and provenance in a way that let corrections propagate across systems. The *Unlabel* workflow streamlined metadata capture for self-releasing artists and libraries, allowing once-only documentation to be reused across distribution and preservation. These cases extend our proposal for *Open Music Registers*, which argued for federated, redundancy-free metadata workflows, into the broader governance framework of this Green Paper, similarly to the CITF three-layer interoperability model[17].

[16] On the divergence of identifiers if not maintained, see (Paskin 2006).

[17] Our modular ontology patterns correspond to CITF's semantic layer, while identifier reconciliation aligns with the foundational layer, and federated registries reflect the technical layer. This mapping demonstrates that music-sector practices can evolve within the broader European framework envisioned by CITF (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča,

Finally, this approach is consistent with work in the heritage sector. The **Heritage Digital Twin Ontology (HDTO)**, developed within the European Cultural Heritage Cloud, uses the same principles of modularity and federation to describe tangible and intangible assets. Where HDTO provides a semantic framework for heritage "digital twins," the Open Music Observatory extends the same logic to music. Both models show how cultural and rights metadata can integrate with wider European data spaces while preserving subsidiarity and institutional diversity.[18]

Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, pp23–33).

[18]The **ECHOES Heritage Digital Twin Ontology (HDTO)** builds on CIDOC CRM extensions to model tangible and intangible heritage with space–time–cultural identity (ECHOES Ontology Task Force 2025).

3 Open Music Observatory: Building a Shared Music Data Space

Open Music Observatory

Our ambition with the development of the **Open Music Observatory** is to provide the technological basis and a practical roadmap for creating a European Music Observatory in a bottom-up, decentralised way. Instead of waiting for a grand, central agreement, any data owners or collectors who satisfy quality and cooperation rules can add their data. Once the Observatory reaches sufficient maturity, its long-term institutional form can be decided.

The Open Music Observatory is a cornerstone task of the OpenMusE project (running until 31 December 2025), delivering data collection, processing, dissemination, and innovative services. It is a digital service provider for the music industry, aligned with the *European Interoperability Framework*, and introduces a unique governance model that adapts best practices from the EU and other sectors.

Transparency note: Following the principles of **Open Policy Analysis**, we have made all key deliverables (including versions 0.99, 1.01, and 1.1 of the *Open Music Observatory* document) publicly accessible to foster broad stakeholder engagement and to provide a clear audit trail. These versions are available at https://zenodo.org/records/11564114, while version 1.0 remains internal and was shared only with OpenMusE evaluators. Minor edits, as well as access to the standardised folders, figures, and bibliographies, can be found at https://github.com/dataobservatory-eu/open-music-observatory. You can access the documentation in PDF, EPUB, and docx [in case you would like to give us comments] here.

Citation note: If you refer to the specification of the *Open Music Observatory* in correspondence, publications, or blog posts, please cite the latest **versioned DOI** available on Zenodo and, if applicable, include the date of access when referring to material on our GitHub repository.[1]

The *Music Ecosystem 2025* report already emphasised that the music sector should be understood as a distributed ecosystem where value and knowledge are held by many small actors (Music Moves Europe 2024, pp6–7). This perspective reinforces why centralised repositories fail and why federated observatories, built on cooperation and interoperability, are more realistic.

[1] Always use the latest versioned DOI when citing this *Open Music Observatory* technical report, available via Zenodo. If you rely on supporting material hosted in the GitHub repository, please add the date of access in your reference.

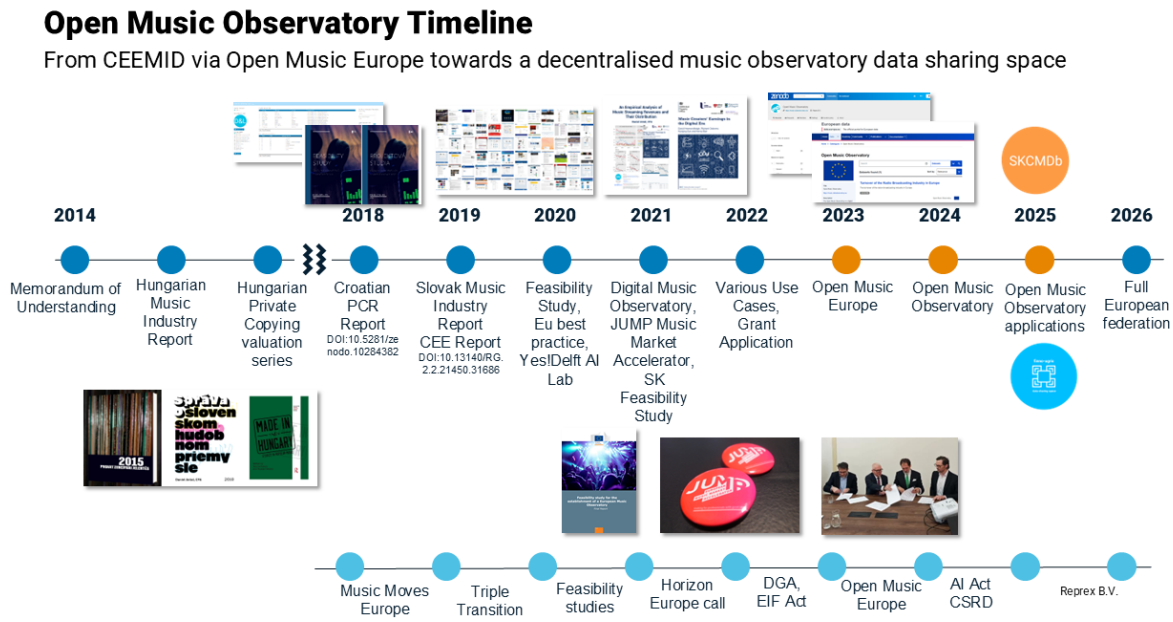


Figure 3.1: Over the past decade, feasibility studies, national reports, and EU pilot projects have laid the foundation for the Open Music Observatory. The roadmap (2014–2026) shows a gradual build-up: from local experiments, through cross-border collaborations, to a European-wide federation aligned with cultural data spaces and interoperability frameworks. This trajectory underlines the Observatory's pragmatic, step-by-step approach to scaling music data infrastructure. DOI: [10.6084/m9.figshare.30073291.v1](https://doi.org/10.6084/m9.figshare.30073291.v1)

3.1 Discussion

> **Caution**
>
> **This will be removed consultation** - EMO feasibility on scarcity/fragmentation and the need for regular, comparable data; EU dataspace thinking (EIF, FAIR); *Music Ecosystem 2025* on systemic view. - Industry positions on centralisation vs. federation; CMOs' reliance on shared infra (e.g., Mint); heritage sector's openness requirements.

3.1.1 Why centralisation is a futile model

Calls for a *centralised European database* of music often reappear in policy debates, but in practice such proposals are neither realistic nor aligned with current EU strategies. Centralisation assumes that highly diverse data sources can be harmonised within a single repository. In an ecosystem where knowledge is held by tens of thousands of micro-

enterprises, NGOs, collective management organisations, and heritage institutions — each operating under distinct legal frameworks — this assumption is untenable.

CITF reaches the same conclusion. It notes that copyright data is inherently distributed across many custodians with incompatible mandates and governance models, and that no single centralised registry can meet the legal, operational, and semantic requirements of modern copyright workflows. Instead, it argues that future-proof infrastructures must rely on federated, lifecycle-aware registries capable of exchanging trustworthy provenance and rights metadata while preserving institutional autonomy.

i Lessons from the Global Repertoire Database

Between 2008 and 2014, European and global stakeholders pursued the **Global Repertoire Database (GRD)** as a solution to the chronic fragmentation of musical works data. Backed by collective management organisations (CMOs), major publishers, and digital service providers, the GRD aimed to establish a **single, authoritative global database** of musical works and rightsholders. Its promise was that licensees—especially online platforms—could obtain reliable rights information from one source, reducing duplication and disputes.
However, the GRD ultimately **collapsed before launch**, despite several years of investment and the establishment of a London-based operating company. A similar project, the *International Music Registry* project, which was backed by the World Intellectual Property Organization, ended with similar results[2].
Post-mortems identified several reasons: - **Governance conflicts**: disagreements between major publishers, CMOs, and other stakeholders over who would control and fund the database.

- **High costs and unclear incentives**: the project's projected maintenance costs exceeded what many participants—particularly smaller CMOs—were willing or able to sustain.
- **Asymmetries of power**: large publishers and CMOs were reluctant to share sensitive commercial data on equal terms with competitors.
- **Lack of trust**: concerns over who would "own" the data and how revenues would be redistributed undermined cooperation.

The failure of the GRD is now widely cited in policy and industry discussions as evidence of the limits of centralised, "single-database" solutions in the music sector. Similar initiatives even failed on national levels.
We can also add that centralisation, even if it was possible, would pose a new risk of creating monopolistic gatekeepers to the music ecosystem.
The predecessor of the Open Music Europe project, CEEMID, was based on the lessons of the following problems and on the insights of a decentralised, dataspace like approach (Antal 2020). In such *federated, interoperable approaches*—where data remains with its custodians but can be linked through shared identifiers, standards, and protocols—have proven more viable. CISAC's CIS-Net, Europeana in the heritage

field, and emerging European data space initiatives exemplify this more **distributed model of governance**.
CITF's analysis reinforces these lessons. It identifies governance opacity, unclear mandates, and incompatible identifier regimes as recurring causes of failure in large-scale copyright registries. It stresses that unless registries adopt transparent governance, open identifiers, and shared semantic profiles, centralised projects inevitably collapse under conflicting incentives[3].

EU infrastructure initiatives have already moved beyond this logic. Since the 2000s, projects such as *Europeana*, the *European Open Science Cloud (EOSC)*, the *European Collaborative Cloud for Cultural Heritage (ECCCH)*, and *DARIAH* have all adopted *federated architectures*, linking distributed collections through shared standards and interoperability frameworks rather than consolidating them into one database. The *Audiovisual Observatory*, established in 1993 as a centralised reporting body, represents an earlier institutional logic that is now being phased out in favour of federation.

The heritage sector, including music heritage, has consistently stressed the need for *open, federated models*. Libraries, archives, and museums use authority files and collaborative platforms (e.g. *VIAF*, *Wikidata*, *Wikibase*) to enable interoperability while preserving institutional autonomy. Commercial infrastructures do the same: the ISRC system, managed by IFPI, is inherently decentralised, while CISAC's *CIS-Net* gives access to rights data without centralising ownership. Even the Mint initiative, launched by CISAC and Armonia Online, shows how shared infrastructure can deliver economies of scale for identifier allocation and metadata management while avoiding dependence on a single repository.[4]

Even official governmental statistics, often seen as centralised, are in reality decentralised. The **ESSnet-Culture** project, coordinated under Eurostat, produced the first comprehensive framework for cultural statistics in 2012, adapted from the UNESCO model, and remains a *"basic reference"* for the field. More broadly, national statistical offices, labour force surveys, and administrative registers each collect partial data, which are harmonised at EU level for comparability. Increasingly, surveys and administrative datasets are complemented by flows from platforms, rights management organisations, and other industry actors. Indicators therefore emerge from **hybrid constellations of public and private data sources**, confirming that decentralisation is a structural feature of European evidence creation.[5]

[3]See for example Goldenfein and Hunter (n.d.); Milosic (2015).

[3]See (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, pp15–18).

[4]On heritage practices, see (Bianchini, Bargioni, and Pellizzari di San Girolamo 2021, p210) and (Sardo and Bianchini 2022, p297), which describe how VIAF, Wikidata, and Wikibase function as authority tools in libraries and archives. On identifiers, the *ISRC Handbook* (International ISRC Registration Authority 2021, p5) explains the decentralised structure of the ISRC system, while CISAC's *Mint Digital Services* (CISAC/SUISA/SESAC 2017) illustrates how federated allocation works in practice. Together, these examples show how distributed stewardship and shared standards underpin global metadata infrastructures.

[5]The *ESSnet-Culture* framework (Commission et al. 2020, p9) demonstrates how cultural statistics are built on national contributions harmonised at EU level, not on central databases. A Slovak pilot (Antal 2023) further illustrates how decentralisation works in practice, integrating public and private sources

3.1.2 Open Data Directive: right without means

The *Open Data Directive* grants a right of reuse for public-sector information and requires that certain "high-value datasets" be made freely available across Europe (*Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-Use of Public Sector Information* 2019). This includes cultural heritage institutions such as libraries, museums, and archives. However, the Directive stops short of providing the means to ensure that such data is actually usable.

