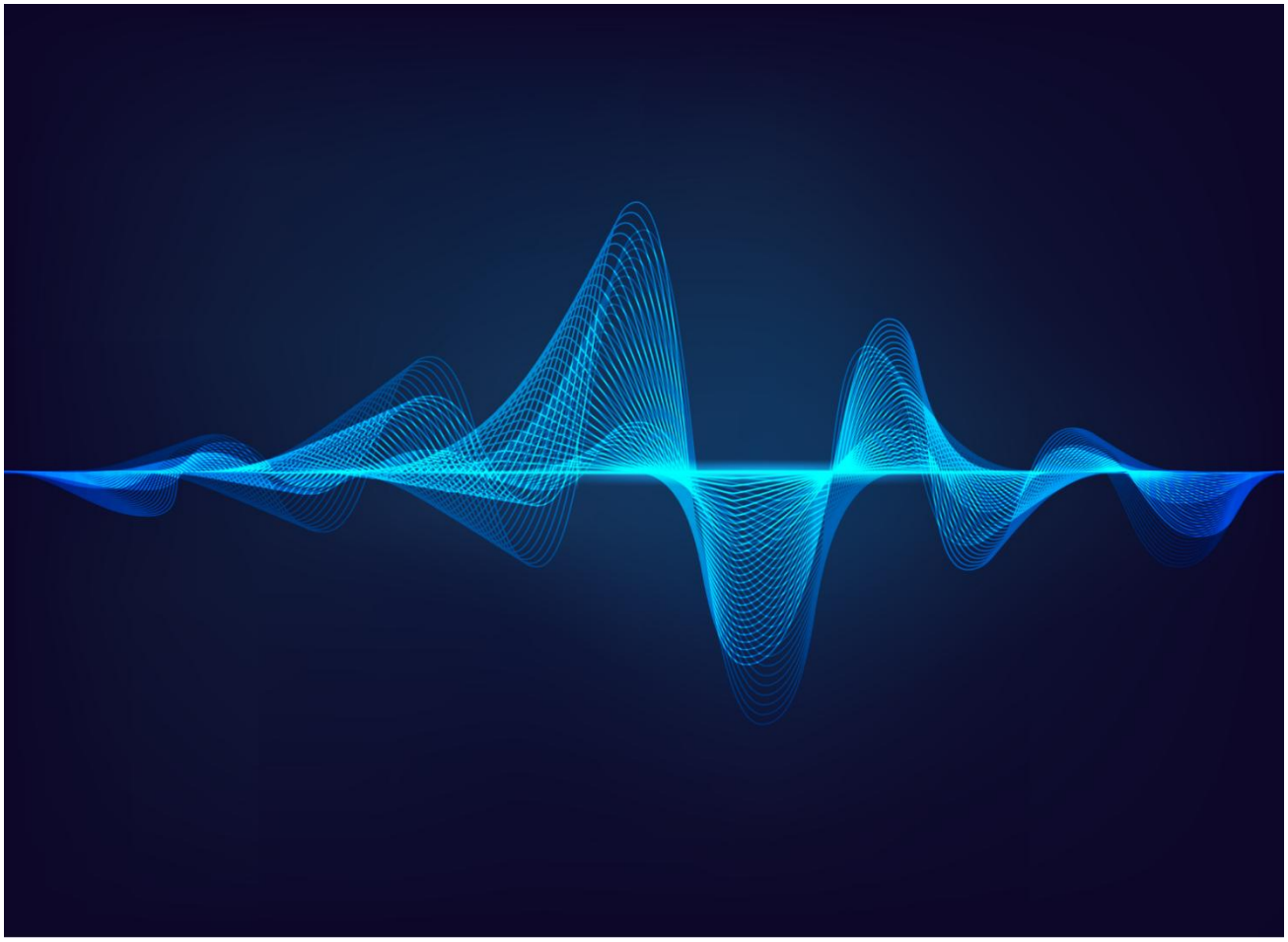


D2.3 – Report on the pilot study on music diversity

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
Reference:	101095295
Type:	Research and Innovation Action
Program:	Horizon Europe
Start:	1st January 2023
Duration:	36 months
Website:	https://www.openmuse.eu/
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)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Deliverable Factsheet

Number:	**D2.3**
Title:	**Report on the pilot study on music diversity and circulation**
Lead beneficiary:	SOZA
Work package:	WP2 Music Diversity and Circulation
Task:	Pilot study on music diversity and circulation
Dissemination level:	Public
Submission date:	29.12.2025
Main author(s):	Antal, Dániel (Reprex)
Contributor(s):	Demčák, Richard (SOZA), Jankovič, Marián (SOZA), Mester, Anna Márta (REPREX), Mikš, Tomaš (SOZA), Nagy, Mihály (REPREX), Semaňáková, Dominika (SOZA), Žilková, Anna (Slovak Music Centre), Zagyva, Natália (Hungarian Heritage House), Záhonyi, Lászó (House of Music)
Quality check:	Dr. Caterina Sganga (SSSA), Rosina Petrova (MusicAutor)

Document history:

Revision	Date	Main modification	Author
0.1	12/17/2025	First summary draft	Daniel Antal
0.4	12/22/2025	Quality check	SSSA, MUSICAUTOR
0.5	12/26/2025	Finalised draft	Daniel Antal
1.0	12/29/2025	Submitted to EC	SINUS

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Disclaimer of Warranties

This project has received funding from the European Union's Horizon Europe, research and innovation programme, under Grant Agreement No. 101095295*.*

This document has been prepared by OpenMusE project partners as an account of work carried out within the framework of the EC-GA contract no 101095295.

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 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 Project Coordinator or any representative of a signatory party of the OpenMusE Project Consortium Agreement, has been advised of the possibility of such damages) resulting from your selection or use of this document or any information, apparatus, method, process, or similar item disclosed in this document.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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)

AI & Data Terms

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

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

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

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)

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)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

identifier authority: trusted organisation responsible for issuing or validating identifiers for agents or assets. Includes registration agencies, national libraries, CMOs, or similar bodies. (Partanen et al. 2025, 85)

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.

holdings: totality of documents in the custody of an information and documentation organization (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)

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 et al. 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).

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)

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)

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 et al. 2025, 33)

Statistical terms

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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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

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

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 et al. 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 et al. 2025, 86)

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

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;

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))

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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.

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.

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.

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

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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.

IFPI: IFPI is the voice of the recording industry worldwide. IFPI and its network of National Groups and Local Associations represents the interests of some 8,000 members across the globe.

REPREX: Data management and trustworthy AI researh and innovation oriented SME in the Netherlands.

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 or MC: Music Centre Slovakia is a music organisation with a mission to promote Slovak contemporary music.

Other abbreviations

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.

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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Contents

Glossary 1
Music terms 1
AI & Data Terms 2
Statistical terms 3
Registers, authorities, standards and identifiers 4
Organisations 5
Other abbreviations 6
Executive Summary 10
1 Introduction 13
1.1 Findings of the Feasibility Phase 14
2 Policy Context and Relationship to the Green Paper 17
2.1 European Cultural Policy Framework: Compass, Data Hub, and Music as a Forerunner 17
2.2 Analytical Foundations: From the CITF Background Study to the OpenMusE Policy Deliverables and Green Paper 18
2.3 The OpenMusE Policy Work in a Nutshell: Section Overview 19
2.4 From Analysis to Implementation: CITF First Project (Books) and Music as a Parallel Case 21
2.5 Diversity and Circulation as the Other Side of the Economic Equation 22
3 From Policy to Implementation: Governance, Architecture, and Semantic Design 24
3.1.1 Governance as an Implementation Method 24
3.1.2 Architectural Commitments under Policy Constraints 27
3.1.3 Semantic Design as Policy Translation 28
4 Methodological Considerations 32
4.1.1 Data Modelling and Representation 32
4.1.2 Interoperability Mechanics 34
4.1.3 Technical Publication and Scalability 36
4.1.4 Cross-Context Validation 37
5 Diversity Goals Without Exposure of Sensitive Data 39
5.1.1 The Diversity & Circulation Aspects of European Music Policy 40
5.2 Application Layers and System Architecture 44
5.2.1 Public-private governance and data protection 45
5.2.2 Public Access Layer 47
6 Case Study: Computing Local Content Indicators from Official Charts 50
6.1.1 Case study: Slovak charts as an empirical testbed 50

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

6.1.2 Aggregation and observatory-level representation 53
6.1.3 Interpretation limits and policy relevance 53
7 Discussion and Outlook 55
8 References 57
A Data Used 62
A.1 Personal Data 62
A.1.1 Location and nationality 63
A.1.2 Repertoire Data 66
A.1.3 Sound recordings 66
A.1.4 Recording Title 66
A.1.5 Identifiers 68
A.1.6 Release 70
B Queries 73
C Governance 74
C.1 Data Sharing Space Memorandum of Understanding 74
C.1.1 Visibility 75
C.1.2 Information Management 75
C.1.3 Authority Control and Mutual Cooperation 75
C.2 Proportionality test for accessing the controller's legitimate interest 77
C.2.1 Introduction 77
C.2.2 Identification of legitimate interest 78
C.2.3 Has the Controller taken additional security measures in connection with the processing of personal data? 82
C.2.4 Evaluation of the proportionality test 82
D Annex D: A Green Paper on AI, Data Governance, and Metadata Policies for Europe's Music Ecosystem 84
D.1 Introduction 84
D.2 Policy context and problem map 91
D.2.1 Three structural pressures 94
D.2.2 National and European pilots as anchors 95
The Slovak Comprehensive Music Database (SKCMDb) 96
D.2.3 Quest for efficiency 100
D.2.4 Potential solutions 103
D.3 Fixing Music Data at the Source 105

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

D.3.1 Discussion 105
D.3.2 Policy proposals 114
D.4 Open Music Observatory: Building a Shared Music Data Space 120
D.4.1 Discussion 121
D.4.2 Policy Proposals 129
D.5 AI that Works for Music, Not Against It 136
D.5.1 Discussion 139
D.5.2 Policy Proposals: Aligning AI with Governance and Value Creation 145
Lowering Documentation Barriers 147
Observatory: European = Open 148
The Open Music Observatory as a Collective Guardrail 148
D.6 What Europe Should Do Next for Music Data & AI 149
D.7 Sources & Further Reading 151

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Executive Summary

Across Europe and other advanced music markets, cultural policy instruments such as local content quotas, diversity targets, and equality pledges have long played an important role in sustaining domestic cultural ecosystems. These instruments reflect a structural asymmetry between global cultural markets and national or regional creative production. While English-language popular music circulates globally with few barriers, music produced in smaller linguistic and cultural markets—such as Slovak music—is consumed predominantly within national borders or among relatively small diaspora communities. For this reason, more than eighty countries worldwide apply some form of preferential treatment to local music and audiovisual production, most commonly through minimum usage shares for public broadcasters and, in many cases, commercial, regional, or community radio stations as well.

In recent European policy debates, these long-standing instruments have acquired renewed significance in the context of the **EU Culture Compass** and the emerging **EU Culture Data Hub**, which aim to support cultural policy with timely, comparable, and evidence-based indicators. Within this framework, diversity and circulation metrics are no longer peripheral reporting tools, but core inputs for monitoring cultural ecosystems, assessing policy impact, and informing strategic decision-making at European level. Yet many of the indicators envisioned in these initiatives—particularly those related to diversity, local representation, and cross-border circulation—have historically been difficult or impossible to measure in a consistent and legally robust manner.

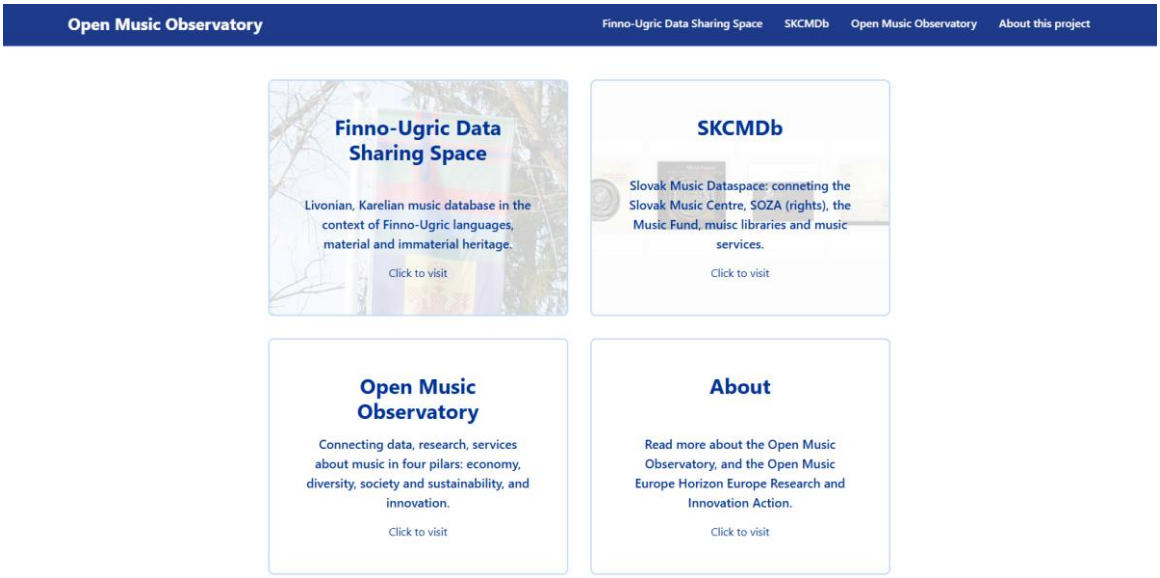
Despite their long history, diversity- and local-content policies have repeatedly struggled at the point of implementation. Broadcasters, festival organisers, and other cultural intermediaries are formally expected to meet quantitative targets, yet they often lack reliable, auditable, and ethically acceptable tools to assess compliance in practice. In the absence of such tools, implementation has frequently relied on informal editorial heuristics, institutional memory, or conservative programming strategies that privilege already well-known repertoire. These practices, while understandable, tend to undermine the intended goals of cultural diversity, renewal, and inclusion, and limit the usefulness of resulting data for European-level aggregation.

The challenge is compounded by scale. Even before the introduction of the General Data Protection Regulation (GDPR), the rapid expansion of global music distribution platforms had already rendered traditional editorial knowledge insufficient. In the 1990s, a highly experienced radio editor or festival curator in a smaller European country might plausibly have maintained working familiarity with a few tens of thousands of recordings circulating through physical distribution channels. Today, global digital platforms provide access to well over one hundred million recordings. No individual editor can meaningfully survey such a repertoire, yet policy objectives increasingly apply across this vastly expanded universe of available content. This scale effect fundamentally alters the nature of categorisation, assessment, and comparability.

These developments expose a structural mismatch between European cultural policy ambitions and the data infrastructures available to support them. The problem is not primarily one of unwillingness or resistance among cultural intermediaries, but of missing coordination mechanisms capable of translating policy criteria into operational, computable indicators that can be aggregated across institutions and borders. At the same time, any viable solution must respect a core principle of European governance: subsidiarity. Cultural production, heritage, and identity are inherently

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

situated, and data infrastructures that feed into European-level instruments such as the Culture Compass must therefore operate across local, national, and supra-national levels without collapsing these distinctions into a single, centralised system.



https://openmusicobservatory.eu/

The work presented in this paper emerges from this insight. It introduces the *Open Music Observatory (OMO)*—and in particular its *Diversity and Circulation pillar*—as a forerunner and model component of a future *EU Culture Data Hub*. Through the concrete implementation of the *Slovak Comprehensive Music Database* (SKCMDb), developed alongside comparable initiatives such as the *Hungarian Music Database* and community-curated Finno-Ugric music databases, the paper demonstrates how diversity and circulation indicators that were previously considered impractical can be computed reproducibly and responsibly. In doing so, it illustrates how national and community-level data infrastructures can function as live, policy-relevant data sources for European instruments such as the Culture Compass.

This paper proceeds as follows.

- Section 1.1 summarises the findings of the feasibility phase that motivated the work. Section 2 situates the project within the European data governance agenda and outlines, in a concise “policy brief in a nutshell”, the relationship between this implementation work, the Green Paper, and current EU initiatives.
- Section 3 provides the path from policy to implementation with explaining governance, architecture and semantic design under the EU’s data, cultural and music, and diversity policy umbrella.
- Sections Section 4 and Section 5 describe the methodological foundations and the separation of internal evaluation from external disclosure.
- Section 5.1 presents the federated implementation and system architecture.
- Section 6 demonstrates the approach through a case study based on official Slovak music charts.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

This deliverable builds on the conceptual and policy analysis developed in Deliverables D2.2 *Report on Music Diversity and Circulation*[1] and D1.3 *Music Metadata Improvement and EU Law*[2]:

- D2.2 critically examines existing EU and national definitions of music diversity, as well as their limitations and contested aspects. The present document does not seek to redefine or normatively assess those concepts. Instead, it focuses on how such definitions—once adopted in a given regulatory or policy context—can be translated into data structures, system components, and operational workflows. In this sense, the deliverable complements D2.2 by providing an implementation-oriented perspective that enables transparency, comparability, and replication across different national contexts.
- D1.3 advocates the establishment of a comprehensive music metadata infrastructure that improves the visibility and accessibility of the entire European music repertoire in digital and algorithmic environments. It also suggests that existing EU legislation on copyright, fundamental rights, and digital markets and dataspaces provide sufficient justification for the establishment of such an infrastructure. The present document demonstrates how improved metadata schemas—once established and acknowledged—could be implemented technically and operationalised in local legal contexts.

The paper concludes by discussing implications for European cultural policy, data governance, and the future development of the *Open Music Observatory* as a prototype European Music Observatory within the EU Cultural Data Hub, or as a component of the EU Cultural Data Hub.

This deliverable was prepared under a constrained and evolving project context, including changes in timelines, dependencies between work packages, and the availability and maturity of upstream inputs. As a result, certain planned validation, integration, or iterative refinement steps could not be carried out to the originally anticipated extent within the project timeframe.

The work presented here therefore prioritises methodological transparency, reproducibility, and clear documentation of implemented components over of cross-work-package integration. Where relevant, limitations are explicitly noted, and all materials are made available to support future extension or reuse beyond the project's formal end date. Additional refinement and replication work was discussed by the partners involved but could not be accommodated within the project timeframe.

The empirical work, implementation activities, and partner contributions documented in this deliverable were carried out in Slovakia and Hungary, where the innovation was developed and implemented. Latvian and Estonian partners contributed data and domain inputs used for methodological validation but did not undertake replication or local implementation of the innovation developed in Slovakia.

We reprint the current version of our policy paper developed in consultation with pan-European, national and regional stakeholders in Annex D, because it was not part of other deliverables, and is relevant to this one.

[1] *D2.2 Report on Music Diversity and Circulation* (Sganga, Contardi, and Cernigliaro 2025)
[2] *D1.3 Music Metadata Improvement and EU Law* (Izyumenko, Margoni, Poort, Senftleben, Szkalej, and Valk 2025)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

1 Introduction

Across Europe and other advanced music markets, cultural policy instruments such as local content quotas, diversity targets, and equality pledges are widely used to sustain domestic cultural ecosystems. As discussed in the OpenMusE Deliverables D2.2 *Report on Music Diversity and Circulation* and D1.3 *Music Metadata Improvement and EU Law*, and the OpenMusE *Green Paper on AI, Data Governance, and Metadata Policies for Europe's Music Ecosystem*, these instruments have repeatedly struggled at the point of implementation, particularly under conditions of platform scale and data protection constraints. Broadcasters, festival organisers, and other cultural intermediaries are expected to meet quantitative targets, yet they often lack reliable, auditable, and legally safe tools to assess compliance in practice.

This paper responds to that policy analysis by presenting the *Slovak Comprehensive Music Database* (SKCMDb) as an implementation-level pilot within the Open Music Observatory. SKCMDb operationalises diversity- and locality-related policy criteria through a federated, governance-by-design data infrastructure. Rather than exposing personal attributes of composers or performers, the system separates internal evaluation from external disclosure, expressing policy-relevant outcomes as properties of cultural artefacts. This approach enables the computation of eligibility and diversity indicators while respecting the General Data Protection Regulation, non-discrimination principles, and institutional competences.

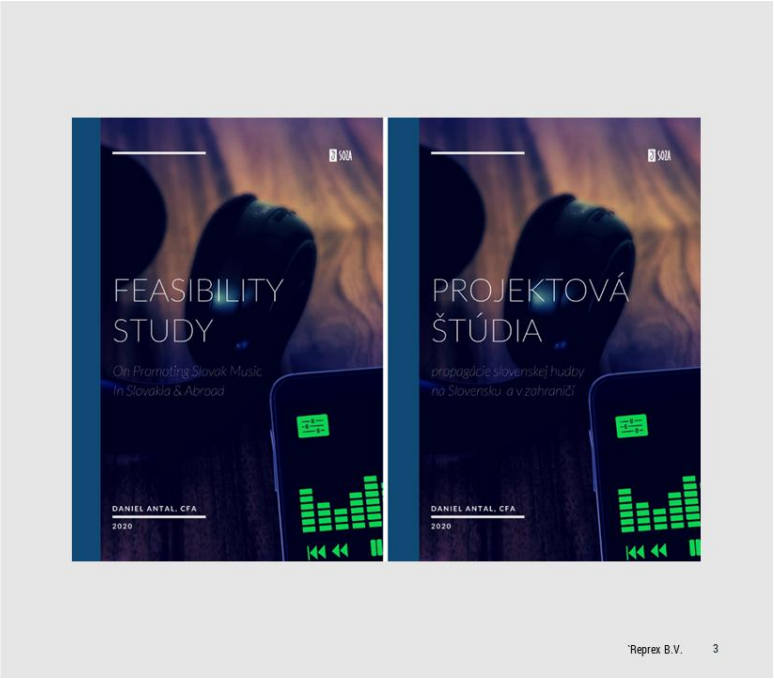
The paper situates SKCMDb within a broader ecosystem of interoperable modules, including the *Hungarian Music Database* and community-curated Livonian and Karelian music databases, illustrating how the subsidiarity principle articulated in the Green Paper can be implemented across national, community, and supra-national levels. Methodologically, the work combines linked-data modelling, modular ontologies, and provenance-aware workflows with lightweight publication strategies.

An empirical case study based on 156 weeks of official Slovak music charts demonstrates how local content indicators can be computed reproducibly and transparently using authoritative reference data, without monitoring individual broadcasters. The results illustrate the feasibility of diversity-aware cultural policy infrastructures envisioned in the Green Paper and support voluntary compliance, editorial decision-making, and comparative analysis. The paper concludes by outlining implications for European data governance, cultural policy, and the future development of the Open Music Observatory as a component of the EU Culture Data Hub.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

1.1 Findings of the Feasibility Phase

The *Feasibility Study On Promoting Slovak Music In Slovakia & Abroad*[3] identified a persistent and structural gap between the normative ambitions of cultural policy and the operational realities of music dissemination. While local content quotas and diversity objectives are typically articulated in legal or institutional terms, they are rarely accompanied by a technical or methodological framework that would allow their consistent application in everyday editorial practice. This gap was not merely practical, but conceptual: policy definitions often relied on categories that could not be coherently mapped onto the objects actually circulated by media platforms.



A central finding concerned the mismatch between **persons** and **cultural artefacts**. Radio stations, streaming platforms, and broadcasters do not distribute composers or performers as such; they distribute recorded fixations of performances, that is, sound recordings. These recordings are cultural artefacts that do not themselves possess personal characteristics. Yet many quota systems implicitly assume that attributes of persons—such as nationality, age, gender identity, ethnicity, or place of residence—can be directly ascribed to the recordings through which music is communicated to the public. This assumption breaks down both conceptually and legally.

Modern European cultural policy does not follow nineteenth-century ethnomusicological traditions that attempted to define what makes music intrinsically "Slovak", "Hungarian", or otherwise national. Contemporary quota systems instead operate through **agent-based criteria**: a recording may qualify as local or national if it is a performance of a musical work composed by a person born in, residing in, or otherwise connected to a given country; in other cases, language of lyrics or place of performance may be decisive. In Slovakia, for example, vocal recordings performed in the Slovak language may qualify irrespective of the performer's place of birth, recording location, or ethnicity. A recording of

[3] The study was published in English as *Feasibility Study On Promoting Slovak Music In Slovakia & Abroad* and *Projektová štúdia propagácie slovenskej hudby na Slovensku a v zahraničí* in Slovak (Daniel Antal 2020a, 2020b).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Slovak folk music performed in Hungary by members of the Slovak ethnolinguistic minority may therefore legitimately count toward both Hungarian and Slovak quota systems.

At the same time, European quota systems are constrained by fundamental legal principles, including the prohibition of discrimination based on ethnicity, nationality, or other protected characteristics. Cultural policy may promote linguistic diversity, local cultural ecosystems, or underrepresented groups, but it cannot lawfully categorise or exclude works on the basis of ethnic identity. This legal constraint is not a weakness of European quota systems; rather, it reflects a normative commitment to pluralism and equal treatment. It also opens space for precisely the kinds of cross-border and minority cases illustrated by Slovak folk music performed in Hungary or Hungarian-language folk traditions maintained by Csángó communities in Romania.

The feasibility work was strongly influenced by insights from **intersectional data feminism**, particularly the recognition that data-related inequalities do not arise along a single axis, but through the accumulation and interaction of multiple structural asymmetries. In the global music industry, the visibility advantage of an American male star cannot be meaningfully compared to the position of a Slovak female composer, nor to that of a Csángó woman singing in a distinct Hungarian dialect in Romania. These cases sit at the intersection of linguistic marginality, gender imbalance, ethnic minority status, and unequal access to global distribution infrastructures. Treating any one of these dimensions in isolation risks reproducing, rather than mitigating, existing inequalities.

Importantly, this intersectional perspective also revealed that diversity-oriented policies must not be framed as opposing “majority” and “minority” interests in simplistic ways. Most European national music ecosystems—regardless of internal gender or ethnic hierarchies—are themselves structurally marginal within the global music economy. From this perspective, the protection of cultural diversity includes not only historically marginalised groups, but also majority male performers operating within small national markets whose repertoire competes with overwhelmingly dominant global catalogues. A European diversity framework must therefore protect all participants in fragile cultural ecosystems, while remaining attentive to internal power asymmetries.

These observations further complicate the already problematic linkage between personal attributes and recordings. A single sound recording may involve multiple composers, lyricists, performers, arrangers, and producers, each situated differently across axes of gender, language, ethnicity, nationality, and residence. Group identities are fluid, and personal circumstances change over time. Attempting to stabilise these attributes at the level of the recording risks both conceptual incoherence and ethical overreach, particularly under data protection law.

The feasibility study therefore concluded that sustainable implementation requires a **decoupling of internal evaluation from external disclosure**, informed by an intersectional understanding of power and inequality. Rather than exposing personal attributes or identity categories, a system must be able to evaluate eligibility internally—within a legally justified, institutionally competent, and ethically governed environment—and expose only the outcome of that evaluation at the level of the artefact. In this model, a recording may be labelled as meeting or not meeting a quota requirement, without revealing which contributor, which protected characteristic, or which intersection of attributes triggered eligibility.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

This conclusion marked a decisive reframing of the problem. Local content and diversity policies could no longer be treated as matters of editorial judgement alone, nor as enforcement exercises aimed at individuals. Instead, they emerged as **intersectional information design problems**: how to construct reference infrastructures that translate policy rules into computable properties of cultural artefacts, while respecting non-discrimination, data protection, institutional competence, and the complex social realities of European cultural production. Crucially, this reframing also opened a path toward indicators that can be aggregated, compared, and reused beyond a single national context, providing the conceptual and methodological preconditions for diversity and circulation metrics capable of informing European-level data infrastructure.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

2 Policy Context and Relationship to the Green Paper

This section situates the present implementation work within the broader European cultural and data policy landscape. It clarifies how the *Open Music Observatory* and its Slovak, Hungarian, Polish, Finno-Ugric subnational and pan-European central modules relate to ongoing EU initiatives, including the *Culture Compass for Europe*, the planned *EU Culture Data Hub*, and recent work on copyright data governance. Rather than restating the policy analysis developed in OpenMusE Deliverables D2.2 *Report on Music Diversity and Circulation* and D1.3 *Music Metadata Improvement and EU Law*, and the OpenMusE *Green Paper*, the section explains how this paper should be read as a complementary, implementation-level contribution that operationalises key policy insights through concrete data infrastructures.

2.1 European Cultural Policy Framework: Compass, Data Hub, and Music as a Forerunner

This paper takes the European Union's emerging cultural policy framework as a given. The *Culture Compass for Europe*, the planned *EU Culture Data Hub*, and recent European Parliament resolutions establish shared political destinations: strengthening cultural and linguistic diversity, improving the visibility and circulation of European cultural content, and enabling evidence-based policy through trustworthy data infrastructures[4]. These initiatives do not prescribe sector-specific implementations, but they clearly signal the need for operational data flows capable of supporting observability at European scale.



> *"Recognising the EU's supporting role in the field of culture, this declaration affirms shared political commitments and responsibilities of the European Parliament, the Council of the European Union and the European Commission - within their respective competencies and in full compliance with EU law - based on the key directions set out in the Commission's 2025 Culture Compass for Europe."*

Within this framework, music constitutes a particularly suitable forerunner domain. Music is highly digitised, deeply platform-mediated, and characterised by complex cross-border circulation. At the same time, it combines strong public-interest objectives—such as diversity, cultural sovereignty, and fair remuneration—with mature sectoral institutions, including collective management organisations, national libraries, and public broadcasters. These characteristics make music an early stress test for the kinds of data infrastructures envisioned by the Culture Compass and the EU Culture Data Hub.

[4] *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) (European Parliament and Council 2022)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

The Open Music Observatory is positioned as a sectoral implementation that can contribute empirical evidence to this broader European architecture. Rather than functioning as a central repository, it operates as a federated observability layer capable of aggregating indicators from national and community-level data sources while respecting subsidiarity and institutional competence. In this sense, the diversity and circulation pillar documented in this paper illustrates how live, policy-relevant indicators can be generated and, in principle, fed into European-level instruments without centralising data or authority.

By demonstrating how previously difficult or impossible-to-measure aspects of music diversity and circulation can be made observable in practice, the music case provides early implementation feedback for the *EU Culture Data Hub* and the *Culture Compass*. It shows how sectoral data spaces can supply continuous, auditable inputs to European cultural policy while remaining grounded in existing institutional ecosystems. The music implementation presented here should therefore be read as a forerunner contribution—informative for other cultural domains, but not prescriptive beyond its sectoral scope.

2.2 Analytical Foundations: From the CITF Background Study to the OpenMusE Policy Deliverables and Green Paper

The analytical foundations of this work build on two Commission-commissioned studies that approach the same structural problem from different entry points: the European Commission *Study on copyright and new technologies – Copyright data management and artificial intelligence* [5], OpenMusE D2.2 *Report on Music Diversity and Circulation,* OpenMusE D1.3 *Music Metadata Improvement and EU Law*, and the OpenMusE *Green Paper on AI, Data Governance, and Metadata Policies for Europe's Music Ecosystem*. While situated in different policy contexts, these studies draw on a shared evidence base and structure that evidence in closely comparable ways.

The 2022 study, prepared in the context of copyright and AI policy, identified fragmented metadata, weak authority control, and the absence of trustworthy, machine-readable reference sources as systemic barriers to lawful AI use, licensing, and cross-border reuse. Crucially, it framed these challenges as infrastructural rather than sector-specific, arguing for layered data ecosystems composed of authoritative sources, interoperability and reconciliation mechanisms, and downstream applications. This analytical framing later informed the work of the Copyright Infrastructure Task Force (CITF) and its first concrete implementation project[6].

The OpenMusE policy deliverables and Green Paper arrives at closely aligned conclusions, but from a different vantage point and at a later point in time. Developed within a large Research and Innovation Action, they are analytical in purpose but explicitly implementation-informed. Their analysis is grounded in several years of practical experience with live data infrastructures in the music sector, including national, community-curated, and federated European-level components. As a result, the OpenMusE policy deliverables and Green Paper do not merely anticipate infrastructural

[5] European Commission. *Study on copyright and new technologies – Copyright data management and artificial intelligence*, 2022 (European Commission et al. 2022).
[6] *Interoperable, Trustworthy, and Machine-Readable Copyright Data in the AI Era: Report of the CITF First Project* (Partanen et al. 2025).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

needs, but reflect on observed constraints, governance trade-offs, and design choices encountered in real deployments.

Specifically, the legal foundations underpinning the governance-by-design approach taken in this deliverable are developed in detail in D1.3 *Music Metadata Improvement and EU Law*, which provides comprehensive analysis of how EU data legislation intersects with copyright law, fundamental rights law, and cultural policy objectives. D1.3 establishes that metadata improvement constitutes a fundamental rights issue—encompassing cultural diversity, linguistic diversity, and artistic freedom—rather than merely a technical or commercial concern. The present deliverable operationalises these legal principles through technical implementation, particularly regarding GDPR compliance through the separation of internal evaluation from external disclosure. Where D1.3 articulates the normative and legal framework, this deliverable demonstrates its practical feasibility.

These endeavours therefore converge on similar policy-relevant insights not because one derives from the other, but because they observe the same underlying evidence. Each recognises that effective policy intervention in the digital cultural domain depends on the availability of authoritative reference data, provenance-aware interoperability layers, and governance arrangements that respect institutional competence and fundamental rights. The difference lies in emphasis and maturity: the 2022 study articulates the problem space and solution logic in a cross-sectoral, anticipatory manner, while the OpenMusE policy deliverables and Green Paper confirm and refine that logic through sector-specific implementation experience in music.

In this sense, the OpenMusE policy deliverables and Green Paper can be understood as playing a role analogous to the CITF background study, while benefiting from later timing and empirical grounding. The present deliverable builds on both analytical foundations. It follows the layered infrastructural logic articulated in the 2022 study and reflected in the OpenMusE policy deliverables and Green Paper, while operating at a level of concreteness comparable to the CITF First Project. The focus on diversity and circulation should therefore be read not as a separate policy agenda, but as one analytically distinct facet of the same data governance challenge that underpins economic sustainability, discoverability, and rights remuneration in the European music ecosystem.

2.3 The OpenMusE Policy Work in a Nutshell: Section Overview

Together, OpenMusE D2.2 *Report on Music Diversity and Circulation,* D1.3 *Music Metadata Improvement and EU Law*, and *Green Paper on AI, Data Governance, and Metadata Policies for Europe's Music Ecosystem* provide a multi-perspectival analysis of existing music data and metadata infrastructures and the enablers for their improvement.