Studies consistently show that open data often remains more of a promise than a reality. In practice, much open data is poorly documented, lacks common identifiers, and is released in unstandardised formats. While it may be free of charge or available at marginal cost, making it interoperable and trustworthy for cross-border use requires significant additional effort. The burden of curation, harmonisation, and enrichment falls on downstream users, which can be prohibitively expensive for smaller organisations. As the CEDAR project put it, *"Public authorities are only required to make existing data available, not to create new data or improve existing systems. This leads to significant disparities in usability and accessibility"* (Project 2023). A recent EU-wide usability study adds that *"many open data portals remain difficult to navigate, poorly documented, and inconsistent in their metadata quality, limiting actual reuse"* (Jachimczyk and Nowak 2024).
These structural weaknesses of open data provision set the stage for the Observatory's role in providing workflow playbooks and redundancy-free registration, discussed in Section 3.2.[6]

3.1.3 Why voluntary workarounds do not scale

The Slovak pilot shows that voluntary workarounds for attribution under GDPR are possible (see Section 2.1.5), but they do not scale. Even with strong communication and opt-in procedures, fewer than 1.3% of authors responded. Every new dataset requires fresh balancing tests, repeated notifications, and continued exposure to legal risk.

For observatories and data spaces, this is untenable. Interoperability requires clarity and legal certainty across borders and institutions. Without guidance from a Data Protection Authority or the European Commission, every national or sectoral initiative risks being

into coherent cultural indicators.

[6] Early modelling stressed the economic potential of open data but also identified major obstacles in practice: lack of availability, uneven quality, and poor usability (Carrara et al. 2015, p7; Huyer and van Knippenberg 2020, p14). Comparative studies show that simply granting a right to reuse rarely produces machine-actionable datasets. In complex domains like music, where attribution depends on precise identifiers, these shortcomings become particularly costly. Cross-sector reviews underline persistent fragmentation: heterogeneous formats and divergent practices across Member States (Buttow and Meijer 2024, p12); variability even in high-value geospatial datasets (Kević, Kuveždić Divjak, and Welle Donker 2023, p3); and sectoral case studies (e.g. mineral intelligence) repeatedly call for shared profiles beyond legal openness (Simoni, Aasly, and Schjøth 2021, p5). Additional evidence shows that preparing legacy administrative data for reuse requires cleansing and enrichment that impose real costs, even when the data are nominally "open" (EuroSDR 2021, p9; Schnurr 2021, p14; Nakos and Tsoulos 2022, p6).

challenged. The result is paralysis: public infrastructures cannot fully attribute works, and private actors refrain from sharing metadata for fear of liability.

In effect, Europe's music data infrastructures remain locked in uncertainty — unable to guarantee attribution, diversity monitoring, or local content compliance. This makes a purely local or voluntary approach insufficient. The solution must be systemic: a **federated data sharing space**, supported by common specifications and clear governance frameworks, so that attribution and interoperability can scale. How such systemic solutions can be embedded into the Observatory's conformance and legal levers is developed in Section 3.2. These unresolved attribution issues ultimately undermine not only observatories but also AI fairness and governance (see Section 4.1.3)[7].

3.1.4 Public infrastructures bypass music's real data flows

Europe has invested heavily in cultural and research data infrastructures such as *Europeana*, the European Collaborative Cloud for Cultural Heritage (ECCCH), and the European Open Science Cloud (EOSC). Yet these initiatives remain poorly aligned with how music metadata is generated and maintained in practice — mostly by private actors such as labels, distributors, and collective management organisations. Unlike archives, museums, or libraries, where digitisation was largely funded with public money, in music and film the bulk of digitisation has been carried out by industry. Public infrastructures therefore miss the systems where music's real data flows originate.

The *Europeana Data Model* (EDM) was designed for library holdings and is well suited to printed works, but it cannot capture the attribution needs of recorded music, which must identify at least three groups of rightsholders: authors, producers, and performers.[8] The *ECCCH report* likewise overlooked music entirely, focusing instead on monuments, archaeology, textiles, and museums.[9] Its first projects — such as **AUTOMATA**, **TEXTaiLES**, **HERITALISE**, and **ECHOES** — developed advanced tools for other heritage assets, but none addressed music directly. Our own attempts to include music datasets in ECHOES' cascading grants illustrate the problem: proposals were screened out early, despite the clear need for music representation.

CITF highlights the same structural blind spot. It observes that national libraries already curate large volumes of copyright-protected material and maintain authoritative identifiers, yet they remain largely decoupled from rights metadata workflows. This is exactly

[7] The Slovak pilot demonstrated that even with careful communication and GDPR balancing tests, participation was below 1.3%, showing the practical limits of voluntary attribution workarounds. Without EU-level guidance, every dataset requires fresh legal reasoning, making scale impossible. Comparable findings in other cultural domains underline the risk: voluntary consent-based models tend to collapse under low response rates and high compliance costs. See also discussions of attribution and AI fairness in (Commission et al. 2020) and (Music Moves Europe 2024).

[8] The EDM builds on DCTERMS, which works well for printed music but not for recordings. It fails to capture neighbouring rights such as those of producers and performers (Europeana 2017).

[9] *Ex–ante impact assessment on the European Collaborative Cloud for Cultural Heritage* (Commission et al. 2022). The first ECCCH pilots (AUTOMATA, TEXTaiLES, HERITALISE, ECHOES) focused on archaeology, textiles, and monuments, leaving music out.

the workflow we tested out in Slovakia, and are going to introduce in our Hungary replication. CITF therefore recommends treating national libraries and cultural institutions as copyright-infrastructure actors, not only heritage custodians, and integrating their registries into federated rights environments, which is exactly what we did in our Slovak national federated module, and what we aim to replicate in Hungary.

Other initiatives show the same bias. The **Polifonia** project created modular ontologies, but it was "blind" to rights management and did not align with ISWC and ISRC identifiers used by industry. As a result, public knowledge graphs and registries do not interoperate smoothly with private-sector identifiers. The result is duplication, costly reconciliation, and under-use of culturally significant catalogues.

EOSC, intended as Europe's backbone for research data, is also relevant. Its federated model provides long-term preservation and persistent identifiers (via Zenodo and OpenAIRE), and music datasets deposited there already attract visibility. But EOSC has no dedicated workflow for music, and industry uptake remains minimal. As with ECCCH, music is underrepresented and rights-aware curation pathways are absent. CITF confirms that lifecycle-based metadata and provenance are prerequisites for integrating cultural and commercial systems[10].

The **European Interoperability Framework (EIF)** helps explain why these gaps persist. Interoperability depends not only on formats but also on legal, organisational, semantic, and technical alignment. Without shared governance and profiles, public and private systems diverge. The **principle of subsidiarity** adds another layer: stewardship over cultural data is distributed across national and regional authorities, as well as private actors. Centralisation is therefore both impractical and politically illegitimate. The challenge is not whether decentralisation should exist, but how to make decentralised contributions work together.[11]
This challenge directly motivates the Observatory's bridging role with EOSC, Europeana, and ECCCH, elaborated in Section 3.2[12].

[10]EOSC provides federated access and persistence through Zenodo and OpenAIRE, but music workflows remain marginal. On EOSC's role, see the *European Strategy for Data* (European Commission 2020). CITF notes that without harmonised identifiers, provenance chains, and semantic profiles, public infrastructures cannot interoperate with private-sector rights workflows, especially in AI contexts where reproduction and transformation rights depend on reliable metadata [Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula (2025), pp20; pp28–33].

[11]The EIF defines layered interoperability (legal, organisational, semantic, technical) (Commission and Digital Services 2017). The *European Strategy for Data* frames subsidiarity as compatible with federation (European Commission 2020). BDVA and the Federation Working Group emphasise that interoperability frameworks are needed to operationalise federation (BDVA/DAIRO 2023; BDVA/DAIRO Federation Working Group 2023).

[12]The EIF defines layered interoperability (legal, organisational, semantic, technical) (Commission and Digital Services 2017). The *European Strategy for Data* frames subsidiarity as compatible with federation (European Commission 2020). BDVA and the Federation Working Group emphasise that interoperability frameworks are needed to operationalise federation (BDVA/DAIRO 2023; BDVA/DAIRO Federation Working Group 2023).

3.1.5 Subsidiarity and infrastructures for scaling music data

The European principle of **subsidiarity** requires that decisions be taken as closely as possible to the citizens they affect. In cultural policy, this means that responsibilities are distributed across multiple levels: in some Member States, culture is managed regionally or provincially; in others, nationally. Beyond public administrations, many important datasets are held by private actors — collective management organisations, platforms, or archives. Any attempt to centralise music data governance would therefore risk losing both legitimacy and local relevance.

Instead, subsidiarity must be built into the design of the Observatory. The **European Interoperability Framework (EIF)** provides a layered model — legal, organisational, semantic, and technical — for reconciling governance across institutions. The **Data Governance Act (DGA)** codifies the same principle: Member States retain stewardship over sensitive datasets, but EU-level standards ensure they can circulate securely and comparably across borders. The **Data Space Support Centre (DSSC)** extends this approach into practice, developing blueprints and building blocks that allow decentralised initiatives to scale. Together, these frameworks show how subsidiarity and federation are not barriers but design principles for data spaces. CITF's three-layer model aligns directly with this reasoning[13].

At the technical level, **Wikidata and Wikibase** provide a proven backbone for collaborative metadata management. They are already embedded in EU infrastructures such as the official EU Knowledge Graph and in national projects like **MetaBelgica** in Belgium. In Flanders, the performing arts field has gone further: since 2017, *Kunstenpunt* and meemoo have published decades of performing arts data on Wikidata, showing how enrichment happens automatically once data becomes part of a wider ecosystem. These pilots illustrate how subsidiarity and federation can work in practice, with decentralised actors maintaining control of their own data while contributing to a shared framework.[14]

The problem of scale makes such infrastructures essential. Large platforms and labels can manage millions of assets cheaply, but small actors cannot. Without shared systems, independent and community-based repertoires remain undocumented because the cost of proper registration exceeds likely revenue. Federated tools — strengthened by automation and AI — are the only realistic way to close this gap.

[13] On subsidiarity and federation: the *Data Governance Act* (European Parliament and Council 2022) and the *European Strategy for Data* (European Commission 2020). On technical frameworks: DSSC's blueprints (Data Spaces Support Centre 2025b, 2025a). On governance: BDVA (BDVA/DAIRO 2023) and the Federation Working Group (BDVA/DAIRO Federation Working Group 2023). CITF's foundational layer concerns authoritative identifiers and registries, its semantic layer provides shared meaning across heterogeneous models, and its technical layer covers APIs, mappings, and resolution services. Together these layers provide a structured approach for embedding subsidiarity into copyright data governance without forcing schema or organisational unification (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, pp23–33).

[14] On official adoption: EU Knowledge Graph (Diefenbach, De Wilde, and Alipio 2021); SEMIC guidelines (SEMIC Support Centre 2023). On Belgian pilots: *MetaBelgica* (Stallmann et al. 2023) and Flemish performing arts enrichment (Magnus and Van D'huynslager 2021).

Finno-Ugric Data Sharing Space

Our pilot with the **Finno-Ugric Data Sharing Space** illustrates subsidiarity in practice (see: https://finnougric.net/). By collaborating with regional NGOs and national archives, we curated and repaired datasets that would have remained invisible in a central repository. The project showed that decentralised actors are best placed to manage their own data, but that interoperability frameworks and shared observability layers can connect them effectively [15].

International comparison confirms this. In the United States, the **Mechanical Licensing Collective (MLC)** was created in 2021 to administer a blanket mechanical license for streaming and downloads. It inherited more than $424 million in unmatched royalties and developed large-scale reconciliation systems to allocate them. By 2022, it had already distributed nearly $700 million. The MLC shows what can be achieved when identifiers such as ISWC and ISRC are used systematically and backed by law. But it also highlights the limits of centralisation: creators must still claim and maintain their records, education gaps persist, and disputes between platforms and rights bodies continue.[16]

The U.S. Mechanical Licensing Collective (MLC)

The *Mechanical Licensing Collective* was created under the U.S. *Music Modernization Act* (2018) to administer a blanket mechanical license for streaming and downloads. It inherited more than $424 million in unmatched royalties from digital services and developed large-scale reconciliation systems to allocate them. By late 2022, it had distributed nearly $700 million.

The MLC shows what can be achieved when identifiers (ISWC, ISRC) are captured systematically and backed by legislation. But it also highlights the limits of centralisation: creators must still claim and maintain their records, education gaps persist, and disputes between platforms and rights bodies continue. For Europe, the lesson is clear: scaling metadata infrastructure is possible, but it must respect subsidiarity and federation rather than rely on a single central clearinghouse (Mechanical Licensing Collective 2021; Varghese 2024).