This section focuses on the *Green Paper*, which of the above outputs is the most immediately aligned with the pilot study aims and work plan. The *Green Paper* is structured as a sequence of five analytical sections. Each section starts from an existing European policy objective or regulatory context, analyses the corresponding data and infrastructure challenges observed in the music ecosystem, and concludes with concrete, implementation-oriented recommendations.

1. Policy Context and Problem Map↗
   This section situates the European music ecosystem within the European Data Strategy, cultural policy objectives, and AI governance debates. It analyses how policy goals such as cultural diversity, market transparency, and lawful AI use are constrained by fragmented and

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

poorly governed music data. It recommends treating music observability as a data infrastructure problem rather than a purely regulatory one.

2. Fixing Music Data at the Source↗
   This section analyses structural weaknesses in music metadata creation and maintenance, including inconsistent identifiers, weak authority control, and misaligned incentives across the value chain. It recommends strengthening authoritative reference sources and improving data quality at the point of creation as a prerequisite for downstream policy, economic, and AI-related uses.

3. Open Music Observatory: Building a Shared Music Data Space↗ This section analyses how European data-space principles—federation, interoperability, and governance-by-design—can be applied in the music sector. It introduces the Open Music Observatory as a shared, federated data space that connects national, community, and European-level data infrastructures. It recommends a layered architecture that enables comparability and aggregation without centralising control.

4. AI that Works for Music, Not Against It↗ This section analyses the interaction between AI systems and music data, focusing on training, discovery, and automated reuse. It examines risks related to provenance loss, bias amplification, and extractive reuse. It recommends provenance-aware, rights-respecting, and auditable data pipelines as conditions for trustworthy AI in the music ecosystem.

5. What Europe Should Do Next for Music Data & AI↗ The final section synthesises the analytical findings and translates them into actionable directions for European institutions, Member States, and sectoral actors. It recommends prioritising implementation pilots, supporting sectoral data spaces, and using concrete deployments to generate empirical feedback for ongoing policy processes.

The present paper focuses on one such implementation pathway. It follows the same analytical logic as the Green Paper but narrows the scope to diversity and circulation, documenting how policy objectives identified in the Green Paper can be translated into working data infrastructures in practice.

The *Green Paper* draws in part on the Slovak, Hungarian, and Finno-Ugric cases presented here, situating them within a wider European policy discussion. It was prepared in consultation with more than ten national and pan-European stakeholders and followed the *Open Policy Guidelines* as a measure of transparency and quality control[7]. Rather than repeating its findings here, we treat the Green Paper as a complementary policy artefact that reflects back on the kinds of implementation-level lessons documented in this paper and connects them to ongoing European debates on data spaces, AI governance, and cultural policy.

[7] The *Guidelines for Open Policy Analysis* form an actionable and practical set of directives that the OpenMusE consoritum was mandated to use under the Grant Agreement. It can be seen as good implementation framework for *Evidence-based policy making in the European Commission* and *The practice of reproducible research* (BITSS 2019; Open Music Europe 2023; J 2015; Kitzes, Turek, and Deniz 2018).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

2.4 From Analysis to Implementation: CITF First Project (Books) and Music as a Parallel Case

This paper documents an implementation that operates at the same level of maturity as the *CITF First Project* in the book sector. In both cases, earlier analytical work is translated into concrete, working data infrastructures that can be assessed under real legal, institutional, and organisational conditions. Neither implementation seeks to set policy or redefine regulatory frameworks; rather, both test how shared analytical principles can be realised in practice.

Both implementations build on a common layered infrastructural logic. Authoritative reference sources form the foundation; interoperability, reconciliation, and provenance mechanisms enable alignment across systems; and downstream applications make the data usable for policy, licensing, analytics, and AI-related purposes. This three-layer model, articulated in the CITF background study and reflected in the Green Paper, provides a shared analytical basis across domains.

The two implementations diverge not in principle, but in the degree of elaboration required by their respective domains. The *CITF First Project* focuses on the book sector, where rights management and data workflows are comparatively linear and where initial use cases often concern non-commercial or low-complexity material, such as academic theses. In this context, the three-layer model can be operationalised with relatively limited emphasis on organisational interoperability. By contrast, the music sector involves substantially more complex workflows. A single musical artefact may simultaneously intersect with libraries, collective management organisations, publishers, record labels, distributors, and platforms, each using overlapping data elements within distinct organisational and legal contexts. For this reason, the Open Music Observatory implementation explicitly follows the European Interoperability Framework (EIF)[8], placing greater emphasis on the organisational layer that mediates between legal requirements and semantic alignment. This choice is not a departure from the CITF model, but a domain-driven extension that becomes necessary as workflow complexity increases.

Governance arrangements in the two cases are likewise comparable. The *CITF First Project* emerged through voluntary, pioneer cooperation among willing Member States and sectoral institutions, without the creation of new mandates or the reallocation of legal authority. The *Open Music Observatory* follows the same governance logic. In Slovakia, cooperation is formalised through a Memorandum of Understanding that establishes shared principles for data stewardship, transparency, and open policy analysis while fully preserving institutional competences[9]. Replication efforts in Hungary and other regions follow similar voluntary patterns, albeit at different levels of maturity. In both cases, governance-by-design and trust-based cooperation are preconditions for implementation.

[8] European Commission. *European Interoperability Framework – Implementation Strategy*, 2017 (Commission and Digital Services 2017).

[9] *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] (Ministerstvo kultúry SR and Open Music Europe 2023)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

The two projects also exhibit complementary strengths. The *CITF First Project* achieves broader geographical reach at an earlier stage, demonstrating that the model can be extended across multiple jurisdictions. The *Open Music Observatory* implementation prioritises depth over breadth, developing a more layered and workflow-sensitive implementation within a smaller number of territories. This trade-off is deliberate and informative: taken together, the two approaches show that the shared analytical model can scale both horizontally, across countries, and vertically, across increasingly complex organisational environments.

Finally, the focus on diversity and circulation in the music case should be understood as analytically inseparable from economic and rights-related concerns. In data-driven cultural ecosystems, discoverability, circulation, and remuneration depend on the same underlying infrastructures of identification, authority control, and provenance. Music that cannot be reliably identified and made visible within such infrastructures cannot circulate effectively and cannot generate sustainable income for rightsholders. Diversity-oriented indicators and economic indicators therefore represent complementary perspectives on the same infrastructural problem.

On this basis, the following sections describe the methodological and architectural choices through which the music implementation was realised, showing how policy objectives can be translated into concrete, interoperable data infrastructures in practice.

2.5 Diversity and Circulation as the Other Side of the Economic Equation

In the music sector, diversity, discoverability, circulation, and remuneration are not separate policy domains but interdependent outcomes of the same underlying data infrastructure. Music that cannot be reliably identified, contextualised, and linked across systems cannot be found by audiences, programmed by intermediaries, or remunerated through rights management systems. From this perspective, diversity indicators and economic indicators represent different views on the same infrastructural capacity: the ability of the ecosystem to make European music visible and actionable in data-driven environments.

Historically, diversity policies have often been treated as normative or corrective instruments, while economic sustainability has been addressed through market mechanisms and rights enforcement. In digital music ecosystems, this separation no longer holds. Algorithmic discovery, platform analytics, and automated rights processing rely on shared metadata pipelines. Where these pipelines are fragmented or unreliable, both diversity objectives and economic outcomes fail simultaneously. A repertoire that is invisible to discovery systems will not circulate, and a repertoire that does not circulate will not generate income, regardless of its cultural value.

The focus on diversity and circulation in this paper should therefore be understood as an analytical entry point into a broader economic problem, not as a departure from it. By examining how local, linguistic, and culturally specific repertoire can be made observable without exposing personal data or essentialising identity, the implementation described here addresses the same infrastructural conditions required for fair remuneration and sustainable markets. Diversity metrics become a stress test for the data infrastructure's ability to support lawful, auditable, and scalable reuse.

In this sense, diversity-oriented observability functions as an early and demanding use case. If an infrastructure can support reliable measurement of circulation and representation under strict legal and ethical constraints, it can also support more conventional economic indicators. The diversity and

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

circulation pillar of the Open Music Observatory thus serves both as a policy-relevant implementation in its own right and as a proxy for the broader economic health of the European music ecosystem.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

3 From Policy to Implementation: Governance, Architecture, and Semantic Design

This section explains how abstract policy objectives are translated into concrete technical and organisational choices. Throughout this report, we focus on implementation feasibility within existing regulatory frameworks, rather than normative evaluation of those frameworks. Our implementation is heavily influenced on how to comply with the Slovak diversity regulation, bearing in mind the possibility for replication in Hungary and elsewhere.

In this section we show how governance arrangements, system architecture, and semantic design are treated not as downstream technical matters, but as primary implementation mechanisms shaped by legal constraints, institutional competences, and the principle of subsidiarity. The subsections that follow describe how governance-by-design, layered architectures, and modular semantic patterns jointly enable policy-relevant observability without centralising authority or exposing sensitive data.

3.1.1 Governance as an Implementation Method

From the outset, the *Open Music Observatory* implementations treated governance as a methodological problem rather than an administrative layer. European cultural policy is structurally multi-level: cultural competences reside primarily at national and regional levels (including federal and Länder-based systems), while the European Union operates through coordination, support, and aggregation. Any viable implementation must therefore align with this reality, enabling subnational, national, and supranational participation without assuming a single centre of authority. Governance, in this context, becomes the method through which policy objectives are translated into workable infrastructures across levels.[10]

At national level, this logic was formalised in Slovakia through a *Memorandum of Understanding* between the Ministry of Culture, its analytical unit (IKP), and the OpenMusE consortium, later amended with a more specific MoU among the *SOZA*, *Slovak Music Centre*, the *Music Fund*, and the *Slovak National Library*, and *Reprex* as the technology partner. Both MoUs deliberately avoid mandating data transfer, standardisation, or compliance. Instead, it establishes shared analytical principles—open policy analysis, methodological transparency, and non-exclusive cooperation—while preserving institutional competences.[11] This governance choice is reflected in the technical design of the Slovak module, which is documented separately in the Open Music Observatory technical documentation and treated as a living specification rather than a fixed standard.[12]

- The **Slovak Comprehensive Music Database** is available at https://hudobnadatabaza.sk/.

Originally, the Slovak use case was planned to be replicated in Bulgaria. After a change in the granted action, first a Hungarian, later a Polish replication was planned.

[10] See the European Data Strategy and the emphasis on coordination, subsidiarity, and sectoral data spaces (European Commission 2020; European Parliament and Council 2022).
[11] *Memorandum o porozumení o využití výsledkov analýz otvorených politík v kontexte slovenského kultúrneho a kreatívneho priemyslu a sektorových verejných politík v spolupráci s konzorciom OpenMuse* (Ministerstvo kultúry SR and Open Music Europe 2023).
[12] *Open Music Observatory: Technical Documentation* (versioned, living documentation) (Daniel Antal 2025c).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

The Hungarian replication planning started in 2024, with a closely resembling plan to Slovakia. A comparable level of centralised coordination could not be achieved at the outset. As a result, the Hungarian implementation followed a deliberately bottom-up trajectory. Building on the Slovak template and its early public articulation in conference presentations, the Hungarian work progressed through a series of bilateral agreements, exploratory workshops, and open professional discussions with individual institutions. This process involved sustained dialogue with archival, heritage, and research actors, rather than a single coordinating authority. While slower and more fragmented, this bottom-up approach proved well suited to the Hungarian institutional landscape and ultimately produced a governance structure aligned with local capacities, traditions, and trust relationships. In this sense, the divergence between the Slovak and Hungarian governance paths is not a deviation from the shared architectural vision, but an expression of the subsidiarity principle in practice[13].

The Hungarian data sharing space does not have yet an end-user facing surface and an API. It exists as a data sharing space that connects several national-level databases, as explained in *Enriching and Futureproofing the Databases of the Hungarian Heritage House*[14], and in the service integration around the databases of the *Hungarian House of Music* (explained later.)

An emerging Polish track further illustrates how this governance pattern can be extended without requiring uniform legal instruments or synchronised institutional readiness. The Polish implementation is at an even earlier stage, and we are in the phase of replication design, and design of data curation workflows.

Below the national level, the *Finno-Ugric Data Sharing Space* (FUDSS) makes subsidiarity operational in its most demanding form. It engages subnational and community-scale actors—regional archives, village museums, research groups, and community organisations—often operating across borders and outside state-scale cultural infrastructures. Such actors are frequently referenced in European policy discourse on cultural diversity, minorities, and regions, yet are rarely integrated into European data infrastructures in practice. The *Finno-Ugric Data Sharing Space* includes community-curated databases for Livonian and Karelian musical heritage. In these cases, the institutional depth of a nation state—such as that available in Slovakia or Hungary—is largely absent. Livonian and Karelian heritage is preserved by a small number of dedicated regional GLAM institutions, research centres, NGOs, and community organisations, including the Livonian Institute.

Under these conditions, a centrally coordinated governance model would be neither feasible nor appropriate. Instead, the implementation necessarily resembles community-driven approaches familiar from infrastructures such as Wikidata, MusicBrainz, or Discogs, albeit combined with more explicit data curation and governance safeguards. Authority is distributed, participation is voluntary, and trust is built through transparent methods rather than formal mandate.

[13] *A szlovák adatkicserélési tér magyarországi föderációjának lehetőségei* [The Hungarian federation possibilities of the Slovak data sharing space], presented in 3-4 April, 2024 in Eger. (Daniel Antal 2024a) Subsequent presentations lead to a different, more archival oriented replication with the House of Hungarian Heritage and the House of Music with its library and archival collection.

[14] A short feasibility study and early manuscript for publication, which is the developing manuscript for a separate publication (Daniel Antal and Zagyva 2025).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

FUDSS functions as a methodological stress test for whether European frameworks can reach the smallest custodians of cultural heritage without forcing them into state-centric, platform-centric, or extractive governance models.[15]

- The **Finno-Ugric Data Sharing Space** is available at https://finnougric.net/.

At the same time, the *Open Music Observatory* explicitly incorporates a supranational layer. Beyond the Slovak, Hungarian, Polish, and Finno-Ugric modules, a central observatory layer is designed to aggregate indicators, support comparison, and feed policy-relevant evidence upward without absorbing authority or data stewardship. This layer mirrors the governance position of European cultural infrastructures such as *Europeana*, the *European Collaborative Cloud for Cultural Heritage (ECCCH)*, the *Culture Compass*, and the planned *EU Culture Data Hub*: coordinative rather than directive, integrative rather than extractive.[16] Comparable governance logics have long been present in European broadcasting and cultural data coordination practices, including those developed within the EBU context.[17]

Taken together, these four modules—subnational (Finno-Ugric), national (Slovak), replicated national (Hungarian and emerging Polish), and supranational—constitute a governance model rather than a single system. They demonstrate how European cultural data infrastructures can operate simultaneously downward and upward: enabling the participation of the smallest cultural actors, supporting national policy implementation, and supplying aggregated observability at European level. In this sense, the Open Music Observatory does not merely align with *Europeana*, *ECCCH*, the *Culture Compass*, or the future *EU Data Hub*; it offers a concrete governance pattern for how such infrastructures can respect subsidiarity while remaining analytically and operationally coherent across Europe.

This governance approach also provides the basis for a direct comparison with the *Copyright Infrastructure Task Force (CITF) First Project*. As discussed in the following subsection, both initiatives translate prior analytical work into concrete, voluntary implementations under real institutional constraints. The comparison is therefore not conceptual but practical: it examines how shared analytical principles—layered infrastructures, respect for competence, and governance-by-design—are realised in different cultural domains, with books and music presenting distinct but comparable challenges.

The CITF First Project offers a reference implementation in a domain with relatively linear workflows and limited actor diversity (see Section 2.4). The Open Music Observatory implementation operates at the same maturity level but addresses a structurally more complex ecosystem involving simultaneous coordination among libraries, collective management organisations, broadcasters, platforms, and community custodians. The governance patterns outlined above—particularly the emphasis on subsidiarity, voluntary participation, and non-extractive aggregation—form the common ground on which this comparison is made.

---

[15] *Federating Open Knowledge through Wikibase: The Case of the Finno-Ugric Data Sharing Space* (Daniel Antal et al. 2025; Dániel Antal et al. 2025).
[16] *Draft Joint Declaration "Europe for Culture — Culture for Europe"* and related Culture Compass materials (European Parliament, Council of the European Union, and European Commission 2025).
[17] Governance practices discussed in European broadcasting and metadata coordination contexts, including CISAC and EBU-related forums (Mikš 2025).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

3.1.2 Architectural Commitments under Policy Constraints

The architectural design of the *Open Music Observatory* is shaped by policy constraints rather than technical convenience. European cultural policy operates under strict limits on centralisation, competence transfer, and personal data processing.

As a result, the architecture does not aim to consolidate music data into a single repository or to replace authoritative sectoral systems. Instead, it is explicitly designed as an observability layer: a structure that computes indicators, exposes aggregates, and links back to authoritative sources while leaving data ownership, stewardship, and legal responsibility in place[18].

Federation is therefore a foundational commitment rather than a transitional solution. The architecture assumes the coexistence of multiple authorities, heterogeneous data sources, and uneven institutional maturity across Europe[19].

National, subnational, community, and European-level components are treated as structurally distinct yet interoperable elements. This mirrors the logic of European cultural infrastructures such as *Europeana* and *ECCCH* and aligns with the design assumptions of the *Culture Compass* and the planned *EU Culture Data Hub*, which require comparability and aggregation without harmonisation by force.

A second core commitment is the strict separation of authority, computation, and publication. Authoritative data remains under the control of competent institutions—libraries, archives, collective management organisations, broadcasters, and community custodians. The Open Music Observatory performs reconciliation, aggregation, and indicator computation without assuming decision-making authority over the underlying data. This separation is essential for lawful participation under GDPR, for maintaining institutional trust, and for enabling voluntary cooperation across public, private, and community actors[20].

The architecture also commits to workflow-aware interoperability. Music data circulates through fundamentally different organisational contexts, each with its own legal obligations, vocabularies, and practices. Rather than imposing a single semantic model, the architecture accommodates overlapping and partially incompatible representations, focusing on alignment where necessary and divergence where meaningful. This approach reflects the European Interoperability Framework's emphasis on organisational and semantic layers and recognises that semantic coherence in complex cultural ecosystems must be negotiated rather than enforced[21].

---

[18] This architectural position reflects core principles of the European Data Strategy and the Data Governance Act, which emphasise interoperability, data sovereignty, and the avoidance of unnecessary centralisation in public-interest data infrastructures (European Parliament and Council 2022).

[19] The federated observability model aligns with the governance logic articulated for *Europeana*, the *European Collaborative Cloud for Cultural Heritage* (ECCCH), and the *Culture Compass for Europe*, all of which rely on aggregation and coordination rather than directive control (BDVA/DAIRO Federation Working Group 2023; European Parliament, Council of the European Union, and European Commission 2025).

[20] This design choice reflects both European data protection law, in particular the *General Data Protection Regulation* (GDPR), and the *CARE Principles for Indigenous Data Governance*, which complement FAIR data principles by foregrounding authority, responsibility, and ethical use in community and heritage data contexts (European Parliament and Council 2016; Carroll et al. 2020).

[21] The emphasis on organisational and semantic interoperability follows the European Interoperability Framework (EIF), which explicitly recognises that legal, organisational, semantic, and technical layers must be

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Finally, provenance and auditability are treated as first-order architectural constraints. Every indicator produced by the observatory is traceable to its source, transformation steps, and assumptions. This is not merely a technical safeguard but a policy requirement: without provenance, indicators cannot support accountability, nor can they safely feed into AI-driven environments or European-level policy instruments. By embedding provenance at architectural level, the Open Music Observatory enables continuous, auditable feedback into European cultural policy while respecting the limits of authority and competence that define the European governance model.

Recent discussions within the European cultural heritage sector further underline the importance of provenance-aware data governance in the age of AI. A joint *Impulse Paper* by the Open Future Foundation and the Europeana Foundation examines how cultural heritage institutions can navigate large-scale data publishing and AI reuse while preserving public accountability, authority, and trust. It highlights that, once cultural data enters automated and AI-driven environments, incomplete or degraded provenance is no longer a marginal archival issue but a structural risk that shapes downstream use at scale. While the focus of that work is cross-domain, its conclusions closely resonate with the challenges observed in music, reinforcing the need to treat authority control, attribution, and lifecycle-aware governance as core infrastructural concerns rather than sector-specific exceptions. (Keller 2025)

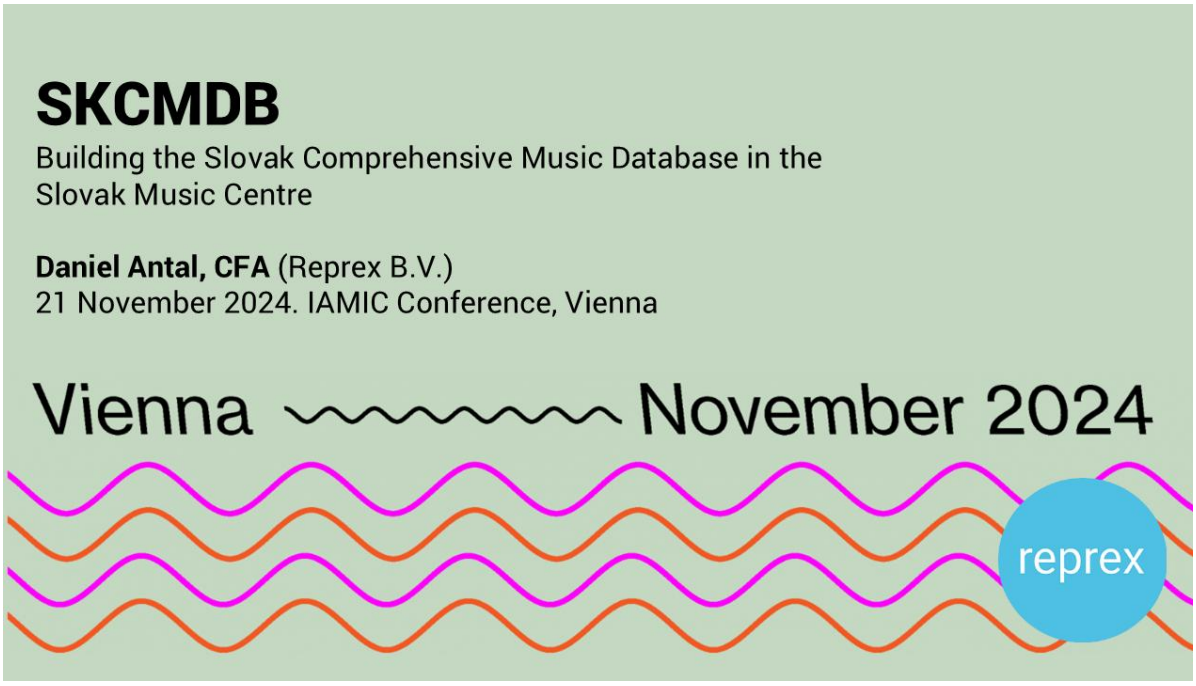
3.1.3 Semantic Design as Policy Translation

In the *Open Music Observatory*, semantic design functions as a mechanism for translating policy objectives into computable form. Cultural policy concepts such as “local repertoire,” “diversity,” “circulation,” or “European works” are not natively machine-readable; they must be operationalised through explicit distinctions between works, recordings, performances, agents, and contexts of use. The semantic layer therefore acts as the point where legal definitions, policy intentions, and sectoral practices are reconciled into shared, auditable meanings.

The design choices adopted here were not developed in isolation, but informed through sustained consultation with sectoral knowledge institutions and rights organisations.

aligned differently in complex, multi-actor public-sector ecosystems (Commission and Digital Services 2017; European Commission 2017).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295



SKCMBDb: Building the Slovak Comprehensive Music Database in the Slovak Music Centre. The presentation was made in the General Assembly and Conference of the International Association of Music Information Centres (IAMIC) on 21 November 2024 as part of the series of stakeholder consultations for the Open Music Observatory. <https://doi.org/10.5281/zenodo.18036464>

Our approach presented here was co-presented jointly with the Slovak Music Centre, an IAMIC member organisation, at the *International Association of Music Information Centres* General Assembly and Annual Conference (Music Austria, Vienna, 21 November 2024). Discussions with music information and documentation centres from several EU Member States focused on the feasibility of replicating the Slovak model—both its governance arrangements and its technical architecture—within different national institutional environments. See the presentation and poster format (Daniel Antal 2024c, 2024d).

We presented and discussed these ideas jointly with the Slovak Music Centre at the *International Association of Music Libraries, Archives and Documentation Centers* (IAML) during their General Assembly and Annual Conference (Salzburg, 7–9 July 2025), within the “Future Libraries” session and the poster session. See the presentation and poster format (Daniel Antal 2025d, 2025e).

The discussion focused on the role of national music libraries and documentation centres as data stewards within federated cultural data spaces. Feedback from IAML participants contributed to clarifying the library-oriented replication pathway and helped accelerate the otherwise slow Hungarian replication process, supported by the positive reception of IAMIC Hungary. The exchange also identified potential future replication interest in Portugal, Finland, and Denmark, although these initiatives are likely to materialise beyond the timeline of the Open Music Europe project. These discussion reinforced the parallels with the first CITF project, highlighting the regulatory and standardisation roles of national libraries are authority control bodies (see Section 2.4).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295



CISAC European Committee Meeting as OpenMusE: Towards a Sustainable Licensing Market for AI Use of Protected Works in Vilnius, on 29 April 2025.

Our work was also presented at the CISAC European Committee Meeting as *OpenMusE: Towards a Sustainable Licensing Market for AI Use of Protected Works* in Vilnius, on 29 April 2025 (Mikš 2025) in comparison with similar metadata governance and repair initiatives from several countries. Our presentation differentiated itself from similar projects in Finland and the United Kingdom with not only fixing the metadata at source *in the future*, but also focusing on legacy metadata repair. Given that the current copyright and neighboring right protection terms will require rights managements for 20th century assets well into the later part of the 21st century, our focus on not only future-proofing, but past-proofing with legacy system interoperability set our approach apart.

These consultations highlighted a recurring tension: the same musical artefact is described, classified, and used differently by libraries, archives, collective management organisations, broadcasters, and platforms. Semantic interoperability therefore cannot mean semantic uniformity.

To address this, the Observatory adopts a modular semantic strategy based on ontological patterns rather than a single comprehensive schema. Core distinctions—such as between musical works, sound recordings, and performances—are stabilised, while domain-specific vocabularies are allowed to coexist at the edges. This enables alignment with library-oriented standards (e.g. authority control and bibliographic description), industry metadata practices (e.g. ISRC, ISWC, DDEX-aligned workflows), and rights-management perspectives, without forcing any actor to abandon its internal epistemology.

This approach directly reflects feedback from IAML and IAMIC consultations, where national music libraries and documentation centres emphasised the need to preserve scholarly and archival integrity, as well as from CISAC-related discussions, where collective management organisations stressed legal precision, provenance, and accountability. Semantic design thus becomes a negotiated space: not an abstract exercise, but a governance-sensitive process that encodes institutional competences and legal responsibilities into the data model itself.

In this sense, semantic design performs the same function for music diversity and circulation as the layered infrastructure does for copyright data in the *CITF First Project* (see Section 2.4). In both cases, policy translation depends on making concepts computable without collapsing institutional differences. The music case demonstrates that, as policy questions become more complex—combining cultural diversity, rights management, and algorithmic mediation—semantic design must explicitly account for organisational workflows and use contexts. This makes it a central implementation method, not a downstream technical detail.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

The diversity and local content indicators operationalised through the SKCMDb could furthermore be situated within the broader social-ecological systems framework (SESF) presented in OpenMusE Deliverable D3.4 *Comprehensive Indicators for the Assessment of European Music Ecosystems*. D3.4 provides a harmonization structure for indicators developed across OpenMusE work packages, adapting Nobel Prize-winning economist Elinor Ostrom's framework to the music sector. In SESF terminology, the diversity and local content indicators demonstrated here would be categorized as Resource variables (properties of the cultural artefacts being distributed) and Governance variables (rules determining eligibility and classification), while the actors involved—composers, performers, CMOs, broadcasters—would be mapped according to D3.4's Actor tier categorization. This alignment ensures that diversity indicators generated through SKCMDb are comparable and combinable with economic indicators (WP1) and citizenship indicators (WP3), facilitating both vertical (across governance levels) and horizontal (across policy domains) coordination within the emerging EU Culture Data Hub. This analysis is taken further in OpenMusE D2.2 *Report on Music Diversity and Circulation*.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

4 Methodological Considerations

This section outlines the methodological foundations that underpin the design and operation of the *Open Music Observatory*. It clarifies the modelling assumptions, interoperability strategies, and publication practices that allow heterogeneous institutional data to be combined into reproducible, auditable indicators. Rather than proposing a universal or exhaustive model of music data, the methodology emphasises selective representation, provenance awareness, and cross-context validation, ensuring that analytical results remain interpretable, legally robust, and transferable across national and community settings.

4.1.1 Data Modelling and Representation

Data modelling in the *Open Music Observatory* is driven by policy questions rather than by aspirations of exhaustive description. The objective is **not** to construct a complete or universal model of music, but to represent those entities and relationships that are necessary to make cultural policy objectives—such as diversity, circulation, and observability—computable under real-world legal and institutional constraints. As a result, modelling choices are deliberately selective, pragmatic, and oriented toward traceability rather than completeness.

In October 2023, a dedicated work session convened by the Polifonia project brought together stakeholders including Podiumkunst.net, the Open Music Observatory, the University of Florence, IC Fonseca School, Joséphine Simonnot/PRISM, Maria Luisa Onida/D'Istruzione Superiore Leonardo Da Vinci, the Carnegie Hall Archive, and the Municipality of Bologna[22]. This consultation played a formative role in refining the modular ontology strategy adopted here, helping to ensure compatibility between scholarly music-knowledge graphs and policy-oriented interoperability requirements, while avoiding premature ontological closure.

At this stage, we started to record user stories for the design of system competences of the *Open Music Europe* software ecosystem and the observatory's semantic architecture[23].

A key methodological insight was that such interoperability could not be achieved through a single, monolithic ontology. Instead, we adopted a **modular approach based on ontological patterns**, inspired in part by the Polifonia Ontology Network but driven by different priorities. While Polifonia emphasises scholarly completeness and expressive richness, our primary objective was sufficient interoperability across existing standards and practices. The resulting data model is intentionally selective. It is designed to interoperate, where necessary, with DDEX-aligned industry metadata, Dublin Core–oriented library and research resources, ISWC-based author identification used by collective management organisations, and ISRC-based recording identification used in neighbouring rights management. This modularity allows different communities to engage with the system at different levels of granularity, without forcing full semantic alignment.

[22] *The Polifonia Ontology Network: Building a Semantic Backbone for Musical Heritage* (Berardinis et al. 2023) and *Engineering Ontologies with Patterns – The eXtreme Design Methodology* (Blomqvist, Hammar, and Presutti 2016)

[23] The user stories were placed in an OPA compliant folder on GitHub. This work was not carried on after the initiated grant agreement change, when the WP4 did no longer participate in building open source software for this purpose, but we did not abandon the competency requirements (Open Music Europe Consortium 2023).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

A foundational methodological decision is the explicit separation between different types of musical entities that are often conflated in practice: musical works, sound recordings, and performances. The Slovak quota legislation defines a 25% quota played between 6 AM and midnight. The quota's definition is the use Slovak musical works. However, works cannot be directly used by a radio station, unless they broadcast a very rare live performance of a composition. They work with sound recordings, which contain the recorded fixations of the performances of the works.

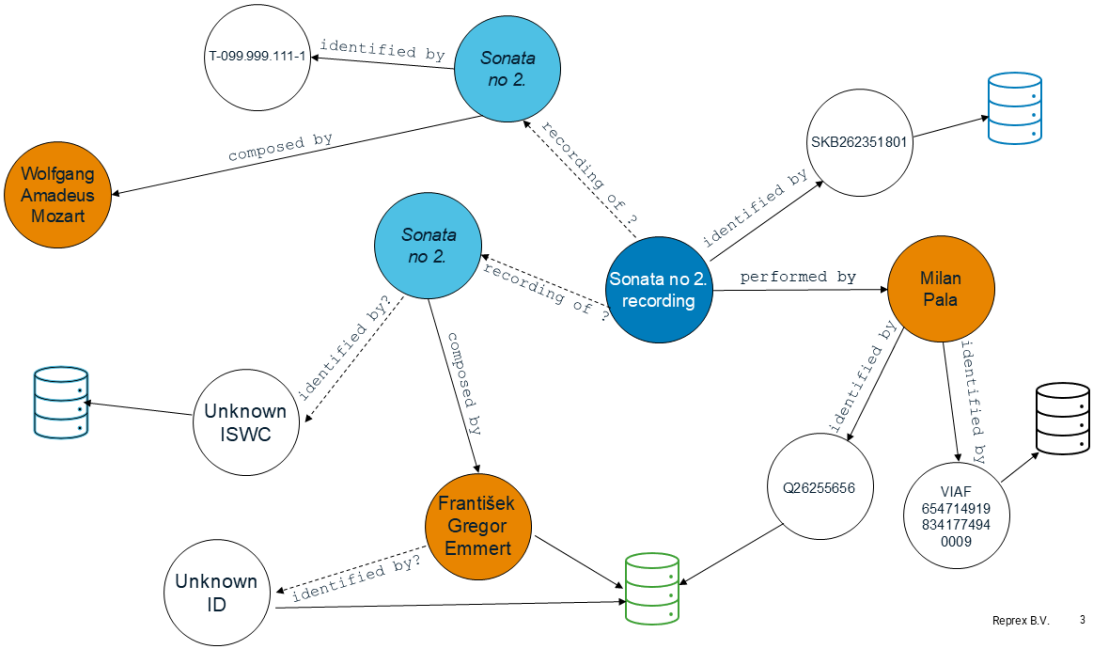
- Sound recordings are the primary artefacts circulated by broadcasters and platforms and therefore form the operational basis for measurement and indicators.
- Musical works, by contrast, are abstract intellectual entities governed by authors' rights and collective management regimes.
- Performances occupy a third category, representing time-bound social events with fluid participation and contextual significance.

Treating these entities as interchangeable would obscure both legal distinctions and policy relevance, and even more the actual implementation and monitoring of the legislative intent. Their separation in the data model precedes any evaluation logic and underpins all subsequent computation. Radio stations do not necessarily know who is the composer of the recording that they put into their playlist; even if they do, they most certainly do not know where does the all the authors of the work live (given that the according to the Slovak legislation, a "Slovak work" has either Slovak lyrics, or at least one author of the text has or had a permanent residence in the territory of the Slovak Republic.)

For the quota to be implementable, we need to link the sound recordings to composers, and evaluate if those composers live or had lived in the territory of the Slovak Republic.