3.1.6 Economies of scale in metadata

Large platforms and major labels can document millions of tracks at very low per-unit cost, because they manage everything in bulk. Smaller actors — independent labels, non-profits, or community archives — face the opposite situation: the cost of registering and maintaining each track is often higher than the revenue it will ever generate. This imbalance explains why so many "frozen" assets remain unregistered and invisible in today's digital ecosystem.

[15] See (Antal et al. 2025; Antal, Pigozne, and Federico 2025).

[16] On the MLC's establishment and operations: (Mechanical Licensing Collective 2021); on contested governance and disputes with platforms: (Varghese 2024).

Without a way to share infrastructure, small actors remain stuck. They cannot afford the per-track cost of full documentation, yet under-documentation ensures their work remains undiscovered. This is not just an accounting issue, but a **structural barrier to diversity** in music data flows. A federated approach, as outlined in Section 3.2.2, is essential to rebalance these inequalities and enable small actors to benefit from the same efficiencies as global players; CITF frames this imbalance as a foundational infrastructure issue.[17]

3.2 Policy Proposals

⚠ Editing reminder

- Open Music Observatory as the convening + conformance + observability layer (not a single database).
- Workflow playbooks: rights→distribution→charting→preservation; change-propagation patterns; provenance trails that survive system boundaries.
- Legal/standards/public investment inline: GDPR legal bases per flow; recommended codes of conduct; lightweight policy for data fitness/quality; funding hooks (ECCCH pilots, national ministries).

ℹ Public–private reconciliation in practice

Reconciling public and private infrastructures: The ALOADED pilot in Latvia
The Unlabel workflow was tested with Latvian archives and the distributor **ALOADED**, showing how public heritage metadata can be reconciled with private supply chains.

- Archival recordings (Hilda Griva's songs and Latvian/Latgalian midsummer songs) were located in the **Latvian Archives of Folklore**.
- Metadata was translated, enriched, and aligned with international authority files.

[17] Comparative research shows that costs per asset decrease sharply with catalogue size, creating scale advantages for majors and global platforms. Without shared infrastructures, small actors are disproportionately disadvantaged. The *Feasibility Study for a European Music Observatory* emphasised this imbalance as a structural barrier (Commission et al. 2020, p9), while the *Music Ecosystem 2025* study highlighted how fragmentation and duplication reinforce these scale inequalities (Music Moves Europe 2024). CITF frames this imbalance as a foundational infrastructure issue: without open, authoritative identifiers and interoperable registries, small actors face disproportionately high documentation costs and cannot benefit from economies of scale. It therefore recommends strengthening the foundational identifier layer as a precondition for fair and efficient copyright ecosystems (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, pp23–28).

- ALOADED extended this metadata with **DDEX-compliant catalogue transfer** and ingested it into Spotify and other platforms.

This demonstrated that reconciliation between **public infrastructures (archives)** and **private infrastructures (distributors and platforms)** is both technically and institutionally feasible, reconnecting suppressed or marginalised repertoires with contemporary audiences.
See a more technical description of what we did here.

Conformance and observability rules in the *Open Music Observatory* should be designed in line with the *European Interoperability Framework* (EIF) and the FAIR data principles. This ensures compatibility with wider European data space initiatives and reduces integration costs for institutions already adapting to these standards (Commission et al. 2020, p9).

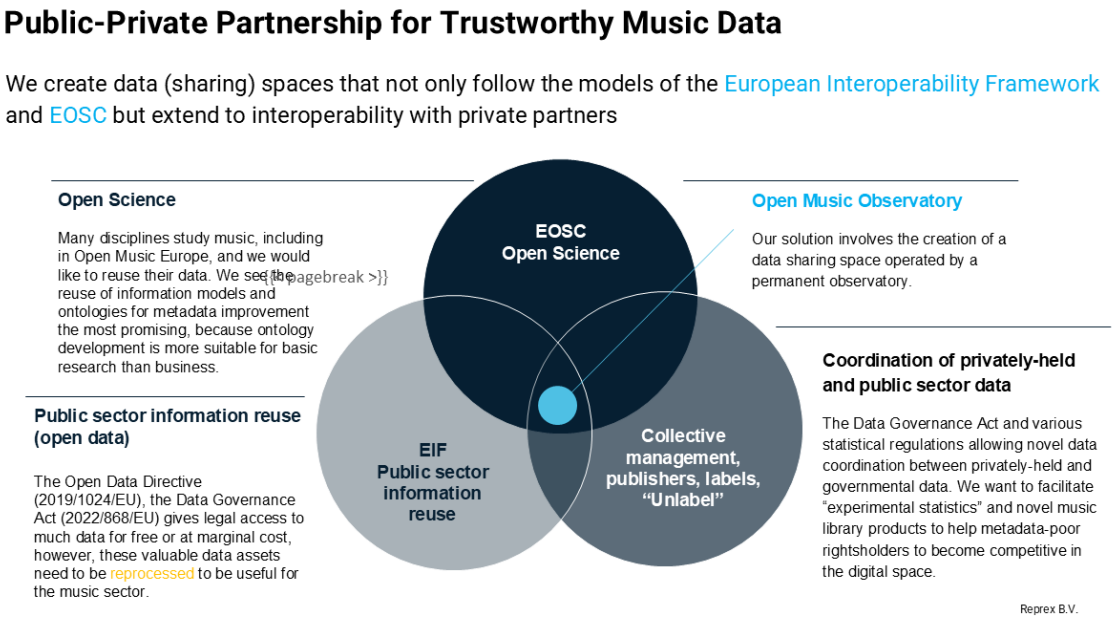


Figure 3.2: The Open Music Observatory sits where open science, public sector information reuse, and music industry workflows overlap. By aligning with the European Interoperability Framework, it creates a shared space where libraries, rights managers, publishers, and researchers can collaborate. This positioning highlights OMO's role as a bridge between cultural heritage, commercial distribution, and open knowledge. DOI: 10.6084/m9.figshare.30073267.v1

3.2.1 Workflow playbooks and provenance trails

The Observatory should not only harmonise data *formats* but also document **workflow playbooks** that capture how metadata flows across the music lifecycle:

- from rights registration,
- to distribution and royalty attribution,
- to charting and visibility,
- to long-term preservation.

Each step should include **change-propagation rules**: if a correction is made in one register, it should ripple through to others. Provenance trails must survive system boundaries, using standards such as PROV-O to show *who did what, when, and under what authority.* This makes corrections auditable, supports cross-border comparability, and prevents "data death" when an asset leaves its original system.

3.2.2 Federated infrastructure as a cost and governance solution

The imbalance described in Section 3.1.6 makes one thing clear: small actors cannot compete on metadata without shared infrastructures. Federation, not centralisation, is the only viable way forward.

A **data sharing space** provides the framework. Instead of forcing everyone into a single metadata schema or legal agreement, it allows organisations to share and reuse data on an *"as-needed"* or *"as-permitted"* basis, while keeping full control of their own assets. For music — where rights, identifiers, and content are dispersed across hundreds of micro-actors and institutions — this model avoids both duplication and dependency. Crucially, it also avoids creating a new *single gatekeeper*: centralisation risks not only technical brittleness but also the emergence of a monopolistic intermediary able to close access or impose conditions on others[18].

Music is one of the most demanding test cases for European data governance. Attribution rules interact with privacy law, identifiers are used unevenly across the sector, and most music enterprises are too small to build their own compliance or documentation infrastructure. If a federated model can function in this environment, it can function anywhere. But decentralisation brings its own challenges: organisations with stronger infrastructures may prefer to protect competitive advantages by withholding data. Effective governance

[18] Our definition here is an extended paraphrase of (Curry 2020) and reflects that a "data [sharing] space is an ecosystem of exchange, processing, sharing and provision of data between trusted partners, for a fee or not" from (EBU and Gaia-X 2022, p16). CITF complements this argument by emphasising that trustworthy provenance, machine-readable rights metadata, and auditable RMI are essential components of a federated copyright infrastructure. It stresses that lifecycle-based provenance chains are necessary not only for attribution but also for AI-era compliance, where training and generation both trigger rights that depend on reliable metadata [Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula (2025), p31; pp101–102].

therefore has to make participation more attractive than isolation — through lower administrative costs, increased visibility, legal clarity, or shared compliance benefits. In practical terms, this means combining *hard alignment* (such as minimal metadata profiles and the use of basic identifiers) with *soft alignment* (such as mappings, crosswalks, and workflow playbooks). This mix allows different actors to negotiate interoperability without forcing anyone into a single model.

For these reasons, the *European Music Observatory cannot be designed as a single central database*. It must act as a convening and observability layer — a place where decentralised contributions can be compared, connected, and reused. Such a structure reduces duplication, lowers costs for smaller actors, improves attribution, and provides a stable governance foundation for trustworthy AI and evidence-based cultural policy.

Europe already has the policy and technical foundations for this. The *European Strategy for Data*, the *Data Governance Act*, and the *Data Act* all define data spaces as **federated by design**, supported by trust frameworks, rulebooks, and shared services.[19] The *Data Spaces Support Centre (DSSC)* has translated these into practical blueprints that can be applied directly to the music sector.[20] Other domains offer concrete precedents: the ISRC system distributes responsibilities across national agencies; CISAC's *CIS-Net* provides access to rights data without centralising ownership; and European statistical systems harmonise indicators through subsidiarity rather than through central repositories. The music sector can — and should — build upon the same logic: distributed stewardship, shared standards, and coordinated interoperability.

Concerns about sovereignty make this design choice even more urgent. Without a European solution, metadata infrastructures risk drifting toward US-style centralisation, such as the *Mechanical Licensing Collective (MLC)*, where market power and legislative mandates converge in a single hub. Deliverable D5.6 of the OpenMusE project explicitly warns against this outcome: unless Europe develops its own federated metadata infrastructures, it risks outsourcing control over visibility, attribution, and royalty data to foreign platforms.[21]

Our proposal for a federated *Open Music Observatory* therefore complements this legal-institutional analysis. Where D5.6 highlights the provisions in EU law that can be mobilised, the present Green Paper demonstrates how a distributed dataspace model can translate them into practice. Taken together, they offer a dual strategy — one legal-institutional, one cultural-sovereignty — for securing Europe's music ecosystems.

In practical terms, this means applying *capture once, reuse many* pipelines across the entire music lifecycle: from registration of works and recordings, through distribution and

[19] The *European Strategy for Data* (2020) defines Common European Data Spaces as federated ecosystems, while the *Data Governance Act* (2022) and *Data Act* (2023) supply the governance and access rules (European Commission 2020; European Parliament and Council 2022).

[20] The *Data Spaces Support Centre (DSSC)*, led by KU Leuven with GAIA-X and BDVA, provides practical blueprints and building blocks for implementing federated data spaces in any domain (Data Spaces Support Centre 2025b, 2025a).

[21] See *Policy Brief 1: Music Metadata Mainstreaming and EU Law* (Senftleben et al. 2024). The brief warns that US-style centralisation (e.g. the MLC) shows the risks of failing to establish European metadata infrastructures. This *Green Paper* complements this by presenting the federated, culture-led model of an Open Music Observatory as Europe's sovereignty-preserving alternative.

royalty attribution, to preservation and cultural statistics. To achieve this, the Observatory must be designed as a *redundancy-free registration space*, aligned with the *European Interoperability Framework* and provenance-oriented models such as *PROV-O*.[22]

Done well, this would rebalance the playing field: lowering costs for small actors, making datasets interoperable across institutions, and ensuring that Europe's cultural and economic policies rest on reliable evidence rather than fragmented silos.

3.2.3 Legal, standards, and funding levers

For these proposals to succeed, they must be backed by legal clarity, lightweight standards, and public investment:

- **Legal:** GDPR legal bases should be specified for each data flow (legitimate interest for attribution; research and cultural heritage exemptions for archives).
- **Standards:** Codes of conduct and minimum profiles should keep conformance achievable even for micro-enterprises.
- **Funding:** ECCCH pilots, national ministries, and EU programmes should explicitly support metadata fitness and data-quality improvements as public-interest infrastructure.

Embedding these levers ensures that interoperability does not remain voluntary but becomes a supported and sustainable practice across Europe.