The representation of agents—composers, performers, ensembles, producers, and institutions—is similarly constrained by policy and legal considerations. Personal attributes are not modelled for public exposure or direct analytical use. Instead, where policy criteria depend on characteristics of contributors, these are evaluated internally by competent institutions and expressed externally only through artefact-level outcomes (for example, whether a recording qualifies under a given policy rule). This modelling strategy supports lawful internal evaluation while avoiding the disclosure or stabilisation of sensitive personal data, aligning representation choices with governance and data protection requirements.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295



The example shows how a knowledge graph can represent incomplete and competing relationships between recordings, works, agents, and identifiers across federated databases without enforcing a single hierarchy. Uncertainty is kept explicit, enabling explainable, human-controlled reconciliation and policy-relevant analysis while avoiding the exposure of sensitive personal data. Reuse the visualisation: https://doi.org/10.5281/zenodo.17897724

To support interoperability across heterogeneous systems, the data model relies on graph-based representations rather than flat relational structures. Graphs provide the flexibility required to express multiple, overlapping relationships between works, recordings, agents, places, languages, and institutions without forcing a single hierarchy. At the same time, the model avoids ontological maximalism. Rather than adopting or extending a single comprehensive ontology, it uses modular patterns that can interoperate with existing standards where needed and remain silent where alignment would add complexity without policy value.

Crucially, representation is designed to distinguish between authoritative assertions and derived analytical statements. Authoritative data—such as identifiers, catalogue records, or rights-related facts—remains anchored to its source institution and is referenced with explicit provenance. Analytical representations, including inferred relationships or eligibility outcomes, are treated as derived artefacts that can be regenerated, revised, or withdrawn as methods evolve. This distinction enables reproducibility, auditability, and institutional trust, while ensuring that the data model remains adaptable to new policy questions and implementation contexts.

The precise data model is available in the data paper *Open Music Europe Recorded Music Datasets* [24].

4.1.2 Interoperability Mechanics

Interoperability in the *Open Music Observatory* is treated as a practical coordination problem rather than a purely semantic or technical one. Music-related data circulates across institutions with fundamentally different mandates, workflows, and legal responsibilities: collective management

[24] *Open Music Europe Recorded Music Datasets* (Daniel Antal, Vieira, and Mester 2025).

organisations, libraries, archives, broadcasters, platforms, and community initiatives often rely on overlapping data elements while assigning them different meanings and operational roles. Full harmonisation—whether through a single schema or a unified database—is neither feasible nor desirable under these conditions. Instead, interoperability is achieved through negotiated alignment across systems that remain institutionally autonomous.

Methodologically, this approach follows the logic of the *European Interoperability Framework* (EIF), which distinguishes legal, organisational, semantic, and technical layers. In the music domain, semantic alignment alone is insufficient unless it reflects organisational realities. For example, a library, a collective management organisation, and a digital distributor may all reference the same musical work or sound recording, yet do so for different purposes, under different legal regimes, and with different tolerance for uncertainty or incompleteness. Interoperability therefore requires making these differences explicit and bridgeable, rather than suppressing them. Semantic mappings are designed to respect divergent workflows while enabling shared reference where policy-relevant questions demand it.

At the technical level, interoperability is implemented through identifier reconciliation, controlled mappings between existing standards, and lightweight ontological patterns rather than comprehensive ontologies. The system is designed to interoperate selectively with DDEX-aligned industry metadata, ISRC-based recording identification, ISWC-based work identification used by CMOs[25], and library- and archive-oriented standards such as *Dublin Core* and *Records in Contexts*[26]. Crucially, these alignments do not imply that all participating actors must adopt the same vocabulary or level of granularity. Instead, they allow specific equivalences to be asserted, qualified, and traced with provenance, enabling downstream computation without forcing upstream conformity. The way we implemented ontological patterns of these music industry, library and archival standards is explained in the technical documentation of the Open Music Observatory[27].

A key design principle is the separation of authoritative data stewardship from interoperability and analysis layers. Authoritative sources—such as national libraries, CMOs, or archival institutions—remain systems of record, governed by their own legal and organisational constraints. Interoperability layers operate on references, mappings, and derived representations that can be regenerated, revised, or discarded without altering the underlying authority. This separation is essential both for governance-by-design and for compliance with data protection principles, as it allows internal evaluation and reconciliation without exposing personal or sensitive data beyond lawful contexts.

Finally, interoperability is treated as an ongoing process rather than a one-off achievement. As new actors join, policy questions evolve, or institutional practices change, mappings and alignments are expected to be revised. The *Open Music Observatory* therefore embeds interoperability mechanisms within transparent workflows that can be inspected, contested, and improved over time. This process-oriented view aligns with the comparison to the *CITF First Project* discussed earlier: both

[25] *International Standard Musical Work Code* (ISWC). ISO 15707:2022 (ISO 2022), *International Standard Recording Code* (ISRC). {ISO} 3901:2019 (ISO 2019a).
[26] In Slovakia, we placed a greater emphasis on libraries following the Dublin Core standards (Core 2020), and in Hungary on archives (Archives Expert Group on Archival Description 2023). Archives only start their transition to RiC from ISAD(G) and related standards, and backwards compatibility is a top priority.
[27] See (Daniel Antal 2025c), particularly the Appendix A — Data Model↗.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

implementations demonstrate that durable interoperability emerges not from technical closure, but from sustained coordination between governance arrangements, organisational realities, and semantic design.

4.1.3 Technical Publication and Scalability

Technical publication in the *Open Music Observatory* is explicitly separated from authoritative data stewardship. Systems of record—such as national libraries, collective management organisations, archives, or institutional catalogues—remain the sole authoritative holders of primary data.

The Observatory does not replicate or replace these systems. Instead, it operates through **derived, disposable publication layers** that are generated from authoritative sources for specific analytical and policy purposes. This distinction is central to both governance and scalability: authority is preserved upstream, while computation and analysis occur downstream on reproducible representations.

The primary publication mechanism relies on Wikibase as an integration and editing environment, combined with structured RDF exports for analysis and reuse. Wikibase provides a pragmatic balance between formal semantic structure and editorial accessibility, enabling institutions and expert communities to participate without requiring deep semantic-web specialisation. At the same time, it supports explicit provenance, qualifiers, and references, which are essential for auditability and institutional trust. Importantly, Wikibase functions here as a *publication and coordination layer*, not as a system of record. As we explain in the Open Music Observatory technical documentation and in our Green Paper, Wikidata is a quasi-standard setter, and hundreds of European cultural institutions, including national libraries, and the EU Publications Office moved to its open-source software system, Wikibase[28].

For scalable dissemination and analysis, RDF exports are further transformed into efficient, queryable formats, including HDT (Header–Dictionary–Triples). HDT enables large graphs to be queried without full materialisation in memory or duplication across systems. This makes it particularly well suited to scenarios where graphs are regenerated periodically and treated as disposable analytical artefacts rather than permanent databases. Publication outputs can therefore be shared, archived, or recomputed without locking institutions into long-term storage or synchronisation commitments.

Scalability in this model is achieved through **reproducible regeneration**, not through accumulation. Rather than continuously appending new data to a growing central store, the Observatory regenerates its publication layers from authoritative sources using documented workflows. This approach supports versioning, comparison over time, and correction of earlier assumptions without technical debt. If a source dataset changes, a policy rule is refined, or a modelling decision is revised, the entire analytical layer can be rebuilt transparently.

This publication strategy aligns with European data governance principles that prioritise subsidiarity, proportionality, and non-extractive reuse. By avoiding central data hoarding, the Open Music Observatory remains compatible with institutional autonomy, data protection constraints, and evolving policy requirements. At the same time, it enables pan-European observability by ensuring

[28] 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

that indicators and analytical results are reproducible, auditable, and scalable across jurisdictions and domains.

4.1.4 Cross-Context Validation

The methodological choices underpinning the *Open Music Observatory* were not validated in isolation, but through sustained application across multiple institutional, cultural, and governance contexts. Validation, in this sense, does not mean optimisation within a single controlled environment, but resilience under variation: methods are considered robust insofar as they continue to function when confronted with different legal regimes, institutional capacities, cultural repertoires, and scales of operation.

The Slovak implementation provides a validation context characterised by relatively strong national institutions, clear cultural policy instruments, and established cooperation between ministries, collective management organisations, libraries, and sectoral bodies. In this setting, methods could be tested against concrete policy questions—such as local content measurement—under realistic governance and data protection constraints. The Hungarian replication introduced a different validation axis. Strong archival traditions, ethnomusicological collections, and divergent institutional histories required adaptation of modelling and governance assumptions, particularly in relation to archival provenance, contextualisation, and rights uncertainty. The fact that core methods remained applicable under these conditions supports their transferability beyond a single national model.

The *Finno-Ugric Data Sharing Space* represents the most demanding validation context. Operating at subnational and community scale, often across borders and outside state-level cultural infrastructures, it tests whether the same methodological framework can function where institutional authority is fragmented or absent. Here, validation depends on the ability to support participatory curation, multilingual enrichment, and differentiated access without collapsing local epistemologies into national or European abstractions. The successful application of the same interoperability, provenance, and publication principles in this context demonstrates that methodological coherence does not require institutional uniformity.

Validation also occurred through exposure to supranational aggregation requirements. At the level of the Open Music Observatory's central module, methods are evaluated not on their capacity for detailed curation, but on their ability to support comparison across countries, repertoires, and policy regimes without reintroducing identity-based categorisation or excessive harmonisation. This upward validation mirrors the role envisaged for European cultural infrastructures such as Europeana, the Culture Compass, and the future EU Culture Data Hub, where observability must coexist with subsidiarity.

Beyond deployment contexts, cross-context validation was reinforced through structured engagement with professional and expert communities. Consultations and presentations at forums such as IAML, IAMIC, CISAC, Wikimedia community meetings, and European data governance events provided iterative feedback on modelling assumptions, governance choices, and interoperability strategies. These exchanges served as peer validation mechanisms, ensuring that methods remained intelligible and credible to practitioners operating under different constraints.

Governance, federation, and data sovereignty issues were discussed in a broader cross-sectoral context at *Dataweek*[24] in Leuven, within the framework of the *Big Data Value Association* and *Gaia-*

X[29]. Presenting the *Open Music Observatory* and its national modules as an example of a European music dataspace allowed the methodological approach to be evaluated against emerging governance expectations for federated data infrastructures, including principles of subsidiarity, non-extractive reuse, and institutional autonomy. Feedback from this forum directly informed the articulation of the federated data-sharing model described in the present paper.

Our work was presented in the Technology sessions of the *Wikimedia CEE Meeting 2024* (Istanbul) and the *Wikimedia CEE Meeting 2025* (Thessaloniki), as well as at *Wikidata Conf 2025* (online) (Daniel Antal 2024b, 2025f; Daniel Antal, Pigozne, and Federico 2025). These forums provided a practical testbed for validating modelling choices, editorial workflows, and governance assumptions against large-scale, community-maintained knowledge infrastructures.

Taken together, these validation contexts demonstrate that the methodological framework of the *Open Music Observatory* is neither tailored to a single institutional configuration nor abstracted away from practice. Its robustness derives from being tested across national, subnational, community, and supranational settings, and from remaining adaptable without losing analytical coherence. This capacity for cross-context operation is essential if music observability is to function as a credible forerunner for wider European cultural data infrastructures.

Through this engagement, we established working relationships with several national Wikimedia chapters and with the Wikidata and Abstract Wikipedia teams, and participated in the *Wikidata Ontology Cleanup Task Force* and the *Wikidata Mereology Task Force*. This involvement helped align our ontological patterns, identifier practices, and documentation strategies with broader open knowledge standards, while also ensuring that our approach to provenance, uncertainty, and institutional responsibility remains compatible with community-driven curation environments.

[29] Jun 5, 2024, Dataweek²⁴, Leuven, Belgium.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

5 Diversity Goals Without Exposure of Sensitive Data

The implementation of the **Slovak Comprehensive Music Database** (available at https://hudobnadatabaza.sk/), HUNDB (collection partly integrated into the OMO main module, partly at a temporary installation https://reprexbase.eu/hundb/index.php which is password protected), the **Finno-Ugric Data Sharing Space** (available at https://finnougric.net/) coincided with a broader shift in European policy thinking on data governance, artificial intelligence, and cultural infrastructures. Over the past decade, it has become increasingly clear that Europe's capacity to pursue cultural, social, and economic objectives in the digital domain depends less on isolated regulatory instruments and more on the availability of trustworthy, interoperable data infrastructures. This shift is reflected in the *European Data Strategy*, the *Data Governance Act*[30], and the growing emphasis on sectoral data spaces as instruments of public policy.

In the cultural and creative sectors, this transformation has been driven in part by the rapid platformisation of music distribution and discovery. Recommendation systems, analytics pipelines, and increasingly generative AI models now mediate access to cultural content at a global scale. These systems rely on metadata that is often incomplete, inconsistent, or governed by commercial incentives that do not align with public-interest objectives. As a result, cultural policy goals—such as diversity, linguistic plurality, gender balance, or local content promotion—are difficult to articulate, let alone enforce, within algorithmic environments dominated by non-European actors.



"Recognising the EU's supporting role in the field of culture, this declaration affirms shared political commitments and responsibilities of the European Parliament, the Council of the European Union and the European Commission - within their respective competencies and in full compliance with EU law - based on the key directions set out in the Commission's 2025 Culture Compass for Europe."

2. Cultural and linguistic diversity are fundamental values on which the EU is founded. We commit to:
• Supporting, celebrating and preserving the rich cultural and linguistic diversity of the European

[30] *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) (European Parliament and Council 2022)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Union and its cultural sovereignty.
• Promoting the full diversity of cultural content created in Europe, notably online and harnessing new technologies to ensure the visibility and discoverability of this content.
Draft Joint Declaration "Europe for Culture — Culture for Europe" (European Parliament, Council of the European Union, and European Commission 2025)

5.1.1 The Diversity & Circulation Aspects of European Music Policy

Against this backdrop, diversity quotas and local content regulations can no longer be understood as standalone cultural policy tools. Instead, they function as **stress tests** for Europe's data governance ambitions. If a policy goal cannot be expressed in machine-readable terms, evaluated using auditable data sources, and implemented without violating fundamental rights, it risks becoming symbolic rather than effective. The monitoring and enforcement gaps identified in earlier quota systems thus foreshadowed a more general problem: the absence of shared reference infrastructures capable of mediating between policy intent, data protection law, and algorithmic practice.

European institutions have increasingly recognised this challenge. Rather than advocating for centralised databases or uniform metadata schemas, recent policy documents emphasise **federation, interoperability, and governance-by-design**. Data spaces are envisaged as socio-technical arrangements in which multiple actors retain control over their data, while agreeing on shared semantics, interfaces, and rules of engagement. This approach is particularly well suited to the cultural sector, where data is distributed across public institutions, collective management organisations, private platforms, archives, libraries, and community initiatives, each operating under different legal and ethical constraints.

Within this framework, the principle of subsidiarity acquires renewed significance. Cultural identity, heritage, and creative production are deeply contextual and cannot be meaningfully governed through purely centralised mechanisms. At the same time, national solutions alone are insufficient in a digital environment characterised by cross-border circulation, global platforms, and transnational audiences. European data governance therefore requires **multi-level infrastructures** that operate simultaneously at community, national, and supra-national scales, without collapsing these levels into a single point of control.

A defining feature of the Slovak Comprehensive Music Database is the strict separation between **internal evaluation** and **external disclosure**. While the system may internally process attributes that are relevant for diversity or local content criteria, it exposes only binary or aggregated outcomes. At the level of the recording, a track is reported as meeting or not meeting a quota requirement, without revealing which contributor, which personal attribute, or which specific criterion triggered eligibility.

This separation is central to the viability of diversity-aware cultural policy under European legal and ethical constraints. It operationalises the principle of data minimisation by ensuring that only the information necessary for editorial, regulatory, or analytical purposes is disclosed. At the same time, it preserves the ability to evaluate complex quota definitions that depend on personal attributes which cannot lawfully or ethically be exposed.

Implementing this distinction requires more than a technical design choice. It is precisely the challenge addressed by **legal, organisational, and semantic alignment** under the European

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Interoperability Framework. The system must be able to lawfully access authoritative data sources, design an organisational workflow that produces a determinate evaluative outcome, and communicate that outcome back into a shared data space using semantics that can be consistently interpreted by diverse actors, including broadcasters, curators, researchers, and media authorities.

In the Slovak context, this workflow is deeply intertwined with rights management. The national data sharing space is organised around SOZA, the collective management organisation with legal competence to identify musical works and connect them to living rightsholders and their heirs. Personal characteristics of composers—such as place of birth, nationality, residence, or ethnicity—are strictly protected under the *General Data Protection Regulation* and are not accessible for disclosure. As a result, while the evaluative logic is conceptually straightforward, its practical implementation is necessarily mediated by institutional trust and competence.

At a high level, the internal evaluation proceeds through a sequence of controlled matches. A sound recording is identified and linked to the musical work it embodies; the work is associated with its composer or composers; and the system evaluates whether any of the relevant personal or work-level characteristics satisfy the applicable quota definition. If at least one criterion is met, the recording is marked as eligible for the purposes of the local content quota. Crucially, the reason for eligibility is never exposed. A composition may qualify because the composer was born in Slovakia, because the composer resides there, or because the lyrics are written in the Slovak language. If any of these conditions are satisfied, all recordings of works by that composer—including co-authored works—are treated as eligible, without disclosing which condition applied.

The result of this internal evaluation is exposed through a simple property at the level of the recording. For example, a recording may be labelled as heritage of Slovakia when it qualifies under the national quota, or as heritage of Slovaks when eligibility is based on linguistic intelligibility for Slovak speakers as an ethnolinguistic group. This modelling choice deliberately abstracts away from individual identity attributes while retaining policy-relevant meaning. It allows quota evaluation to operate at the level of cultural artefacts rather than persons.

Language of lyrics occupies a particular position within this framework. Unlike personal attributes, language is an audible and editorially relevant characteristic of a recording, yet it remains culturally mediated rather than strictly objective. In SKCMDb, lyrics are treated as Slovak when they are intelligible to a native Slovak speaker, even if they are performed in dialectal or archaic forms, as is common in folk traditions. Such determinations may be established by native-speaker curators, performers, or composers. This approach recognises linguistic variation without essentialising identity or excluding historically rooted forms of expression.

The is heritage of property used to express these outcomes is defined in the ECCCH *Heritage Digital Twin Ontology*[31], which subclasses CIDOC CRM and is widely used in museum and heritage contexts. In SKCMDb, this property—also used under the alias local to—represents a form of community stewardship governed primarily by ethical frameworks such as CARE, rather than by legal entitlement alone. In the Hungarian replication, the same property is used to refer not only to Hungarians as a national group, but also to linguistically distinct communities such as Székely or Moldavian Hungarian speakers in present-day Romania. A Szekler folk song may thus be treated as part of the Hungarian

[31] *Heritage Digital Twin Ontology (HDTO) – First Draft* (ECHOES Ontology Task Force 2025)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

national quota for policy purposes, while retaining a more precise ethnolinguistic designation for ethnomusicological analysis.

The same abstraction is applied at the level of nation-states. A recording is treated as heritage of or local to Slovakia or Hungary if it meets the respective national quota regulations. (For a critical discussion of the scope, limits, and normative assumptions underlying these diversity definitions, see *Deliverable D2.2*, which addresses these issues from a legal and policy perspective; in this report we model how the existing framework can be modeled for compliance and monitoring.)

This modelling choice enables the computation of statistical indicators—such as the proportion of locally relevant repertoire in charts or playlists—without exposing personal data or identity categories. It also allows the same underlying evaluation logic to support multiple policy objectives, including radio and television quotas, festival programming diversity, and international initiatives such as gender equality pledges.

It is important to note that the impact of the General Data Protection Regulation on the cultural heritage, music, and digital humanities sectors extended far beyond the specific context of diversity quotas. While GDPR contains differentiated provisions for libraries, archives, and other GLAM institutions, its introduction nevertheless came as a structural shock to the sector. Long-standing practices of metadata publication, authority sharing, and collaborative documentation—often developed in good faith and under earlier legal regimes—suddenly became legally ambiguous or non-compliant. This applies in particular to retroactively affected archival materials and to cross-institutional authority initiatives, such as shared name files or federated registries, which rely on the controlled exchange of personal data.

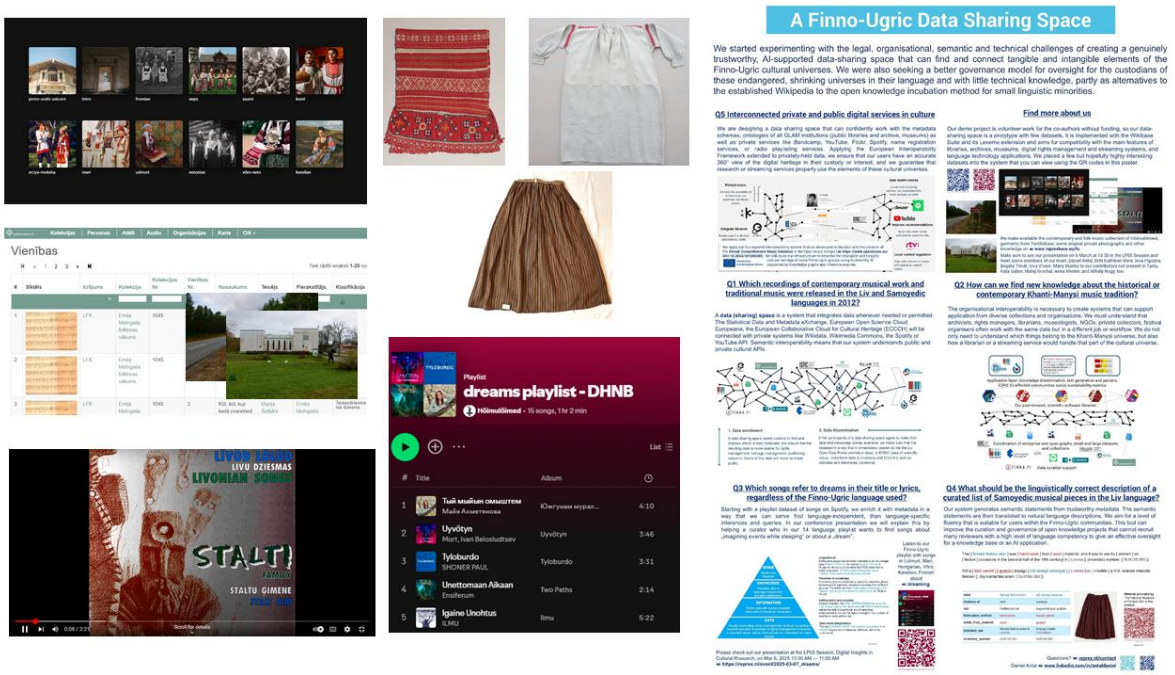
Within this context, the architecture developed for SKCMDb supports not only policy evaluation, but also institutional compliance and remediation. By enabling the lawful use of GDPR-protected data within trusted environments to correctly link agents to archival, library, museum, and rights registry holdings, the system helps participating institutions to repair metadata, clarify authorship and neighbouring rights relationships, and address potential moral rights issues. Crucially, this work remains internal to competent institutions, while its outcomes—such as corrected links, clarified work–recording relationships, or eligibility evaluations—can be safely exposed at an abstracted level.

In this sense, the approach taken here does not merely limit disclosure to avoid legal risk. It creates the conditions under which heritage institutions may, after careful analysis, increase the educational, research, or public visibility of their holdings in a legally and ethically defensible manner. Rather than treating GDPR as an external obstacle to cultural data work, the system incorporates it as a design constraint that, when properly addressed, enables more responsible sharing, clearer rights attribution, and more sustainable interoperability across the music and heritage sectors.

By design, this approach prevents the system from becoming a tool for profiling or surveillance. Editors and analysts receive determinate, interpretable results, while the underlying personal data remains protected within competent institutions. In this way, diversity goals are operationalised not through the exposure of identity, but through the careful design of evaluative workflows that respect legal constraints, ethical commitments, and the complex social realities of European cultural production.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

From a policy perspective, the *Finno-Ugric Data Sharing Space* functions as a subsidiarity and epistemic-justice stress test for European data governance concepts. Unlike the Slovak or Hungarian cases, it operates largely outside state-scale cultural infrastructures and engages communities whose heritage has historically been documented, classified, and interpreted through external—often post-imperial—institutional frameworks. This makes it particularly relevant for assessing whether European principles such as federation, non-extractive reuse, governance-by-design, and respect for cultural and linguistic diversity can be operationalised in low-scale cultural ecosystems.



See our original poster in Digital Humanities in the Nordic and Baltic Countries (DHNB 2025), Tartu, 5-7 March 2025, and the subsequent paper in the following footnote.

At the same time, the turn toward Finno-Ugric heritage was motivated by a pragmatic methodological consideration. While the Slovak and Hungarian implementations aimed for broad coverage across hundreds of thousands of works and recordings, such scale makes it difficult to observe, test, and refine governance and modelling assumptions in a controlled manner. By contrast, Livonian and related Finno-Ugric repertoires can be described using hundreds rather than hundreds of thousands of items. This reduced scale enables exhaustive modelling, close inspection of provenance and rights statements, and direct community engagement, providing a conceptually tractable environment in which architectural and governance choices can be stress-tested before being applied to larger national or pan-European systems[32].

In this sense, the Finno-Ugric module is neither peripheral nor exceptional. It functions as a *conceptual*and *methodological laboratory* for European data-space design, demonstrating how principles developed under conditions of maximal institutional fragmentation and linguistic diversity can inform the construction of scalable, trustworthy infrastructures in more complex cultural ecosystems. By enabling communities to participate in the repair, contextualisation, and reuse of

[32] See our accepted paper in preprint: *Federating Open Knowledge through Wikibase: The Case of The Finno-Ugric Data Sharing Space* and the original poster presentation *A Finno-Ugric Data Sharing Space* (Daniel Antal et al. 2025; Dániel Antal et al. 2025)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

authoritative cultural data without recentralising control, the module aligns European data-space logic with commitments to non-discrimination, proportionality, and fundamental cultural rights.

5.2 Application Layers and System Architecture

This section describes how components interact conceptually; detailed technical implementation is provided for replication teams with data expertise.

This section explains how the different parts of the system fit together and why each part is needed. Rather than focusing on software code or technical implementation details, it describes the roles of the main components and how information moves between them. The purpose is to help readers understand the logic of the system at a conceptual level; detailed technical specifications are provided separately for teams that will implement or replicate the system.

The layered system architecture of SKCMDb is a direct consequence of the federated governance model described in the preceding sections. The separation between data stewardship, interoperability, public access, and analysis is not merely a technical design choice, but an operationalisation of institutional competence, legal responsibility, and subsidiarity. In particular, it reflects the need to support policy-relevant evaluation and analytical reuse while ensuring that sensitive data remains under the control of the institutions legally entitled to process it.

At the foundation of the architecture is a **private or trusted layer**, operated by institutions with formal legal competence over authoritative data. In the Slovak case, this layer is anchored in the systems of the collective management organisation SOZA and, where relevant, other trusted institutional partners. This layer contains detailed information on musical works, rightsholders, and personal attributes that may be necessary for internal evaluation, rights clearance, or metadata repair, but which are protected under data protection law. Access to this layer is governed by contractual, legal, and organisational safeguards, reflecting the governance arrangements established through the national Memorandum of Understanding described in Section 2. Its contents are never exposed directly to external users or applications.

Behind the scenes, the trusted private layer performs metadata reconciliation, linking, and repair tasks using AI-assisted workflows within a data sharing space built on a knowledge graph. This model was chosen because it preserves full human control over both data sharing and inference: institutions retain authority over what data is processed and disclosed, while any inferred relationships—such as links between a musical work and its recorded fixations for royalty allocation—remain transparent, traceable, and auditable. By embedding AI within a governed knowledge graph rather than opaque pipelines, the architecture ensures that automation supports, rather than replaces, institutional responsibility, legal accountability, and human oversight[33].

8. The use, development and governance of artificial intelligence (AI) systems should foster human creativity through a fair human-centric and rights-based approach. It should: respect cultural rights, accessibility, inclusivity and cultural diversity; develop and promote discoverability of European, national and local content; foster competitiveness; counter digital divides; and foster digital inclusion. We commit to:

[33] See more policy context in our *Green paper*, and more technical details in the Open Music Observatory technical documentation (Daniel Antal 2025c, 2025a).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

• Promoting human creation and European cultural and linguistic ligital sovereignty, and addressing ethical risks of biases and cultural homogenisation.
• Protecting intellectual property rights, by addressing the impacts of AI on creators' remuneration, while embracing innovation. *Draft Joint Declaration "Europe for Culture — Culture for Europe"* (European Parliament, Council of the European Union, and European Commission 2025)

The following subsection documents how the private or trusted layer of the observatory was implemented in practice in the Slovak pilot. Rather than proposing a general legal model, it describes the concrete data protection reasoning, lawful bases, safeguards, and governance mechanisms adopted by SOZA and its partners to enable authority control and controlled data sharing in compliance with EU data protection law. It is relatively straightforward to replicate in another EU country, because GDPR applies directly in the entire European Union. Outside of the European Union, significant changes may be necessary.

5.2.1 Public-private governance and data protection

Prior to the actual establishment of the private or trusted layer as the first layer, data protection challenges needed to be addressed. SOZA collects and processes the personal data of its members for the purposes of collective management. Such collection and processing are based on consent. Although the concept of "collective management" as a lawful basis is sufficiently broad, for reasons of transparency and legal certainty SOZA considered it necessary to identify a new lawful basis for the further processing of personal data within the context of the SKCMDb.

Initially, SOZA prepared a consent-based participation scheme. However, due to the large number of members and the low response rate to participation requests, it became necessary to identify an alternative lawful basis for the intended processing. Consequently, a legitimate interest was selected as the lawful basis for further processing. In this context, a legitimate interest assessment, in the form of **a proportionality test**, was required and prepared.

The core of this proportionality test[34] consisted of assessing the potential impact of further processing of personal data on the rights and freedoms of the data subjects, namely SOZA's members. If the potential impact on these rights and freedoms were to override SOZA's legitimate interest, legitimate interest could not serve as a lawful basis for the intended further processing.

The personal data held by SOZA that is crucial for the proper functioning and establishment of the SKCMDb was divided into two groups, depending on whether the data would be made publicly accessible or accessible upon request to selected subjects. The categories of personal data concerned include name, surname, pseudonym, date of birth, place of permanent residence, and sex. Name, surname, and pseudonym are made publicly accessible, while date of birth, place of permanent residence, and sex are accessible upon request to selected subjects. This division of personal data into two tiers of access—public and upon request—plays a significant role in the

[34] The proportionality test was conducted in accordance with Article 6(1)(f) of Regulation (EU) 2016/679 (General Data Protection Regulation, short: GDPR), which permits the processing of personal data where it is necessary for the purposes of the legitimate interests pursued by the controller, provided that such interests are not overridden by the interests or fundamental rights and freedoms of the data subject. (European Parliament and Council 2016)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

management and protection of personal data. In this context, maintaining control over the data and its extraction is crucial for SOZA.

Based on these tiers of access, users of the SKCMDb can also be divided into two groups: data contributors and data extractors. Data contributors provide personal data, thereby contributing to the content of the SKCMDb and improving its quality. Data extractors exclusively extract data from the SKCMDb, either from the publicly available layer or from the upon-request layer. It should be noted, however, that data contributors also act as data extractors with respect to each other's data.

Data contributors include SOZA, HC, HF, and SNK, all of which are signatories to the *second Memorandum of Understanding* with Reprex, the technical data and knowledge processor is concluded, among other purposes, to enable controlled and legally sound data sharing[35]. This structure is very similar to the first project of the Copyright Infrastructure Task Force introduced in Section 2.4.

Data extractors include all data contributors, as well as various music streaming platforms, online knowledge databases, encyclopaedias, and similar services.

SOZA's legitimate interest was identified as including the following objectives:

- improving and streamlining the performance of collective rights management carried out by SOZA for its members;
- increasing the accuracy and availability of information about the musical works of SOZA's members;
- promoting Slovak music, musical works, musical recordings, and knowledge about Slovak music by making them more easily accessible and searchable; and
- promoting authors of musical works and music publishers represented by SOZA.

In light of the above, and in the context of the establishment of the SKCMDb, the potential benefits of further processing of personal data were assessed as not overriding SOZA's legitimate interest. Accordingly, legitimate interest could be employed as a lawful basis for the intended further processing of personal data.

As the lawful basis for further processing is legitimate interest rather than consent, an opt-out option must be provided to data subjects, namely SOZA's members. Notifications of the planned further processing of personal data, including information about the opt-out option, were distributed to data subjects via SOZA's internal user interface, an email newsletter, and a permanent notification on SOZA's website. The opt-out can be exercised by sending an email to SOZA's Data Protection Officer. In this context, the opt-out option serves as an important mechanism for protecting the rights and freedoms of data subjects.

In connection with the processing of personal data, cooperation with an external data processor was established. The external data processor was selected based on its technical expertise,

[35] *Memorandum o porozumení vo vzťahu k Slovenskej súhrnnej hudobnej database* - [Memorandum of Understanding in relation to the Slovak Comprehensive Music Database] (SOZA, Hudobné centrum, Slovenská národný knižnica, Hudobný fond and Reprex)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

trustworthiness, and implementation of appropriate data protection and security measures. The cooperation is governed by an agreement on the processing of personal data, which serves as the legal basis for the relationship and defines the mutual rights and obligations of the parties. The agreement also imposes obligations relating to confidentiality, security, and data protection. The SKCMDb is continuously updated and subject to ongoing technical and security supervision, ensuring compliance with EU data protection regulations and resulting in a fully functional private or trusted layer.

Above this sits an **intermediate interoperability layer**, which functions as a controlled interface between trusted data holders and the wider data sharing space. In the SKCMDb implementation, this role is fulfilled by a SPARQL endpoint implemented using Apache Jena Fuseki. This layer hosts derived RDF graphs generated from authoritative sources and refreshed on a regular basis. Crucially, these graphs are treated as reproducible publication artefacts rather than systems of record. . Due to ease of transferability between source and SPARQL endpoints, such graphs can be rebuilt in a matter of minutes ensuring that changes in the source material can be relayed to the data sharing space in near-real time. This modular architecture can also accommodate multiple graph databases that can be set up with ease using limited computational resources. Most importantly, these graphs contain only non-personal data and outcome-level properties resulting from internal evaluation, such as eligibility flags or heritage associations. In this way, the interoperability layer translates institutionally governed microdata into a form that can circulate safely across organisational and national boundaries without exposing sensitive underlying information.