3.2.4 Alignment with the European Open Science Cloud

Bridge cultural clouds and market workflows via a federated Music Data Sharing Space. Position the Open Music Observatory as the *convening + conformance + observability* layer that connects ECCCH/Europeana and GLAM authority files with industry pipelines. Concretely:

1. **Capture once, reuse many** across creation→registration→distribution→preservation.
2. Require **minimal profiles** that smaller actors can actually implement.
3. Prioritise **identifier crosswalks** (ISRC ISWC ISNI VIAF/Wikidata) and change-propagation.
4. Use **Wikibase/Wikidata** as a low-friction backbone where appropriate.
5. Govern with EIF/FAIR-aligned rules, auditability, and PPP participation so rights-holders and memory institutions keep stewardship while interoperating.

[22] The EIF ensures interoperability across legal, organisational, semantic, and technical layers (Commission and Digital Services 2017). The W3C's PROV model and PROV-O ontology offer a standard way to connect actors, activities, and entities in chains of attribution (W3C 2013b, 2013a). Applied together, they enable consistent tracking of economic and cultural flows without centralising databases.

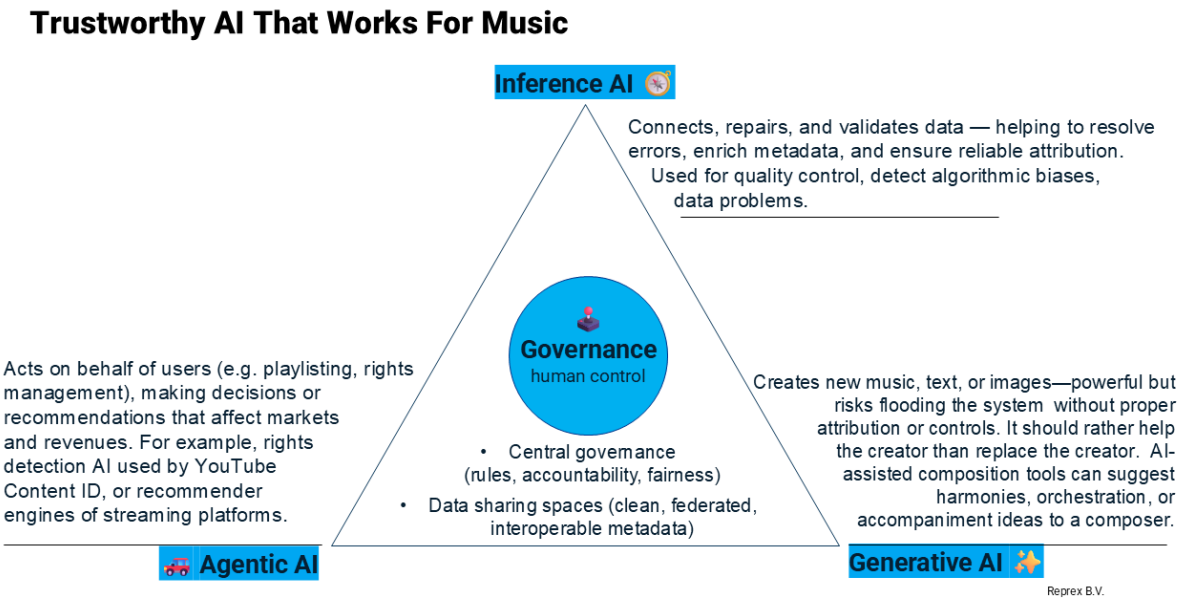
This reframes Europe's investments from siloed repositories into a shared **data space** that lowers reconciliation costs, respects subsidiarity, and makes cultural metadata usable across public and commercial contexts — the practical foundation for any future European Music Observatory.

4 AI that Works for Music, Not Against It

Most AI projects fail because they chase hype. MIT's Project NANDA found that 95% of enterprise initiatives with generative AI delivered no measurable value. Budgets were spent on flashy pilots in sales or marketing, while the real potential — reducing back-office costs, prolonging the life of legacy systems, and avoiding constant IT churn — was overlooked.

Our approach is different. We do not see AI as "for its own sake." Instead, we treat it as a way to reduce IT churn, keep legacy systems alive longer, and cut both capital and operating expenses. Where once every new regulation, distributor change, or catalogue migration required costly upgrades, **curative AI can patch outputs from existing software, extend the lifespan of old systems, and make them interoperable with new ones**. Shared infrastructures make this practical for micro-enterprises, NGOs, and collective management organisations (CMOs), who could never maintain such capacity in-house.

The European Parliament's resolution on the music streaming market warns of the risks that AI-generated content poses for *discoverability, attribution, and fair remuneration* if metadata remains incomplete or unreliable. At the same time, the *Music Ecosystem 2025* study highlights that AI will be both a disruption and an opportunity: while it can overwhelm systems with synthetic material, it also offers tools to automate documentation, reduce costs, and strengthen evidence-based policymaking (Music Moves Europe 2024, 23–24).



Artificial intelligence is therefore central to the future of Europe's music ecosystem. On one hand, it threatens to exacerbate existing inequalities by concentrating technological advantages in platforms and major rights holders. On the other, it can repair, enrich, and automate processes that are otherwise prohibitively costly for small actors. The challenge is not whether AI will be used, but whether its benefits will be distributed fairly across the ecosystem.

European policy provides guidance for this balancing act. The *Ethics Guidelines for Trustworthy AI* underline that AI must be lawful, ethical, and robust throughout its lifecycle (Commission, Directorate-General for Communications Networks, and Technology 2019). The *Getting the Future Right* report by the Fundamental Rights Agency stresses the need to align AI with fundamental rights, especially where vulnerable groups and cultural participation are concerned (European Union Agency for Fundamental Rights 2020). Most recently, the **AI Act** enshrines a risk-based regulatory framework, defining obligations for providers and deployers of AI systems while reaffirming the principles of subsidiarity and proportionality in EU digital policy (European Parliament and Council 2024).

Our own engagement with these issues began with the *Listen Local* feasibility study in 2020. By experimenting with the Spotify API, we discovered that Slovak users were rarely recommended Slovak music — not because Spotify was at fault, but because the data about local repertoire was sparse. Spotify's open API was, in fact, uniquely transparent compared to competitors, and it enabled us to see a larger policy problem: without structured, machine-readable knowledge of diverse repertoires, algorithms cannot deliver fair outcomes. This lesson has guided our work ever since: improving metadata and interoperability is the first step to better AI governance.CITF reaches the same conclusion[1].

[1]The CITF report stresses that AI-related obligations cannot be met unless rights metadata, identifiers, and provenance chains are trustworthy and repairable, and that fragmented RMI makes it impossible to

Case Study: Lessons from the Spotify API

During the *Listen Local* feasibility study (2020), we experimented with the Spotify API and found that Slovak listeners were rarely recommended Slovak music. This was not Spotify's fault. In fact, Spotify's open API and conceptual documentation gave us more insight than any competing platform (Deezer and Apple never even replied to our requests).
That transparency revealed a deeper policy issue: without structured, machine-readable metadata on diverse repertoires, even the most advanced recommender systems cannot deliver fair results. This insight shaped the rest of our work — showing that better outcomes depend not on blaming algorithms, but on supplying them with the right knowledge.

Today, the same dynamics are playing out across the ecosystem. **Agentic AI** powers recommender systems and rights management tools; **Generative AI** is spreading rapidly, raising fears for the economic basis of music; and **Inference AI** offers a path to guardrails, cross-checking recommendations against cultural policies (such as local content quotas) or verifying generative outputs against copyright and attribution rules. Our proposal is triangular: agentic, generative, and inference AI should complement and, when necessary, correct each other — always with the human in ultimate control.

Generative AI is threatening the livelihoods and existing rights of rights holders. The *Open Music Europe* consortium did not study generative AI in-depth; our initial focus was on untrustworthy use of agentic AI[2]. We want to highlight again our analysis from Chapter 2 that metadata is never neutral.

Against this backdrop, the Observatory proposes AI not as a substitute for human creativity or governance, but as a **shared utility**: a way to pool curative, agentic, and generative services within a federated infrastructure. This ensures that SMEs, non-profits, and community archives gain access to trustworthy AI capacities, reducing costs and risks while preserving diversity and accountability in the European music ecosystem.

ensure fair attribution or compliant AI outputs [Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula (2025), pp20; p31].

[2]The term *agentic AI* refers to AI systems that operate with some degree of autonomy and goal-directed behaviour, and it is not present in the language of the European AI Act. In academic and technical contexts, this covers *agent-based frameworks* (where AI systems act in an environment, plan, and adapt), which can be applied either to *generative models* (e.g. creating music or images) or to *workflow automation* (e.g. metadata correction, licensing negotiation). By contrast, recommender systems such as collaborative filtering — widely used in music platforms — are not agentic, since they do not act independently of user input. We retain the shorthand "agentic AI" in this Green Paper for accessibility, but recognise that multiple sub-classes of AI systems are involved.

4.1 Discussion

4.1.1 Structural problems for music businesses to apply AI

1. **AI benefits are unevenly distributed.**
   Music businesses operate in value chains where platforms and large intermediaries already use agentic, generative, and even inference AI. Most platforms already rely on agentic workflows (matching, recognition, playlisting, claim resolution). These actors reap most of the benefits, while smaller publishers, labels, and managers may not even be aware that AI is shaping outcomes in discoverability, rights management, and revenue flows.

2. **AI impacts the bottom line in multiple ways.**

- *Operating costs (OPEX):* Most European music is released by self-publishers or very small labels who cannot afford dedicated staff for documentation or accounting. They save costs by using Excel or freelance accountants, but per unit this is very expensive and leads to poor metadata. As a result, their works perform badly on agentic platforms where poor documentation means poor sales. AI could sharply reduce documentation and claims costs — but deploying it is not easy.
- *Capital costs (CAPEX):* Investing in proper IT or ERP systems is rarely viable at small scale. A system that pays off when managing a million works is wasteful when managing 3,000. Curative AI could extend the life of outdated IT and reduce the need for costly replacements.
- *Working capital:* Many rights holders experience late or missing royalty payouts, even for well-known artists, because the cost of claiming is high compared to the low value of claims. This ties up cash between payment periods. AI could accelerate claims processing and improve matching, smoothing liquidity.
- *Sales:* While dedicated "sales AI" projects are often prone to failure, in music most transactions already run through agentic AI on platforms like Spotify, YouTube, TikTok, and Apple Music. Simply providing these agents with better documented music can improve sales outcomes without the need for standalone sales AI.

3. **Generative AI is only part of the problem.**
   Public debate often focuses on generative AI flooding the market with unlimited non-copyrighted music, which can devalue existing repertoires. This is a real issue, but it is not the only one. Agentic AI in distribution platforms has been shaping the market for at least 14 years, determining who gets discovered, listened to, and paid — long before generative AI became a concern[3].

[3]Surveys and management research confirm these patterns. PwC's *Global CEO Survey* shows how quickly generative AI rose from a marginal issue in 2023 to a central boardroom concern by 2024–25, though most executives expressed only "bounded optimism" (PwC 2024). Bloomberg and BCG's *CEO Radar* tracked quarterly earnings calls in 2025, reporting a 100% increase in references to AI and machine learning, but also rising caution about productivity claims (Bloomberg and Boston Consulting Group 2025). MIT's *Project NANDA* concluded in August 2025 that 95% of enterprise generative AI initiatives

4. **Severe talent shortages.**
   Recruiting and integrating digital expertise is difficult across industries, but especially in music where most enterprises are micro-enterprises. A Chief Data Officer (CDO) is often recommended, yet unrealistic for most publishers, labels, or agencies. Even Fortune 500 companies — far larger than Europe's 50,000 "large" enterprises — report persistent difficulties in filling CDO and AI leadership roles. With 23 million SMEs in Europe, and several hundred thousand music entities, usually with less than 2 people in full-time positions, this AI and data talent shortage cannot be solved on an individual business level[4].

4.1.2 European regulation that misses the point

Europe prides itself on having some of the world's strictest AI rules. Compared to the United States and China, the EU has adopted a risk-based framework in the **AI Act**, with strong obligations for high-risk systems (such as self-driving cars) and lighter rules for low-risk ones. But this framework is poorly suited to music.

Music was classified as "low-risk" on the assumption that nobody is harmed by being offered a bad song. This framing ignores how **agentic AI governs the marketplace itself**. If recommendation systems consistently fail to show music created by women, small nations, or minorities, they devalue those repertoires to zero by depriving them of discoverability. Copyright value is based on the present value of expected royalty flows; if works are never recommended, those flows vanish, and with them the rights protected under EU law and international treaties.

In other words: Europe regulates AI strictly where physical safety is at stake, but does not protect cultural diversity, women's authorship, or the economic rights of creators. What is framed as "low-risk" can in practice be *systemically high-risk* for the music ecosystem. This problem is then reflected in the actual design of commercial or institutional AI systems.

The problem with this categorisation is even more problematic with the rise of large language models (LLMs) and their applications like ChatGPT, Gemini, Llama. The "Human Artistry Campaign" was initiated by a coalition of 150+ organisations, including major music industry bodies (IFPI, RIAA, BPI) and artist representative groups (AIM, Featured Artists Coalition, Impala), establishing a collaborative effort to advocate for responsible AI development within the creative sector.

---

failed to deliver measurable value, with back-office automation offering the clearest returns (MIT Sloan School of Management 2025). These findings mirror evidence from talent studies: Gartner's *CDO Survey* reports persistent shortages in chief data officer and AI leadership roles, even among Fortune 500 companies (Gartner, Inc. 2024), while PwC's *Digital IQ* survey highlights the difficulties of capturing ROI on digital transformation and AI investments (PwC 2023).

[4]According to Eurostat's *Culture statistics — 2023 edition*, cultural and creative industry (CCI) enterprises in the EU are overwhelmingly micro-enterprises. More than 95% employ fewer than 10 people, and the average enterprise size across the sector is below two employees (Eurostat 2023). This structural feature explains why most music publishers, labels, and agencies lack in-house IT, accounting, documentation, or HR functions — and why recruiting specialised AI or data talent is unrealistic without shared infrastructures.

However, recent research questions if the EU's AI Act is even practically applicable as a legal framework to Generative AI. The Act's risk-based categorisation may struggle to capture the emergent behaviour of LLMs and their potential for misuse, and it is highly questionable that human oversight or human control is possible with LLM alone[5].

i Metadata repair may increase generative AI risks

Last, but not least, we want to highlight that efforts at **metadata repair and publication** — as we propose in earlier chapters — also increase the risk of generative AI misuse.
A prompt like `create me an ABBA-like disco hit` is likely to combine:

- **Core musical learning** from audio/MIDI (raising clear copyright and GDPR risks), and
- **Metadata signals** that guide the model's interpretation.

Core musical learning (from audio/MIDI):
From training on ABBA's catalogue (and related artists), a model learns to reproduce:
- Harmony → diatonic progressions (e.g. I–V–vi–IV), bright major keys.
- Melody → catchy, stepwise motifs (e.g. *Mamma Mia* hooks).
- Rhythm & texture → steady 4/4 grooves, piano/guitar foundations, layered vocals.
- Structure → verse–chorus–bridge arcs with memorable refrains.
- Production cues → lush vocal overdubs, polished pop arrangements, disco influences.

How metadata sharpens the imitation:
- *Artist metadata ("ABBA")* → links to Swedish pop, Eurovision history, chart success. Tags like "Europop," "disco-pop," "vocal harmony group" cue specific stylistic markers. Models may also trace producers and collaborators to expand training.
- *Genre/award metadata ("disco-pop hit")* → narrows toward 1970s–80s tropes: syncopated basslines, string pads, tambourine.
- *Chart/award metadata* → biases output toward catchy, chorus-driven songs resembling global hits.

The paradox:
- Metadata repair makes ABBA's legacy more discoverable across platforms (archives, streaming, Wikidata, Wikipedia).
- But richer metadata also helps AI **pinpoint and reproduce their exact style**, making prompts like "ABBA-style Eurovision anthem" feasible.

Metadata empowers and exposes: it restores visibility for heritage and repertoire, but also creates new pathways for substitution by generative AI.
For this reason, metadata governance is a **policy concern**, not just a technical task[6].

[5] *Towards Responsible AI Music: an Investigation of Trustworthy Features for Creative Systems* is an excellent review of the theoretical or practical applicability of the EU's trustworthy AI paradigm for generative AI (Berardinis et al. 2025). For more information on the *Human Artistry Campaign*, see https://www.humanartistrycampaign.com/.

[6] CITF frames this as a core AI-era requirement. It argues that lifecycle-based provenance and machine-readable RMI are essential for assessing lawful uses in training and generation, and that without

4.1.3 Policy issues at the intersection of AI, copyright, and GDPR

AI in music does not operate in a legal vacuum. It interacts with existing European law on intellectual property, author's rights, moral rights, and data protection. In practice, this creates tensions and unresolved policy gaps that directly undermine cultural policy goals. This governance problem builds directly on the interoperability failures described in Section 3.1.3.

1. **Attribution vs GDPR.**
   The Treaty on the Functioning of the European Union enshrines protection of intellectual property. European copyright law gives authors moral rights, including attribution. Yet GDPR may prohibit storing or publishing the same identifying data needed to respect these rights. In the absence of jurisprudence from the Court of Justice of the EU or guidance from competent data protection authorities, actors who try to give proper attribution risk GDPR penalties. This legal uncertainty has direct implications for AI:

   - If attribution is blocked, it becomes impossible to test whether AI systems treat authors fairly.
   - More broadly, GDPR makes it difficult to safeguard against algorithmic discrimination if information on gender, nationality, or other attributes cannot legally be used[7], and the current announced revision of GDPR by the Commission is the best moment to address this problem.

2. **Local content protection gaps.**
   In broadcasting, local content quotas were established in line with WTO rules to safeguard cultural diversity (e.g. Slovak private radios playing at least 25% Slovak music). Similar obligations now exist in audiovisual streaming. But in music streaming there are no binding European diversity or local content rules. This creates two problems:

   - AI-driven distribution platforms can crowd out local repertoire with global catalogues, depriving smaller nations of audiences.
   - Even where voluntary quotas exist, compliance depends on knowing the origin of repertoire. If we cannot know whether a work is Slovak, French, or by a young author, quotas or diversity targets cannot be implemented.

3. **Voluntary compliance is impractical.** Current practice relies on voluntary measures by radio editors, festival curators, or platform users to include local or diverse content. But without accessible data, this becomes unworkable. Our own experiments with GDPR balancing tests and opt-ins show the futility of this approach.

interoperable identifiers, neither attribution nor AI governance can scale [Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula (2025), pp28–33; pp101–102]. This tension is also recognised in recent policy research.

[7] CITF identifies this same contradiction. It notes that attribution data is simultaneously necessary for copyright, required for RMI, and treated as personal data under GDPR, creating legal uncertainty that directly undermines AI governance, lifecycle compliance, and fairness testing (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, pp12–18).

Fewer than 1% of artists responded to requests to consent to attribution data — even prominent Slovak artists, puzzled at being asked to consent to rights they already legally hold.

In short, **AI cannot be made trustworthy for music without resolving these legislative and policy blocks**. The AI Act currently misplaces risk, treating music as "low-risk" while ignoring systemic harms. GDPR, in practice, blocks data use that would enable fairness testing. And the absence of local content rules in streaming removes a cornerstone of cultural policy. AI in music will remain misaligned with European policy goals unless these conflicts are addressed.

4.1.4 AI design without awareness of limits

AI systems are not usually designed with an awareness of their own conceptual limits.

- **Agentic AI** systems (recommenders, playlist builders, rights-management bots) operate without recognising the biases or incompleteness of the datasets they learn from. Because European legislation deems the agentic use of AI in music "low risk", currently there are no real expectations to address this problem.
- **Generative AI** produces synthetic material without constraints, and its training processes seldom acknowledge gaps or skew in the underlying data. This problem touches upon various issues that we discussed earlier in this paper: author's rights and performer rights are assigned to natural persons (and their heirs), as well as sometimes producer's rights, too. GDPR currently appears to conflict both with designing safer AI systems and with providing proper attribution without legal risk to creators of protected work. We could technically guardrail generative AI to not produce plagiarism, but not without giving it access to whose work is forbidden.
- Even **Inference AI**, which is supposed to reason from formal rules, can miss the point: ontological relativity and incompleteness are structural limits and not optional refinements. We discussed in the Chapter 2, and just as well as database designers must be aware that that no ontology or schema is ever complete, AI engineers must realise that they train algorithms that cannot capture all perspectives. Without this awareness, AI will silently reproduce exclusions — whether of women, minorities, or smaller repertoires — while appearing "intelligent."

This is a design issue: the guardrails must be built in from the start, not bolted on afterwards. We see a lot of promise in building Inference AI tools, perhaps in a public-private partnership, that can actually provide help for human-in-control principles for the use of agentic and generative AI.

4.1.5 Unfreezing frozen assets

Many music assets remain "frozen" because their documentation costs exceed their current commercial value. This applies to non-commercial repertoires, small-label releases, and culturally valuable but low-market recordings. Without affordable workflows, these works cannot enter modern distribution systems, regardless of their cultural or artistic significance.

The *Unlabel* pilot illustrates this problem: by treating catalogue transfers and documentation as high-cost, high-friction processes, valuable repertoires remain locked away. AI-assisted metadata repair and DDEX-compliant catalogue transfer workflows provide a pathway to lower costs and bring neglected repertoires back into circulation.

Note

Example: Old SQL Database in a Cultural Institution

- A label or archive has a recording stored in a **20–30 year-old SQL database**, built on a schema that was never fully documented. The system's author is retired (or no longer alive).
- The institution wants to re-release the recording, but to distribute it today, the metadata must be expressed in **DDEX Catalogue Transfer messages** — a completely different schema, designed decades later.

Curative AI

- Acts at the *system level*: it can "read" the old database structure, infer undocumented field meanings, and **patch outputs** so the legacy database can still talk to modern pipelines.
- Instead of rebuilding or migrating the old database (expensive, risky), curative AI *extends its lifespan* by making its outputs usable.

Reparative AI

- Acts at the *metadata/epistemic level*: it can detect inconsistencies or missing fields (e.g., composer names stored in free-text notes, titles in mixed languages) and **reformat or enrich** them into structured DDEX-compliant fields.
- This not only enables distribution but also **restores visibility** for works that might otherwise remain trapped in inaccessible formats.

The policy point

- Without curative/reparative AI, such recordings risk becoming "frozen assets": legally owned but practically undistributable because the metadata cannot be transformed.
- By investing in these AI uses, Europe can preserve access to cultural heritage, reduce IT churn, and ensure that both **heritage archives** and **independent labels** can connect to modern digital value chains.

Unlike U.S.-style copyright, Europe's author's rights regime contains a moral component. Authors (and, for a period, their heirs) retain certain rights over how their works are used, even after economic rights expire. This recognises that works are part of a creator's moral and cultural heritage, not only economic assets. Various legal norms, for example, local content guidelines, also gave tool earlier to national or ethnic communities to provide some guardrails to the use of their shared heritage, even this means community stewardship and not *inheritance* in legal terms.

Metadata repair and publication strengthen visibility, but also create risks that generative AI will use these works in ways that undermine **moral rights**, where heirs object to uses they see as distorting or trivialising an author's legacy and **community stewardship norms**, where groups perceive their folk or minority heritage as being misappropriated, even when no legal infringement occurs.

While we do not identify these challenges at this point as similarly actionable public policy challenges as the problems of GDRP and the creation of trustworthy music AI, regulators do face political risk if **ethical expectations of communities** around cultural stewardship are not addressed. Even if no author's rights or other legal norms are breached, the ability to create "fake" Livonian, Latvian or Basque folk songs may strongly conflict with the expectation of communities on the ethical use of AI.

4.1.6 AI support for investment into new repertoire assets

While generative AI that disregards human repertoires can undermine cultural value, AI also has constructive roles. Just as photographers benefit from embedded AI in tools like Photoshop or GIMP, musicians and producers can use AI to reduce the costs of composition, recording, and documentation. In practice, this means that creating new works and registering them with identifiers can become less burdensome and more accessible.

This perspective aligns with the European Parliament's call for "metadata from birth" (European Parliament 2024), but it goes further. AI can not only generate metadata automatically at the moment of creation, but also support sound recording, scoring, and archiving processes directly, ensuring that new assets enter circulation with complete, interoperable metadata.

4.2 Policy Proposals: Aligning AI with Governance and Value Creation

Generative, agentic, and inference AI are now woven into the global creative economy. But value is not created by algorithms alone — it comes from **governance, curated data, and institutions that ensure trust**. Policy interventions are needed on three levels: **EU**, **industry**, and **organisational**.

Our focus is the **metadata and data needs of the music ecosystem** — labels, distributors, publishers, managers, CMOs, archives — not the creative act of composing music itself.