5.2.2 Public Access Layer

The **public access layer** builds on this interoperability interface to support exploration, navigation, and reuse by a broader audience. In several modules of the Open Music Observatory ecosystem, including the Slovak implementation, this layer is realised through a Sampo semantic browser.

The Sampo Semantic Browser is implemented using the Sampo-UI framework, a full-stack JavaScript framework for developing semantic portal user interfaces. These interfaces are designed primarily for cultural heritage, research, and policy contexts[36], and support exploratory access to graph-based data through faceted search and multiple analytical perspectives. In the observatory, this layer allows users to browse repertoires, explore relationships between works, recordings, agents, and institutions, and apply faceted filters based exclusively on non-sensitive attributes. From a governance perspective, this layer embodies the commitment to transparency and public value articulated in the Slovak MoU, while remaining firmly decoupled from any processing of personal data.

The computational resources required to set up a Sampo-UI instance are low as actual data retrieval is managed by the intermediate interoperability layer. The layout and filtering mechanisms constrain queries such that the resulting query result set retrieved from that layer remains minimal. Building on this architecture, the practical effort required to deploy a Sampo-UI instance is intentionally kept

[36] Sampo-UI was developed by the Semantic Computing Research Group (SeCo) at Aalto University in collaboration with the Helsinki Centre for Digital Humanities (HELDIG). The framework has been applied in numerous national and international semantic portals in the cultural heritage and digital humanities domains. We are grateful to the Sampo team for their openness, documentation, and constructive feedback during the design of the observatory's semantic browsing layer. More information: *Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces* (Ikkala et al. 2019)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

low, as the framework is designed around a declarative configuration model rather than bespoke interface development. Core user-interface functionality—such as faceted search, result views, and data visualizations—is provided through reusable, generic components that can be adapted to a specific domain largely by defining JSON-based configuration files and SPARQL query templates. This allows a functional semantic browser to be instantiated rapidly once a compliant interoperability layer is available, without extensive front-end programming. In methodological terms, Sampo-UI thus shifts development work from implementation to configuration, reducing technical overhead, facilitating reproducibility, and making the framework particularly well suited for architectures operating under constrained resources.

Our use of Sampo-UI benefited substantially from guidance and feedback provided by the Sampo development team. In particular, discussions during a workshop on presenting Finno-Ugric cultural data at the *Digital Humanities in the Nordic and Baltic Countries* (DHNB) 2025 conference in Tartu helped clarify design choices and interaction patterns for semantic browsing in a multilingual and cross-institutional context. From a governance perspective, the semantic browser layer supports transparency and public value, as articulated in the Slovak Memorandum of Understanding, while remaining clearly separated from any processing of personal data and from the private or trusted layers of the observatory architecture.

The **analytical layer** provides tools for computation, aggregation, and visualisation, and is explicitly constrained to operate on the same outcome-level data exposed through the interoperability layer. In the Slovak case, this includes computational notebooks built in R and applications built with Shiny, which enable editors, analysts, or researchers to upload playlists or chart data and compute diversity or local-content indicators based on matches against the SKCMDb graph. These analytical tools do not require, and do not permit, direct access to trusted institutional data. Their design reflects the separation between internal evaluation and external use described in Section 5 ensuring that policy-relevant analysis can be performed without expanding the circle of access to protected data.

This layered architecture also enables replication across jurisdictions and institutional contexts. A similar configuration can be deployed elsewhere by identifying appropriate data stewards for the trusted layer, exposing derived graphs through a standardised interoperability interface, and reusing or adapting existing public and analytical applications. Because the analytical logic operates on outcome-level properties rather than raw personal data, the same tools can be reused across national implementations even where underlying legal regimes, institutional arrangements, or data sources differ. This was precisely the case in the Hungarian implementation, where a bottom-up governance model required tighter institutional boundaries, and in the Finno-Ugric data spaces, where trusted layers are distributed across NGOs and regional GLAM institutions rather than central authorities.

Importantly, the architecture also supports asymmetry between modules. Not all implementations are required to expose the same interfaces or support identical analytical workflows. In contexts where statistical monitoring is a priority, such as national broadcast quotas, the analytical layer may focus on time-series aggregation and indicator computation. In archival or ethnomusicological contexts, by contrast, public interfaces and analytical tools may prioritise qualitative exploration, provenance tracing, or competence questions that are not immediately reducible to numerical

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

indicators. The architecture accommodates these differences without fragmentation because all layers remain anchored in shared semantic models and governance principles.

By deriving its technical structure from governance and policy constraints rather than imposing them retroactively, the SKCMDb architecture ensures that diversity-aware cultural policy objectives can be supported without compromising data protection, institutional autonomy, or ethical commitments. At the same time, it creates a clear and auditable pathway from authoritative microdata to aggregated indicators. This makes it possible to demonstrate the system's capabilities through concrete empirical applications. The following section presents such an application, applying the SKCMDb infrastructure to the computation of local content indicators derived from official Slovak music charts.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

6 Case Study: Computing Local Content Indicators from Official Charts

This section presents an empirical application of the SKCMDb infrastructure to demonstrate how diversity- and locality-related indicators can be computed reproducibly, lawfully, and without exposing sensitive personal data. The case study applies the methodological, governance, and architectural principles developed in the preceding sections to an authoritative, real-world dataset: the official Slovak music charts. Its purpose is not to evaluate individual actors or enforce compliance, but to show how a shared reference infrastructure can support informed, voluntary, and legally robust engagement with cultural policy objectives.

6.1.1 Case study: Slovak charts as an empirical testbed

The empirical case study is based on the ingestion and evaluation of two official Slovak chart datasets: **SK – RADIO – TOP 100** and **SK – SINGLES DIGITAL – TOP 100**, using the 50th chart week as a reference point. These charts were added to the Slovak Comprehensive Music Database (SKCMDb) as complete, unfiltered corpora and evaluated against the two criteria defined in §215 of the Slovak Media Services Act: (a) authorship-based eligibility and (b) language-based eligibility.

Both charts were treated as bounded, authoritative reference sets that reflect actual repertoire circulation in different usage contexts. The radio chart represents editorially curated broadcast output, while the digital singles chart aggregates performance across multiple formats, including downloads and streaming. Although statutory quota obligations apply only to radio broadcasting, analysing both datasets allows comparison between regulated and unregulated segments of the music ecosystem.

Evaluation results

For the **SK – RADIO – TOP 100**, the evaluation showed that **13%** of the chart entries met the statutory definition of a Slovak musical work. Of these, approximately **10%** qualified on the basis of Slovak-language lyrics, while a further **3%** qualified through authorship-based eligibility despite being performed in English. This result falls significantly below the statutory quota thresholds defined for radio broadcasters.

By contrast, in the **SK – SINGLES DIGITAL – TOP 100**, the share of Slovak-eligible recordings reached **approximately 25%**, thereby meeting the quota threshold—even though no such obligation applies to digital or streaming distribution. In this dataset, eligibility was driven almost entirely by Slovak-language recordings, with comparatively few cases of authorship-based qualification without Slovak lyrics.

This pattern mirrors observations from other national contexts. In Hungary, for example, MAHASZ reports domestic versus international sales breakdowns that consistently show a higher share of local repertoire in overall consumption than in radio airplay. Taken together, these findings suggest that there is **room for growth in Slovak radio programming** with respect to domestic repertoire, and that current broadcast practices may not fully reflect audience demand or the available pool of eligible works.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Interpretation limits

The case study does not assess compliance by individual broadcasters, nor does it attempt to explain causal mechanisms behind programming choices. The results are descriptive and indicative. Their purpose is to demonstrate that quota-relevant indicators can be computed transparently and reproducibly from shared reference data, and that differences between broadcast and consumption contexts can be made visible without accessing playlists or monitoring individual actors.

Data sources, legal criteria, and observability

The Slovak charts used in this case study are curated and published by *Česká národní skupina Mezinárodní federace hudebního průmyslu, z. s. (ČNS IFPI)*, which represents Czech and Slovak record labels and acts as the official chart producer for both Czechia and Slovakia. As part of its role in supporting domestic repertoires, ČNS IFPI produces national and domestic charts for industry benchmarking, media use, and public communication. This role, however, should not be conflated with regulatory authority. ČNS IFPI does not operate under the mandate of the Slovak Media Services Act, nor is it empowered to determine compliance with statutory quota obligations.

Determining whether a recording fulfils the legal definition of a "Slovak musical work" would require the collection and processing of personal data, most notably information on the current or former permanent residence of authors or co-authors. Such data processing is neither required nor permitted within the scope of chart production. Consequently, while the charts constitute an authoritative and widely recognised industry reference, they cannot function as legally determinative instruments for quota implementation or monitoring.

Under §215 of the Media Services Act, a Slovak musical work is defined as a work that either (a) has at least one composer or lyricist with a current or former permanent residence in the Slovak Republic, or (b) contains lyrics in the Slovak language. These criteria operate at different conceptual levels: authorship and residence are properties of the musical work and its creators, while language is a property of the performed recording. Crucially, neither criterion is fully observable in routine industry or broadcast metadata. Information on authors' residence constitutes personal data protected under the GDPR and is not publicly available. Language, while generally recognisable to a native Slovak speaker, also presents edge cases, including deliberate mixtures of Slovak and Czech (for example in "Czechoslovak rap") or traditional repertoires using Western Slavic dialects that diverge from contemporary standardised Slovak.

Language classification in this case study was therefore not treated as a purely automatic attribute. Recordings were cross-checked by a native-speaker poet and songwriter, Terézia Fracerová, who validated whether songs were performed in Slovak, Czech, or other languages. This expert curation step was used to resolve ambiguous cases and to ensure that language-based eligibility assessments reflect linguistic practice rather than surface assumptions or platform metadata alone.

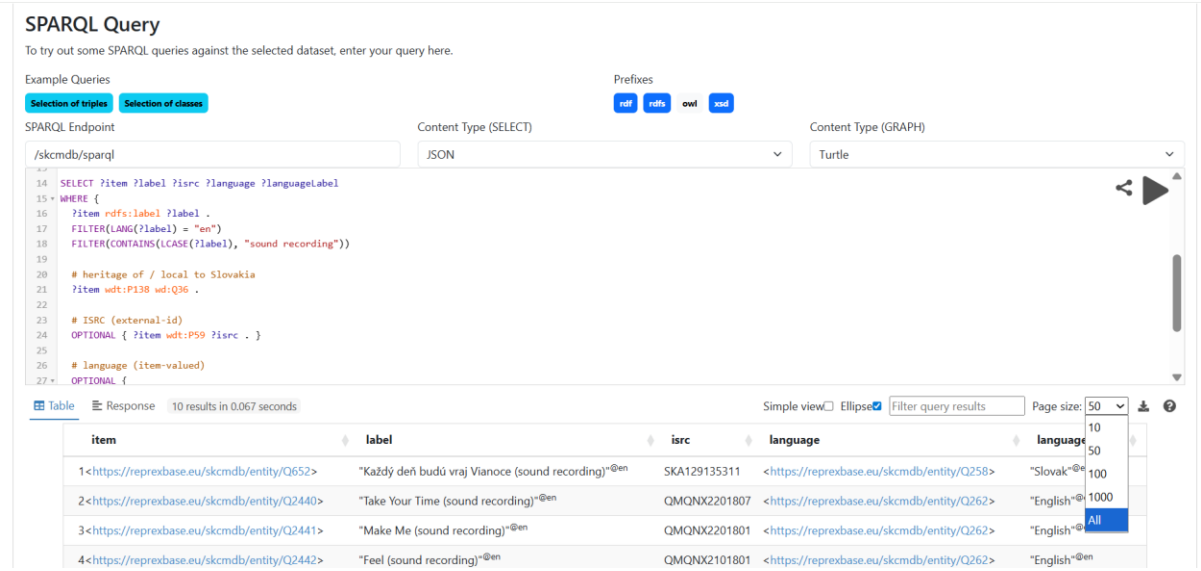
How the system works in practice

Operationally, the case study follows a reproducible workflow implemented within SKCMDb. Each chart entry is first resolved to a **sound recording**, using standard identifiers such as ISRC where available, complemented by platform identifiers and curated matches where necessary. Recordings are then linked to their underlying **musical works**, identified through rights management registries

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

(notably SOZA–CISNet and ISWC data), as eligibility under §215 partly depends on authorship attributes that belong to the work rather than the recording.

Once recordings and works are linked, eligibility criteria are evaluated internally. If at least one criterion defined by the law is met—either language-based or authorship-based—the recording is marked as eligible at the artefact level. Importantly, the system does not record or expose which specific criterion triggered eligibility, nor which individual contributor satisfied it. The outcome is expressed solely as a property of the recording, ensuring consistency across co-authored works while avoiding person-level disclosure.



We can provide API level access to contracted participants to the https://hudobnadatabaza.sk/ system.

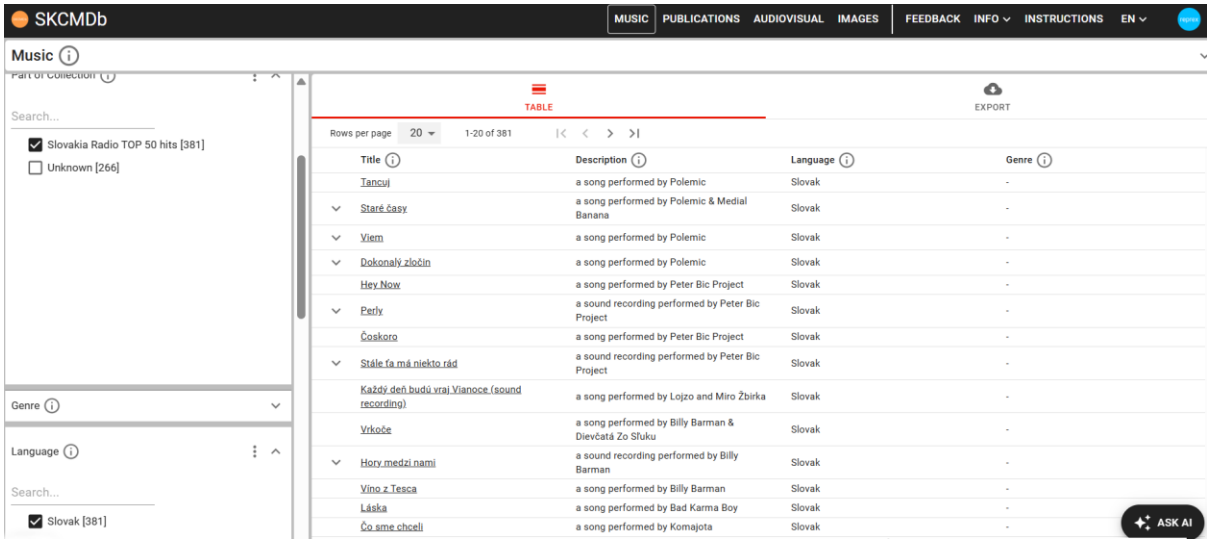
The results of this evaluation can be retrieved through predefined queries. As demonstrated in the Annex, a SPARQL query returns a structured table containing titles, identifiers, language classification, and quota eligibility status, which can be downloaded and used as a reference or filtering table for playlists stored in common formats such as text files, CSV, or spreadsheets.

From case study to continuous service

Beyond the one-off analysis, the case study demonstrates how the approach can be scaled into a continuous, incremental service. New recordings enter circulation every week, and the value of the system increases as additional charts, playlists, and releases are added. A logical service workflow would allow broadcasters, curators, or cultural institutions to upload playlists or repertoire lists. Where a recording is already present in SKCMDb, quota eligibility and language attributes can be resolved immediately. Where it is not, a curation workflow can be triggered to establish identifiers, link recordings to works, and evaluate eligibility.

Over time, this incremental process reduces identification and transaction costs and progressively builds a more comprehensive reference database. Such a service would be relevant not only for radio broadcasters, but also for festival programmers, rights managers, cultural policymakers, the Slovak Music Centre, and the Music Fund, all of whom share an interest in tracking repertoire dynamics and improving the visibility of Slovak music domestically and internationally.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295



Our semantic browser uses dynamic queries with simple, faceted searches that can narrow down the list of known recordings that fulfil the Slovak quota. This service is currently available freely, and we plan to add an even more intuitive, chatbot like interface to query the database.

6.1.2 Aggregation and observatory-level representation

Eligibility outcomes were aggregated at two temporal resolutions: **weekly**, corresponding to individual chart publications, and **annual**, corresponding to calendar-year summaries. For each period, the proportion of chart entries qualifying under the Slovak-language and Slovak-origin indicators was computed.

These aggregated results were then expressed as **statistical data cubes** within the central modules of the Open Music Observatory. Time was modelled explicitly as a dimension, with measured values representing the respective shares. By lifting the results from item-level evaluations into a statistical representation, the case study demonstrates how national-level reference data can contribute to supra-national observatory functions without transferring underlying microdata or sensitive attributes.

Importantly, the observatory-level representation contains no information that would allow the reconstruction of personal characteristics or editorial decisions at the level of individual works or contributors. What circulates upward is a set of interpretable indicators, not the data used to compute them.

6.1.3 Interpretation limits and policy relevance

The purpose of this case study is not to assess compliance by individual broadcasters, nor to evaluate the effectiveness of Slovak cultural policy. The indicators presented are *descriptive*, not normative, and they do not imply causality. Their function is to demonstrate that diversity- and locality-related metrics can be computed in a manner that is auditable, reproducible, and compatible with European data protection and non-discrimination principles.