4.2.1 EU-Level Policy: Compass and Guardrails

- **Embed cultural sectors in the EU AI Act & Data Spaces** so music and cultural industries are not treated as "low risk."
- **Subsidise shared AI utilities** for identifier reconciliation, metadata repair, and fraud/plagiarism detection.
- **Adopt "metadata from birth" principles**: embed ISNI/ISWC/ISRC identifiers at the point of creation.
- **Tax incentives for onboarding frozen assets**, supporting digitisation and enrichment of under-documented catalogues.
- **Resolve attribution vs GDPR conflicts** through legal clarification or jurisprudence, enabling fairness testing and copyright compliance.

4.2.2 Industry-Level Policy: Standards and Collaboration

- **Codes of conduct for AI in music**, modelled on GDPR codes.
- **Identifier crosswalks** across ISRC, ISWC, ISNI, VIAF, etc.
- **Federated AI services** for claims, reconciliation, multilingual enrichment.
- **Training and reskilling** to close the AI/data talent gap.
- **Working capital optimisation** through AI-assisted claims and faster distributions.

These principles do not stand in isolation: they echo and extend ongoing work such as the **Responsible AI Music framework**, ensuring that sector-specific practices in Europe are consistent with emerging international standards.[8]

[8]The *Responsible AI Music framework* (RAIM) sets out principles for transparency, fairness, sustainability, and accountability in the use of AI in music (Herremans, Sturm, et al. 2025). Several of the codes of conduct proposed here — such as clarity around data provenance, safeguards for attribution, and limits on exploitative recommendation practices — align closely with RAIM's recommendations. Where RAIM defines broad principles, this *Green Paper* provides concrete mechanisms for their operationalisation within European music data spaces and observatories.

4.2.3 Organisational-Level Policy: Playbooks for CMOs, Publishers, Archives

- **Embed AI in workflows** so metadata is generated and validated during creation/distribution.
- **Capture once, reuse many times**, reducing redundant re-entry.
- **Invest in knowledge capital, not IT churn** (ontologies, vocabularies, multilingual enrichment).
- **Subscribe to shared AI utilities** instead of bespoke in-house builds.
- **Develop internal AI governance** — even small actors can appoint an "AI steward."

4.2.4 Curative AI and Reparative AI as a Remediation Solution

While data spaces establish rules for *new* data flows, they do not address the **legacy backlog** of poorly formatted or incomplete open data. Here, **curative AI** provides a complementary solution.
AI-assisted services can detect duplicates, infer missing identifiers, reconcile heterogeneous formats, and enrich metadata with multilingual descriptions. In effect, they transform datasets that are legally open but practically unusable into resources that can circulate across the ecosystem.

Note

Curative AI as regeneration, not replacement

Figure 4.1: (Ise Grand Shrine): a wooden sanctuary in continuous use for 1,600 years thanks to regeneration practices handed down through generations.

The Ise Grand Shrine in Japan has been in continuous use for 1,600 years — not because its wooden beams never rotted, but because the knowledge of renewal was embedded and transmitted across generations. The true asset was the **embedded know-how of regeneration**, not any single plank of wood.
Curative AI can play the same role in the digital domain:
- Extend the life of **legacy systems** by fixing patchy outputs from old ERPs, catalogues, or distributor software.
- Preserve the **methods of repair**: how to reconcile corrupted records, reshape data for new systems, and upgrade databases while remaining compatible with older formats.
- Transform investment logic: instead of constant capex for new IT systems, shared data infrastructures with curative AI reduce costs, smooth opex, and deliver **future-proof and past-proof services**.
Our pilots — such as **Unlabel** and **SKCMDb** — show that new value can be created without additional IT investment or system upgrades by the participating companies, libraries, and rights management agencies.

Thus, governance and remediation are two sides of the same coin:
- **Data sharing spaces** ensure that *new* data is created in interoperable ways.
- **Curative AI** repairs the *inherited stock* of legacy and low-quality datasets.

Together, they close the gap between the *right of reuse* (granted by the Open Data Directive) and the *means of reuse* required for music, culture, and AI-driven innovation.

4.2.5 Lowering Documentation Barriers

We propose to adapt *Unlabel*'s approach as a model for unfreezing frozen assets. By leveraging AI-assisted metadata repair and DDEX-compliant catalogue transfer workflows, documentation costs can be reduced enough to enable non-profits, small labels, and community archives to register and redistribute neglected repertoires. Public support should subsidise onboarding costs, create standardised pipelines, and incentivise low-friction reuse of metadata across systems.

4.2.6 Observatory: European = Open

When we call for a **European Music Observatory**, the adjective "European" should not be read as a cultural filter that limits scope to European repertoires. Music is, and always has been, global. The task of the Observatory is not to create an insular archive of "European music," but to build a governance and data architecture rooted in **European values**:

- **Data sovereignty** — ensuring that creators, communities, and institutions have meaningful control over how their metadata and works are represented.

- **Subsidiarity** — solutions should be built at the lowest effective level, allowing national archives, collective management organisations, and industry actors to contribute without being absorbed into a single monolith.

- **Inclusiveness** — minority repertoires, independent artists, and small markets must be equally visible alongside the global catalogues of multinational platforms. Our Finno-Ugric case studies show how fragile metadata can be repaired without erasing community perspectives — a model that must be embedded at Observatory scale[9].

This is why we chose the name **Open Music Observatory (OMO)**. Even if the policy framework ultimately labels it the "European Music Observatory," the essential principle must remain openness — of infrastructure, of governance, and of participation. The Observatory should be a federated, open knowledge space, not a centralised database.

Europe has an opportunity to take a step that resonates beyond its borders. The U.S. *Music Industry Licensing Collective (MILC)* demonstrated how a single initiative could set standards and ripple globally. An **Open Music Observatory**, grounded in European governance but open to the world, could play a similar role — aligning sovereignty with interoperability, and showing how collective data architectures can provide guardrails for AI in a truly global music ecosystem.

4.2.7 The Open Music Observatory as a Collective Guardrail

AI will only create sustainable value for music when governance, interoperability, and human capital are aligned.
But building effective **guardrails for agentic and generative AI** cannot be done by individual firms or even national markets.
- At the **business level**, companies lack the scale and incentives to police AI use of metadata.
- At the **industry level**, cooperation is necessary but often fragmented by competing interests.

This is where the **European Union can play a decisive role**:
- Coordinating and aligning existing investments in **Europeana, the European Collaborative Cloud for Cultural Heritage (ECCCH), and the new data sharing spaces**.
- Anchoring these initiatives in an **Open Music Observatory (OMO)** built around federated, Wikibase-compatible knowledge graphs.
- Ensuring that metadata repair and publication feed into collective data architectures that double as guardrails — improving attribution and interoperability while reducing the risk of generative AI misuse.

[9] We have created the second federated module of the Open Music Observatory with contemporary popular and authentic folk music of European Finno-Ugric minorities who do not have a nation state. (Antal et al. 2025)

Wikidata Embedding Project: An Open Model for AI Guardrails

In 2024–25, Wikimedia Deutschland, in collaboration with Jina.AI and DataStax, launched the **Wikidata Embedding Project**.
- Its goal is to add **vector-based semantic search** to Wikidata, combining its multilingual knowledge graph with modern embedding models.
- This enables **context-aware retrieval** for AI systems while anchoring results in a **public, verifiable knowledge base**.

Why it matters for music policy
- Shows that **guardrails for generative AI** can be built on **open, community-managed graphs** rather than proprietary black boxes.
- Demonstrates how semantic search and retrieval-augmented generation can:
- Reduce hallucinations by grounding outputs in human-verified data.
- Combat misinformation with verifiable references.
- Amplify underrepresented knowledge by balancing global visibility.

Implication for the Open Music Observatory (OMO)
- By adopting **Wikibase-compatible knowledge graphs** and existing ontological patterns, OMO can build similar guardrails for music.
- This positions metadata repair and publication not just as technical fixes, but as part of a **collective data architecture** that keeps AI accountable.

The OMO model would provide:
- **Compass and coordination** at the EU level.
- **Standards and shared utilities** through industry cooperation.
- **Flexible governance and playbooks** for organisations.

With this architecture, AI becomes an infrastructure for *continuous renewal*: prolonging legacy systems, unfreezing frozen assets, and supporting both heritage and new repertoires — while embedding guardrails against substitution and misappropriation into the very data fabric of Europe's music ecosystem.

5 What Europe Should Do Next for Music Data & AI

Europe's music ecosystem is under pressure. Streaming pays in micro-royalties, metadata mistakes cost real money, and AI threatens to overwhelm platforms with untracked content. But solutions are within reach. This Green Paper sets out a path forward, built on three pillars: better metadata, shared data spaces, and AI that works for everyone. (See Chapter 1 for the background and policy context.) These priorities align closely with the emerging framework proposed by CITF: trusted identifiers, lifecycle-aware provenance, and interoperable, federated rights metadata.

The first step is to **fix metadata at the source**. Rights societies, platforms, labels, libraries, and archives all capture fragments of information about works and recordings. Today this is done in parallel, wasting effort and creating errors. Smarter pipelines, shared identifiers, and pragmatic exchange patterns can make documentation "capture once, reuse many." This is not just a technical upgrade — it is the foundation for fair royalties, legal certainty, and cultural visibility. It also directly supports the CITF view that robust identifiers and transparent RMI are prerequisites for trustworthy AI and lawful reuse.

See Chapter 3 for how shared infrastructures can make this possible.

The second step is to build **federated data sharing spaces**. Instead of a single giant database, Europe should connect what already exists: collective management systems, heritage archives, and platform catalogues. Each actor stays in control of its own data but agrees to shared profiles, identifiers, and rules. This approach lowers costs, improves trust, and makes cross-border reuse realistic. The Open Music Observatory is our proposal for such a space: not a central repository, but a convening layer that makes decentralisation work. This federated model reflects CITF's conclusion that copyright infrastructures must remain distributed, but linked through shared semantics, APIs, and provenance trails.

See Chapter 4 for how artificial intelligence can be used to strengthen, not weaken, this foundation.

The third step is to treat **AI as a shared utility**. Big platforms already use AI to document millions of tracks and to steer attention. Smaller players cannot compete unless Europe provides common tools: AI to reconcile identifiers, repair legacy datasets, enrich metadata in multiple languages, and help creators embed information "from birth." If deployed in a federated way, AI reduces costs and unfreezes neglected repertoires — while respecting rights, attribution, and diversity.

Taken together, these steps close the gap between the *right of reuse* granted by the Open Data Directive and the *means of reuse* that the music industry actually needs. They also bring Europe's approach into alignment with the emerging international consensus: that cultural data infrastructures must be trustworthy, distributed, and provenance-rich in order to support both human creativity and responsible AI.

Europe should therefore:

- Support metadata capture and cross-domain identifiers, ensuring that attribution is reliable, repairable, and legally secure.

- Invest in federated data sharing spaces like the Open Music Observatory, enabling decentralised actors to work together.

- Provide pooled AI services for reconciliation, repair, enrichment, and documentation, accessible to SMEs, CMOs, and heritage institutions alike.

This is how Europe can make its music ecosystem fair, efficient, and future-proof — not only safeguarding its cultural heritage, but ensuring that AI strengthens rather than erodes the foundations of musical creativity, diversity, and economic sustainability.

Sources & Further Reading

Abing, Hannah. 2024. "Report Finds More Music Is Released in a Day in 2024 Than in All of 1989 Combined." November 20. https://www.rareformaudio.com/blog/more-music-released-in-a-day-2024-than-1989.

Albanese, Davide, Matthias Balliauw, Alexis Macedo-Rouet, et al. 2023. "ChoCo: An Ontology and Knowledge Graph for Musical Chords." *Scientific Data* 10 (1): 91. https://doi.org/10.1038/s41597-023-02410-w.

Antal, Daniel. 2019. *Slovak Music Industry Report [Správa o slovenskom hudobnom priemysle]*. 2019. https://doi.org/10.17605/OSF.IO/V3BE9.

———. 2020. *Central And Eastern European Music Industry Report 2020*. CEEMID, Consolidated Independent. https://doi.org/10.13140/RG.2.2.21450.31686.

———. 2021. *An Empirical Analysis of Music Streaming Revenues and Their Distribution*. Version 1.0. Zenodo. https://doi.org/10.5281/zenodo.5554089.

———. 2022. *The Music Industry Value Chain*. Figshare. https://doi.org/10.6084/m9.figshare.19174310.v1.

———. 2023. *Pilot Program for Novel Music Industry Statistical Indicators in the Slovak Republic*. https://doi.org/10.5281/zenodo.10372026.

———. 2024a. "A szlovák adatkicserélési tér magyarországi föderációjának lehetőségei." In *Az oktatás, a kutatás és a közgyűjtemények digitális transzformációja felsőfokon : NETWORKSHOP 2024 : 33. Országos Informatikai Konferencia : 2024. április 3–5. Eszterházy Károly Katolikus Egyetem, Eger*, edited by József Tick, Károly Kokas, and András Holl, 192–98. Budapest: HUNGARNET Egyesület. https://doi.org/10.31915/NWS.2024.25.

———. 2024b. *Building a Music Data Sharing Space with Wikibase*. Version 1.0. Open Music Observatory. https://doi.org/10.5281/zenodo.17078911.

———. 2024c. *Open Music Observatory*. Version 1.1. Digital Music Observatory. https://doi.org/10.5281/zenodo.16539570.

———. 2024d. *Trustworthy AI and Data-Sharing Spaces for the Slovak Music Centre. Poster Presentation at the IAMIC Conference 2024 on November 21, 2024, at Music*

Austria, Vienna. Open Music Observatory. https://doi.org/10.5281/zenodo.16540605.

———. 2025a. *SKCMDb. Interoperability of Music Libraries and Archives with Public and Private Music Services. Presentation at the IAML 2025 Conference in Salzburg, Austria Held on the 7th of July 2025.* Open Music Observatory. https://doi.org/10.5281/zenodo.16634558.

———. 2025b. *Slovak Music Data Sharing Space. Poster Presentation at the IAML 2025 Conference in Salzburg, Austria Held on the 7th of July 2025.* Open Music Observatory. https://doi.org/10.5281/zenodo.15814286.

———. 2025c. *Wikibase as a Data Sharing Space: Connecting Rights, Communities, and GLAM Through Federated Infrastructures. Presentation on Wikidata Conf 2025.* Open Music Observatory. https://doi.org/10.5281/zenodo.17496740.

Antal, Daniel, Ieva Pigozne, and Asmah Federico. 2025. *WikiMuseum = GLAM Wiki + Wikibase + Data Sharing Space*. Finno-Ugric Data Sharing Space. https://doi.org/10.5281/zenodo.17237793.

Antal, Dániel, Ieva Pigozne, Kata Gábor, and Bogáta Tímár. 2025. "Federating Open Knowledge Graphs: Finno-Ugric Cultural Heritage and Semantic Infrastructure Facilitated Through Wikibase [Preprint]." *Proceedings of the Digital Humanities in the Nordic and Baltic Countries (DHNB) Conference* 7.

BDVA/DAIRO. 2023. *Data Sharing Spaces and Interoperability: BDVA Discussion Paper*. BDVA/DAIRO. https://www.bdva.eu.

BDVA/DAIRO Federation Working Group. 2023. *Federated Data Spaces: Position Paper*. BDVA/DAIRO. https://www.bdva.eu.

Berardinis, Jacopo de, Valentina Anita Carriero, Nitisha Jain, Nicolas Lazzari, Albert Meroño-Peñuela, Andrea Poltronieri, and Valentina Presutti. 2023. "The Polifonia Ontology Network: Building a Semantic Backbone for Musical Heritage." In *The Semantic Web – ISWC 2023*, edited by Terry R. Payne, Valentina Presutti, Guilin Qi, María Poveda-Villalón, Giorgos Stoilos, Laura Hollink, Zoi Kaoudi, Gong Cheng, and Juanzi Li, 302–22. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-47243-5_17.

Berardinis, Jacopo de, Lorenzo Porcaro, Albert Meroño-Peñuela, Angelo Cangelosi, and Tess Buckley. 2025. *Towards Responsible AI Music: An Investigation of Trustworthy Features for Creative Systems*. https://arxiv.org/abs/2503.18814.

Bianchini, Carlo, Stefano Bargioni, and Camillo Carlo Pellizzari di San Girolamo. 2021. "Beyond VIAF Wikidata as a Complementary Tool for Authority Control in Libraries." *Information Technology and Libraries* 40 (2). https://doi.org/10.6017/ital.v40i2.12959.

Big Data Value Association. 2019. *Towards a European Data Sharing Space: Enabling Data Exchange and Unlocking AI Potential*. BDVA. https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V1.pdf.

Blomqvist, Eva, Karl Hammar, and Valentina Presutti. 2016. "Engineering Ontologies with Patterns – the eXtreme Design Methodology." In *Ontology Engineering with Ontology Design Patterns*, 23–50. IOS Press. https://doi.org/10.3233/978-1-61499-676-7-23.

Bloomberg, and Boston Consulting Group. 2025. *CEO Radar: AI and Machine Learning References in European Earnings Calls*. Bloomberg L.P.; Boston Consulting Group. https://www.bcg.com/publications/2025/ai-machine-learning-mentions-earnings-calls.

Buttow, Carolina V., and Albert Meijer. 2024. "Public Sector Data Openness in the Crafting of the Data-Driven Society: The Co-Constitutive Role of Regulation and Innovation." *Swiss Yearbook of Administrative Sciences* 15 (1): 12–27. https://serval.unil.ch/resource/serval:BIB_118FD2F827C6.P002/REF.

Carrara, Wendy, Wae San Chan, Sandra Fisher, and Eva van Steenbergen. 2015. *Creating Value Through Open Data. A Study on the Impact of Re-Use of Public Data Resources*. Publications Office of the European Union. https://data.europa.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf.

Carriero, Valentina Anita, Aldo Gangemi, Maria Letizia Mancinelli, Andrea Giovanni Nuzzolese, Valentina Presutti, and Chiara Veninata. 2021. "Pattern-Based Design Applied to Cultural Heritage Knowledge Graphs." *Semantic Web* 12 (2): 313–57. https://doi.org/10.3233/SW-200422.

CISAC/SUISA/SESAC. 2017. *Mint Digital Services: A Transnational Licensing Platform*. CISAC. https://www.cisac.org/Newsroom/society-news/sesac-and-suisa-launch-mint-digital-services.

Colavizza, Giovanni, Tobias Blanke, Charles Jeurgens, and Julia Noordegraaf. 2022. "Archives and AI: An Overview of Current Debates and Future Perspectives." *ACM Journal on Computing and Cultural Heritage (JOCCH)* 15 (1): Article 4, 1–15. https://doi.org/10.1145/3479010.

Commission, European, and Directorate-General for Digital Services. 2017. *New European Interoperability Framework. Promoting Seamless Services and Data Flow for European Public Administrations*. Luxembourg: Publications Office of the European Union. https://ec.europa.eu/isa2/sites/default/files/eif_brochure_final.pdf.

Commission, European, Content Directorate-General for Communications Networks, and Technology. 2019. *Ethics Guidelines for Trustworthy AI*. Publications Office of the European Union. https://doi.org/doi/10.2759/346720.

Commission, European, Sport Directorate-General for Education Youth, Culture, M Clarke, P Vroonhof, J Snijders, A Le Gall, et al. 2020. *Feasibility Study for the Establishment of a European Music Observatory : Final Report*. Publications Office. https://doi.org/doi/10.2766/9691.

Commission, European, Directorate-General for Research, Innovation, P. Brunet, L. De Luca, E. Hyvönen, A. Joffres, et al. 2022. *Report on a European Collaborative Cloud for Cultural Heritage – Ex – Ante Impact Assessment*. Publications Office of the European Union. https://doi.org/doi/10.2777/64014.

Cruz, Maria, and Clifford Tatum. 2021. *NWO Persistent Identifier Strategy*. Zenodo. https://doi.org/10.5281/zenodo.4674513.

Curry, Edward. 2020. "Dataspaces: Fundamentals, Principles, and Techniques." In *Real-Time Linked Dataspaces: Enabling Data Ecosystems for Intelligent Systems*, 45–62. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-29665-0_3.

Data Spaces Support Centre. 2025a. "Data Spaces Blueprint V2.0 — Cross-Data Space Interoperability Considerations in Data Space Design and Operation." https://dssc.eu/space/BVE2/1071252241/Cross-data+space+interoperability+considerations+in+data+space+design+and+operation.

———. 2025b. "Data Spaces Blueprint V2.0 — Introduction: Key Concepts of Data Spaces." https://dssc.eu/space/BVE2/1071251613/Introduction+-+Key+Concepts+of+Data+Spaces.

DataCite. 2021. *DataCite Metadata Schema: Mapping to Dublin Core*. DataCite. https://schema.datacite.org/meta/kernel-4.4/doc/DataCite_DublinCore_Mapping.pdf.

Diefenbach, Dennis, Max De Wilde, and Samantha Alipio. 2021. "Wikibase as an Infrastructure for Knowledge Graphs: The EU Knowledge Graph." In *The Semantic Web – ISWC 2021*, 12922. Lecture Notes in Computer Science. Cham: Springer. https://doi.org/10.1007/978-3-030-88361-4_37.

Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-Use of Public Sector Information. 2019. European Union. https://eur-lex.europa.eu/eli/dir/2019/1024/oj.

EBU, and Gaia-X. 2022. *Dataspace for Cultural and Creative Industries. Position Paper. V.2.0*. Gaia-X. https://gaia-x.eu/wp-content/uploads/2022/10/EBU_position-paper_Media-Data-Space.pdf.

ECHOES Ontology Task Force. 2025. *Heritage Digital Twin Ontology (HDTO) – First Draft*. Technical Report. Version 0.1. ECHOES Project / European Collaborative Cloud for Cultural Heritage (ECCCH). https://github.com/ECHOES-ECCCH/

HDTO-Heritage-Digital-Twin-Ontology.

European Commission. 2020. *A European Strategy for Data*. European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020DC0066.

European Parliament. 2024. *European Parliament Resolution of 17 January 2024 on Cultural Diversity and the Conditions for Authors in the European Music Streaming Market (2023/2054(INI))*. P9_TA(2024)0020. European Parliament. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0020_EN.pdf.

European Parliament and Council. 2022. "Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act)." In *Official Journal of the European Union*, L 152, 1–44. https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng.

———. 2024. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

European Union Agency for Fundamental Rights. 2020. *Getting the Future Right. Artificial Intelligence and Fundamental Rights*. Luxembourg: Publications Office of the European Union. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf.

Europeana. 2017. *Definition of the Europeana Data Model V5.2.8*. Europeana. https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf.

EuroSDR. 2021. *Sustainable Open Data Business Models for NMCAs. EuroSDR Official Survey Report*. https://lirias.kuleuven.be/retrieve/696771.

Eurostat. 2023. *Culture Statistics — 2023 Edition*. Publications Office of the European Union. https://doi.org/10.2785/537327.

Gangemi, Aldo. 2005. "Ontology Design Patterns for Semantic Web Content." In *The Semantic Web – ISWC 2005*, edited by Yolanda Gil, Enrico Motta, V. Richard Benjamins, and Mark A. Musen, 262–76. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. https://doi.org/10.1007/11574620_21.

Gartner, Inc. 2024. *2024 Chief Data and Analytics Officer (CDAO) Survey: Talent, Budgets, and Priorities*. Gartner Research. https://www.gartner.com/en/insights/cdao.

Goldenfein, J., and D. Hunter. n.d.

Herremans, Dorien, Bob L. Sturm, et al. 2025. *Responsible AI Music (RAIM) Framework*. arXiv preprint arXiv:2503.18814. https://arxiv.org/abs/2503.18814.

Hull, Geoffrey P., Thomas W. Hutchison, Richard Strasser, and Geoffrey P. Hull. 2011. *The Music Business and Recording Industry Delivering Music in the 21st Century*. New York: Routledge.

Huyer, Esther, and Laura van Knippenberg. 2020. *The Economic Impact of Open Data. Opportunities for Value Creation in Europe*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2830/63132.

International ISRC Registration Authority. 2021. *International Standard Recording Code (ISRC) Handbook. 4th Edition*. London: United Kingdom: International ISRC Registration Authority. https://www.ifpi.org/wp-content/uploads/2021/02/ISRC_Handbook.pdf.

ISO. 2012. *International Standard Musical Work Code (ISNI). ISO 27729:2012*. 27729. Version 2012. International Organization for Standardization. https://www.iso.org/standard/44292.html.

———. 2013. *ISO 17369:2013(en) Statistical Data and Metadata Exchange (SDMX)*. London:United Kingdom: International Organization for Standardization. https://www.iso.org/obp/ui/en/#iso:std:iso:17369:ed-1:v1:en.

———. 2017a. *ISO/IEC 19941:2017(en), Information Technology — Cloud Computing — Interoperability and Portability*. London:United Kingdom: International Organization for Standardization. https://www.iso.org/obp/ui/#iso:std:iso-iec:19941:ed-1:v1:en.

———. 2017b. *ISO/IEC 5127:2017(en), Information and Documentation — Foundation and Vocabulary*. London:United Kingdom: International Organization for Standardization. https://www.iso.org/obp/ui/en/#iso:std:iso:5127:ed-2:v1:en.

———. 2019a. *International Standard Recording Code (ISRC). ISO 3901:2019*. 3901. Version 2019. International Organization for Standardization. https://www.iso.org/standard/64817.html.

———. 2019b. *ISO/IEC 20546:2019 Information Technology — Big Data — Overview and Vocabulary*. London:United Kingdom: International Standards Organisation. https://www.iso.org/obp/ui/en/#iso:std:iso-iec:20546:ed-1:v1:en.

———. 2020. *ISO/IEC 22624:2020(en), Information Technology — Cloud Computing — Taxonomy Based Data Handling for Cloud Services*. London:United Kingdom: International Organization for Standardization. https://www.iso.org/obp/ui/en/#iso:std:iso-iec:22624:ed-1:v1:en.

———. 2021. *ISO 10957:2021 - Information and Documentation — International Standard Music Number (ISMN)*. 10957. Version 2021. International Organization for Standardization. https://www.iso.org/standard/83122.html.

———. 2022. *International Standard Musical Work Code (ISWC). ISO 15707:2022*. 15707. Version 2022. International Organization for Standardization. https://www.iso.org/standard/83125.html.

———. 2023a. *ISO/IEC 11179-1:2023(en), Information Technology — Metadata Registries (MDR) — Part 1: Framework*. London:United Kingdom: International Organization for Standardization. https://www.iso.org/obp/ui/en/#iso:std:iso-iec:11179:-1:ed-4:v1:en.

———. 2023b. *ISO/IEC 2382:2015(en), Information Technology — Vocabulary*. London:United Kingdom: International Standards Organisation. https://www.iso.org/obp/ui/en/#iso:std:iso-iec:2382:ed-1:v2:en.

ISO/IEC. 2022. *Information Technology — Artificial Intelligence — Concepts and Terminology. ISO/IEC 22989:2022*. 22989. Version 2022. International Organization for Standardization. https://www.iso.org/standard/74296.html.

———. 2023. *Information Technology — Artificial Intelligence — Management System. ISO/IEC 42001:2023*. 42001. Version 2023. International Organization for Standardization. https://www.iso.org/standard/81230.html.

Jachimczyk, Bartosz, and Katarzyna Nowak. 2024. "Exploring Open Government Data (OGD) Portals: A Usability Study Across the EU." *arXiv Preprint*, 2024. https://arxiv.org/abs/2406.08774.

Jegan, Robin, Leon Fruth, Tobias Gradl, and Andreas Henrich. 2023. "Integrating Access to Authority Data for Improved Interoperability of Research Data in the Digital Humanities." *Datenbanksysteme Für Business, Technologie Und Web (BTW 2023)*, 2023, 829–36. https://doi.org/10.18420/BTW2023-54.

Kević, Kristijan, Ana Kuveždić Divjak, and Frederika Welle Donker. 2023. "Benchmarking Geospatial High-Value Data Openness Using GODI Plus Methodology: A Regional Level Case Study." *Geographies* 12 (6): 222. https://doi.org/10.3390/geographies12060222.

Leurdijk, Adnra, and Nieuwenhuis Ottilie. 2012. *Statistical, Ecosystems and Competitiveness Analysis of the Media and Content Industries. The Music Industry*. 25277 EN. Edited by Jean Paul Simon. Luxembourg: Publications Office of the European Union, 2012: Joint Research Centre Institute for Prospective Technological Studies (IPTS). https://doi.org/10.2791/796.

Leurdijk, Andra, and Nieuwenhuis Ottilie. 2012. *Statistical, Ecosystems and Competitive-*

ness Analysis of the Media and Content Industries. The Music Industry. 25277 EN. Edited by Jean Paul Simon. Luxembourg: Publications Office of the European Union, 2012: Joint Research Centre Institute for Prospective Technological Studies (IPTS). http://ftp.jrc.es/EURdoc/JRC69816.pdf.

Magnus, Bart, and Olivier Van D'huynslager. 2021. "Podiumkunstendata Op Wikidata: De Stap Naar Echte Linked Open Data." Kunstenpunt / Flanders Arts Institute, February 11. https://www.kunsten.be/nu-in-de-kunsten/podiumkunstendata-op-wikidata-de-stap-naar-echte-linked-open-data/.

Mechanical Licensing Collective. 2021. *2021 Annual Report*. https://www.themlc.com/annual-report-2021.

Mikš, Tomáš. 2025. *OpenMusE: Towards a Sustainable Licensing Market for AI Use of Protected Works*. April 29–30, 2025. Vilnius, Lithuania: Slovak Performing; Mechanical Rights Society (SOZA); OpenMusE Consortium; Presentation at the CISAC European Committee Meeting. https://www.openmuse.eu/wp-content/uploads/2025/05/20250427_CISAC_EC_2025_OpenMusE_final.pdf.

Mikš, Tomáš, and Dániel Antal. 2025. *Open Access Music Dataspaces – Open Music Observatory*. Open Music Observatory. https://doi.org/10.5281/zenodo.17669739.

Milosic, Klementina. 2015. "The Failure of the Global Repertoire Database (GRD)." *Hypebot*, August 2015. https://www.hypebot.com/hypebot/2015/08/the-failure-of-the-global-repertoire-database-effort-draft.html.

Ministerstvo kultúry SR, and Open Music Europe. 2023. *Memorandum o porozumení o využití výsledkov analýz otvorených politík v kontexte slovenského kultúrneho a kreatívneho priemyslu a sektorových verejných politík v spolupráci s konzorciom pre výskum a inovácie s názvom OpenMuse. [Memorandum of Understanding on utilizing the Open Policy Analysis results of the OpenMuse Research and Innovation Consortium in the context of Slovak cultural and creative industries and sectors' public policies]*. https://www.crz.gov.sk/zmluva/7645338/.

MIT Sloan School of Management. 2025. *Project NANDA: Enterprise Generative AI Value Creation*. Massachusetts Institute of Technology. https://mitsloan.mit.edu.

Music Moves Europe. 2024. *Music Ecosystem 2025: Study on the Music Ecosystem*. Publications Office of the European Union. Luxembourg: European Commission, Directorate-General for Education, Youth, Sport; Culture. https://doi.org/10.2766/95340.

Nakos, Basil, and Lazaros Tsoulos. 2022. "Web-Based Nautical Charts Automated Compilation from Open Hydrospatial Data." *Journal of Navigation* 75 (6): 1–19. https://doi.org/10.1017/S0373463322000489.

Open Music Europe Consortium. 2025. *Policy Brief: An Open, Scalable Data-to-Policy Pipeline for European Music Ecosystems*. EU Horizon Europe Deliverable D5.7. Open Music Europe Consortium. https://openmuse.eu/.

OpenRefine Community. 2021. *OpenRefine Reconciliation API Standard*. https://reconciliation-api.github.io/specs/latest/.

Partanen, Niko, Philippe Rixhon, Karīna Bandere, Jānis Ziediņš, Pawan Kumar Dutt, Matīss Bolšteins, Matias Frosterus, Mona Lehtinen, Inta Miklūna-Žukeviča, Deniss Ozerskis, Päivi Maria Pihlaja, Jogita Sauka, Katerina Sornova, and Aija Uzula. 2025. *Interoperable, Trustworthy, and Machine-Readable Copyright Data in the AI Era: Report of the CITF First Project*. Publications of the Ministry of Education and Culture, Finland 2025:23. Helsinki: Ministry of Education; Culture, Finland; National Library of Finland; National Library of Latvia; Culture Information Systems Centre (Latvia); Tallinn University of Technology (Estonia); Valunode OÜ. https://julkaisut.valtioneuvosto.fi/.

Partanen, Niko, Philippe Rixhon, Karīna Bandere, Jānis Ziediņš, Pawan Kumar Dutt, Matīss Bolšteins, Matias Frosterus, Mona Lehtinen, Inta Miklūna-Žukeviča, Deniss Ozerskis, Päivi Maria Pihlaja, Jogita Sauka, Katerina Sornova, and Aija Uzu. 2025. *Interoperable, Trustworthy, and Machine-Readable Copyright Data in the AI Era: Report of the CITF First Project*. Research report. Helsinki: Ministry of Education; Culture. https://urn.fi/URN:ISBN:978-952-415-143-6.

Paskin, Norman. 2006. "Identifier Interoperability: A Report on Two Recent ISO Activities." *D-Lib Magazine* 12 (4): 1–20. https://doi.org/10.1045/april2006-paskin.

Pomerantz, Jeffrey. 2015. *Metadata*. The MIT Press Essential Knowledge Series. Cambridge, MA, USA: MIT Press.

Project, CEDAR. 2023. "A Hitchhiker's Guide to High Value Datasets." https://cedar-heu-project.eu/articles/hitchikers-guide-high-value-datasets.

PRS for Music. 2023. "PRS for Music Expands Pioneering Nexus Programme." September 6. https://www.prsformusic.com/press/2023/prs-for-music-expands-pioneering-nexus-programme.

PwC. 2023. *Digital IQ 2023: Driving ROI on Digital Investments*. PricewaterhouseCoopers International Limited. https://www.pwc.com/gx/en/industries/technology/digital-iq-survey.html.

———. 2024. *27th Annual Global CEO Survey*. PricewaterhouseCoopers International Limited. https://www.pwc.com/gx/en/ceo-agenda/ceosurvey/2024.html.

Quine, Willard Van Orman. 1968. "Ontological Relativity." *The Journal of Philosophy* 65 (7): 185–212. https://doi.org/10.2307/2024305.

Sardo, Lucia, and Carlo Bianchini. 2022. "Wikidata: A New Perspective Towards Universal Bibliographic Control." *JLIS.it : Italian Journal of Library and Information Science* 13 (1): 291–311. https://doi.org/10.4403/jlis.it-12725.

Schnurr, Daniel. 2021. *Open Government Data in Digital Markets: Effects on Innovation, Competition and Societal Benefits*. SSRN Working Paper. https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=3743648.

SEMIC Support Centre. 2023. *Wikidata and Wikibase — SEMIC Support Centre*. https://interoperable-europe.ec.europa.eu/collection/semic-support-centre/wikidata-and-wikibase.

Senftleben, Martin, Thomas Margoni, Joost Poort, Kacper Szkalej, and Etienne Valk. 2024. *Policy Brief 1: Music Metadata Mainstreaming and EU Law*. EU Horizon Europe Deliverable D5.6. OpenMusE Consortium. https://www.openmuse.eu/.

Simoni, Marco U., Kristin A. Aasly, and Frode Schjøth. 2021. *MINERAL Intelligence for Europe (Mintell4EU) – Case Study Overview*. GeoERA. https://geoera.eu/wp-content/uploads/2021/10/D4.1-Mintell4EU-Case-Study-Overview.pdf.

Stallmann, Claudia, Koen Deneckere, Ruben Verborgh, et al. 2023. "MetaBelgica Project: A Linked Data Infrastructure Between Federal Scientific Institutes in Belgium." *Proceedings of the 19th Extended Semantic Web Conference (ESWC 2023)* (Cham), 2023. https://doi.org/10.1007/978-3-031-33455-9_24.

Teosto. 2024. "New ISNI Identifier Creates Better Opportunities for International Author Identification." March 12. https://www.teosto.fi/en/new-isni-identifier-creates-better-opportunities-for-international-author-identification/.

Varghese, Jacob. 2024. "Beyond the Metadata: How Can We Solve the Black Box Royalty Mystery?" Noctil, August 14. https://independentmusicinsider.com/editorial-articles/4820/.

Virág, Barnabás. 2024. "Our Library's Music Collection in the Era of Streaming Services." *Canadian Journal of Information and Library Science / La Revue Canadienne Des Sciences de l'information Et de Bibliothéconomie* 47 (2): 175–87. https://doi.org/10.5206/cjils-rcsib.v47i2.17436.

W3C. 2013a. *PROV-o: The PROV Ontology*. Edited by Satya AND McGuinness Lebo Timothy AND Sahoo. W3C. https://www.w3.org/TR/prov-o/.

———. 2013b. *PROV-Overview: An Overview of the PROV Family of Documents*. Edited by Paolo Moreau Luc AND Missier. W3C. https://www.w3.org/TR/prov-overview/.

World Intellectual Property Organization (WIPO). 2023. "Project Nexus: A Data Matching Project of PRS for Music." April 19. https://www.wipo.int/edocs/mdocs/mdocs/

en/wipo_webinar_cr_2023_6/wipo_webinar_cr_2023_6_presentation.pdf.