From a policy perspective, the case study illustrates how authoritative industry data, rights-mediated evaluation, and federated data governance can be combined to support evidence-informed discussion without creating new surveillance or enforcement mechanisms. From a methodological

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

perspective, it shows that the same infrastructure can support both fine-grained microdata curation and high-level statistical aggregation, depending on the level at which questions are asked.

In this sense, the charts analysis serves as a *proof of implementation* rather than as a standalone analytical result. It demonstrates that the SKCMDb architecture is capable of translating abstract policy criteria into computable indicators, while preserving institutional autonomy, protecting sensitive data, and remaining interoperable within a European data space. It shows that our approach can make a future EU Culture Compass and EU Culture Data Hub serve very high-granularity, tailored monitoring tools for various local, national or EU-wide policies.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

7 Discussion and Outlook

This paper set out to address a persistent gap between the ambitions of diversity-oriented cultural policy and the practical realities of its implementation in contemporary music ecosystems. Through the development and empirical demonstration of the *Slovak Comprehensive Music Database*, embedded within the Open Music Observatory framework, we have shown that this gap is not primarily a matter of regulatory design or editorial goodwill, but of missing data infrastructures capable of translating policy criteria into computable, ethically governed outcomes.

A central contribution of this work lies in reframing local content quotas and diversity targets as information design problems rather than enforcement challenges. By distinguishing clearly between cultural artefacts and persons, and by separating internal evaluation from external disclosure, the SKCMDb approach demonstrates that complex eligibility rules can be operationalised without exposing sensitive personal data or violating principles of non-discrimination. This reframing is particularly important in the European context, where cultural policy objectives must coexist with strong data protection regimes and fundamental rights protections.

The case study based on official Slovak music charts illustrates how this approach functions in practice. By relying on an authoritative, neutral reference corpus, the analysis avoids the pitfalls of monitoring or evaluating individual broadcasters, while still demonstrating how policy-relevant indicators can be computed reproducibly and transparently. The resulting indicators are not instruments of enforcement, but shared points of reference that can support voluntary compliance, editorial reflection, and evidence-informed discussion across the music ecosystem. In this sense, the charts analysis serves as a proof of implementation rather than a policy verdict.

At the same time, the broader architecture of SKCMDb highlights the importance of subsidiarity and pluralism in cultural data governance. The Slovak and Hungarian implementations, together with the Livonian and Karelian community-curated databases, demonstrate that methodological coherence does not require uniform analytical outputs. National broadcast monitoring, ethnomusicological competence questions, and community heritage stewardship can coexist within a shared framework, provided that governance, semantics, and responsibilities are clearly articulated. This pluralistic design reflects both the diversity of European cultural production and the institutional realities of the cultural sector.

The supply-side infrastructure developed through SKCMDb is complemented by demand-side analysis in OpenMusE Deliverable D3.2 *Report on Music, Society and Citizenship*, which presents representative survey data on music discovery behaviour, consumption patterns, and participation in musical life. Where this deliverable focuses on enabling broadcasters and cultural intermediaries to compute and report diversity indicators, D3.2 examines how citizens experience cultural diversity in their music engagement. The Slovak survey data linked D3.2 provides important context for interpreting the circulation patterns observed in official charts, revealing audience preferences, discovery mechanisms, and barriers to accessing diverse repertoires. Together, these deliverables provide supply-side (what is distributed) and demand-side (what is consumed) perspectives on cultural policy assessment.

Several limitations of the present work should be acknowledged. First, the indicators demonstrated here are necessarily reductive abstractions of complex cultural phenomena. They do not capture

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

qualitative dimensions of musical innovation, artistic value, or audience engagement, nor do they address causal relationships between policy interventions and cultural outcomes. Second, the reliance on authoritative institutional data implies a dependency on existing infrastructures, which may vary in completeness and accessibility across countries and sectors. Finally, while the approach is designed to be extensible, its successful replication depends on the willingness of competent institutions to engage in shared governance and to invest in sustainable data stewardship. Despite these limitations, the implications of this work extend beyond the immediate Slovak context. The methodological and architectural principles outlined here are applicable to other cultural domains and policy areas in which diversity, representation, and fairness intersect with large-scale data processing. In particular, they offer a pathway for integrating cultural policy objectives into algorithmic environments—such as recommender systems and analytics platforms—without reproducing the extractive or surveillance-oriented practices often associated with global digital platforms.

Looking forward, this work contributes to the ongoing development of the *Open Music Observatory* as a prototype for a future European Music Observatory within the EU Culture Data Hub. Its diversity pillar, of which SKCMDb constitutes a substantial part, demonstrates how European values of data governance—subsidiarity, proportionality, data sovereignty, and ethical stewardship—can be operationalised in practice. The inclusion of Finno-Ugric musical heritage, extending beyond the geographical boundaries of Europe, further underscores that “European” in this context denotes a governance philosophy rather than a territorial limitation. Future research and development efforts will focus on expanding the range of indicators, refining rights-aware modelling, and exploring integrations with emerging AI governance frameworks. In doing so, the aim is not to replace editorial judgement or cultural expertise, but to augment them with infrastructures that make diversity goals computable, auditable, and actionable under real-world conditions. By treating cultural policy as a shared infrastructural challenge rather than a zero-sum regulatory contest, the approach presented here offers a constructive path toward more inclusive, resilient, and transparent cultural ecosystems in Europe and beyond.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

8 References

Antal, Daniel. 2020a. “Feasibility Study on Promoting Slovak Music in Slovakia & Abroad.” https://doi.org/10.5281/zenodo.6427514.

———. 2020b. “Projektová štúdia propagácie slovenskej hudby na Slovensku a v zahraničí [Feasibility Study On Promoting Slovak Music In Slovakia & Abroad].” Translated by Dáša Bulíková. https://doi.org/https://doi.org/10.5281/zenodo.6427556.

———. 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.” Open Music Observatory. https://doi.org/10.5281/zenodo.17078911.

———. 2024c. “SKCMBDb: Building the Slovak Comprehensive Music Database in the Slovak Music Centre.” Open Music Observatory. https://doi.org/10.5281/zenodo.18036465.

———. 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. “A Green Paper on AI, Data Governance, and Metadata Policies for Europe’s Music Ecosystem.” Open Music Observatory. https://doi.org/10.5281/zenodo.17244314.

———. 2025b. “Graph-Based Representation of a Recording–Work Reconciliation Problem in a Federated Music Data Space.” Figshare. https://doi.org/10.6084/m9.figshare.30928814.

———. 2025c. “Open Music Observatory.” Open Music Observatory. https://doi.org/10.5281/zenodo.17897724.

———. 2025d. “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.

———. 2025e. “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.

———. 2025f. “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, Kata Gabor, Ieva Pigozne, and Bogáta Tímár. 2025. “Federating Open Knowledge Through Wikibase: The Case of the Finno-Ugric Data Sharing Space.” Finno-Ugric Data Sharing Space. https://doi.org/10.5281/zenodo.17938081.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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].” In *Proceedings of the Digital Humanities in the Nordic and Baltic Countries (DHNB) Conference*. Vol. 7. CEUR-WS.

Antal, Daniel, and Reprex. 2023. “Open Music Europe Diversity Pillar Work in Slovakia Repository.” OpenMusE Consortium. October 24, 2023. https://github.com/dataobservatory-eu/listen-local-slovakia.

Antal, Daniel, Fabio Vieira, and Anna Márta Mester. 2025. “Open Music Europe Recorded Music Datasets.” Open Music Observatory. https://doi.org/10.5281/zenodo.17985665.

Antal, Daniel, and Natália Zagyva. 2025. “Enriching and Futureproofing the Databases of the Hungarian Heritage House.” Open Music Observatory. https://doi.org/10.5281/zenodo.17759777.

Archives Expert Group on Archival Description, International Council on. 2023. “Records in Contexts–Conceptual Model. Version 1.0.” International Council on Archives. https://www.ica.org/app/uploads/2023/12/RiC-CM-1.0.pdf.

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.

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.

BITSS. 2019. “Guidelines for Open Policy Analysis.” Berkeley Initiative for Transparency in the Social Sciences. http://www.bitss.org/wp-content/uploads/2019/03/OPA-Guidelines.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.

Carroll, Stephanie Russo, Ibrahim Garba, Oscar Luis Figueroa-Rodríguez, Jarita Holbrook, Raymond Lovett, Simeon Materechera, Mark Parsons, et al. 2020. “The CARE Principles for Indigenous Data Governance.” *Data Science Journal* 19 (1): 1–12. https://doi.org/10.5334/dsj-2020-043.

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.

Core, Dublin. 2020. “DCMI Metadata Terms.” http://dublincore.org/specifications/dublin-core/dcmi-terms/2020-01-20/.

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

European Commission. 2017. “European Interoperability Framework – Implementation Strategy. Communication from the Commission to the European Parliament the Council the European Economic and Social Committee and the Committee of the Regions.” https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52017DC0134.

———. 2020. “A European Strategy for Data.” Communication from the Commission. European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020DC0066.

European Commission, Directorate-General for Communications Networks, Content and Technology, Crowell&Moring, IMC University of Applied Sciences Krems, Philippe Rixhon Associates, Technopolis Group, and UCLouvain. 2022. *Study on Copyright and New Technologies – Copyright Data Management and Artificial Intelligence*. Publications Office of the European Union. https://doi.org/doi/10.2759/570559.

European Parliament and Council. 2016. “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation).” *Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2016/679/oj.

———. 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).” *Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng.

European Parliament, Council of the European Union, and European Commission. 2025. “Draft Joint Declaration ‘Europe for Culture – Culture for Europe’.” Brussels: European Commission. https://culture.ec.europa.eu.

Ikkala, Esko, Eero Hyvönen, Heikki Rantala, and Mikko Koho. 2019. “Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces.” *Semantic Web* 10 (1): 69–84. https://doi.org/10.3233/SW-180303.

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.” International Organization for Standardization. https://www.iso.org/standard/44292.html.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

———. 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.” 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.

———. 2022. “International Standard Musical Work Code (ISWC). ISO 15707:2022.” 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.” International Organization for Standardization. https://www.iso.org/standard/74296.html.

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

J, Wilson. 2015. *Evidence-Based Policy Making in the European Commission*. Edited by Elisabeth Lannoo. CIC Report 7440. Oslo (Norway): CICERO Centre for International Climate; Environmental Research. http://www.cicero.uio.no/en/posts/news/report-from-science-to-policy-how-to-improve-the-dialogue/.

Keller, Paul. 2025. “Publishing Cultural Heritage Data in the Age of AI.” Policy report. Open Future; Europeana. https://openfuture.eu/wp-content/uploads/2025/12/251202PublishingCulturalHeritageDataInTheAgeOfAI.pdf.

Kitzes, Justin, Daniel Turek, and Fatma Deniz, eds. 2018. *The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences*. 1st ed. University of California Press. http://www.practicereproducibleresearch.org/.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Media Services Act. 2022. “Zákon č. 264/2022 Z. z. o mediálnych službách a o zmene a doplnení niektorých zákonov (zákon o mediálnych službách)” *Official Collection of Laws of the Slovak Republic.* https://static.slov-lex.sk/static/SK/ZZ/2022/264/20251201.html.

Mikš, Tomáš. 2025. “OpenMusE: Towards a Sustainable Licensing Market for AI Use of Protected Works.” Vilnius, Lithuania: Slovak Performing; Mechanical Rights Society (SOZA); OpenMusE Consortium; Presentation at the CISAC European Committee Meeting. https://doi.org/10.5281/zenodo.17944096.

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/.

Open Music Europe. 2023. “Open Music Europe (OpenMusE) – An Open, Scalable, Data-to-Policy Pipeline for European Music Ecosystems.” https://doi.org/10.3030/101095295.

Open Music Europe Consortium. 2023. “Open Music Europe User Stories: OPA-Compliant User Stories for System Competences and Semantic Architecture.” https://github.com/dataobservatory-eu/open-music-europe-user-stories/. https://github.com/dataobservatory-eu/open-music-europe-user-stories/.

Partanen, Niko, Philippe Rixhon, Karīna Bandere, Jānis Ziediņš, Pawan Kumar Dutt, Matīss Bolšteins, Matias Frosterus, et al. 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/.

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

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.

Sganga, Caterina, Magali Contardi, and Luigi Cernigliaro. 2025. “Report on Music Diversity and Circulation: OpenMusE Deliverable D2.2.” Horizon Europe Deliverable D2.2. Open Music Europe (OpenMusE). https://www.openmuse.eu/.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

A Data Used

A.1 Personal Data

Personal data was received in eight Excel files based on explicit GDPR consent. These files are not available directly, but the persons' data was worked into the Slovak Comprehensive Music Database.

Some of this personal data is also available on an opt-out basis to support SOZA's collective rights management activities, and reduce the cost and potential error rate of distributing the artist's royalties. The opt-out data is handled by a separate agreement between SOZA and Reprex.

Generally, names are treated in two ways: in the graph database, they are used as a *label* of the graph node that represents the artist; and either as a full name, or as given and surname, they are added separately to this graph.

The *label* is always a preferred name, whenever possible, based on our MoU with the Slovak national library, we use the preferred name of SNK's authority file. In the alias field, we collect known surface forms of the name to match music service or bibliographical records.

For example, the 19th century composer's name is often spelled as *Franz Liszt*, *Liszt Ferencz* (contemporary 19th century spelling) or *Liszt Ferenc*, and as *Liszt, F*. Liszt was one of the most famous musicians of the Austro-Hungarian dual monarchy, and he lived and worked in the current territory of the Slovak Republic, Hungary, and Austria. According to the customs of Austria-Hungary, he used his name in a German and Austrian form.

Currently, the Slovak national authority files and the German authority files use *Liszt, Franz* as the preferred name. The Hungarian, Estonian, Czech, and Lithuanian national systems prefer the Hungarian *Liszt, Ferenc* version. The Russian national library uses *Liszt, F*, or its Cyrillic transliteration.

In the 21 century, it is less usual to have translated given names, however, it remains a fact that Hungarian names, including minority Hungarian names in Slovakia are following the Eastern name order: *Kurtág Ferenc*, which would be *Ferenc Kurtág* in the Western name order, and libraries in Hungary and most of the world would use the *Kurtág, Ferenc* unambiguous ordering.

We see two modelling solutions for this problem, explicitly writing in *Ferenc* as a given name, and indicating that its historical form is *Ferencz* and historically it was often translated to *Franz*, and treat *Liszt* as a surname. This aligns most with the generally used FOAF ontology.

The other solution, which may replace or complement the previous solution is the use of several labels and aliases. Our system has labels for every language, therefore, in the Slovak interface, the main label is *Liszt, Franz,* while in the Hungarian it is *Liszt, Ferenc.*

In this case, for interoperability between Slovak and Hungarian sources, we use *Liszt, Franz* as a preferred label on the Slovak interface, and *Liszt, Ferenc* as an alias (alternative name); in the Hungarian language interface, we do the opposite. In ontological terms, *"Liszt, Franz"@sk* is the rdfs:label for Slovak language use and *"Liszt, Ferenc"@hu* is the rdfs:label for Hungarian language use.

Last but not least, artists may chose to detach their private and public personas and use pseudonyms. In this case, unless we are explicitly consented, we only reveal the pseudonym of a

living artist. This aligns with the practice of authority files in national libraries, CISAC and IFPI affiliates.

Given name

variable name	given_name
description	Given name of the person consenting
ontological alignment	omo:P22 - aligned with FOAF
note	

Surname

variable name	surname
description	Family name of the person consenting
ontological alignment	omo:P22 - aligned with FOAF
note	

A.1.1 Location and nationality

We do not collect, unless consented, data on ethnicity. However, for the purpose of Slovak cultural policy and media regulation, we need to know if the works of a person meet the local content guidelines, which is derived from the nationality, place of birth or residence status of the artist.

As explained in Section 5, by principle, we do not expose this sensitive data, but we infer if the person's works meet the local content regulation.

- In case of people, we create a special location property which states that the person is local to Slovakia.
- In the case of works, we apply a special property, is heritage of Slovakia; or is heritage of Slovaks.

Historically, many Czech, Austrian and Hungarian nationals lived in Slovakia, and currently the Slovak Republic provides temporary protection to many Ukrainian (and other) nationals. Regardless of ethnicity, these people may fall under the local content guidelines, and for this media regulatory purpose we claim that they are local to Slovakia. We use qualifiers to ensure that this is not part of their personal identity, but a legal fact about their compositions.

In the case of works, a work and its recorded fixation meets the Slovak local content regulation if it uses the Slovak language, or it is composed by a person who meets the local content regulation's definitions, regardless of ethnicity or nationality. When we claim that a work is a heritage of "Slovaks", we use it in a narrow sense: for composed works, we refer to Slovaks as an ethnolinguistic group who use the Slovak language. The heritage of "Slovakia" is any work composed by qualifying persons.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Last, but not least, non-vocal authentic folk music of Slovaks is also a heritage of Slovakia. In this case, to avoid misunderstandings, we use “authentic Slovak folk music” as a genre, i.e., a property of the music, and not its unknown composer or performer.

City

variable name	adress_city
description	The city where the person currently resides
ontological alignment	location omo:P22 - aligned with DCTERMS, RiC, CIDOC
note	

Date of birth

variable name	date_of_birth
description	Date of birth
ontological alignment	date of birth omo:P22 - aligned with FOAF with qualifier
note	

Country of birth

variable name	country of birth
description	The country where the person was born
ontological alignment	location omo:P73 - with qualifier, aligned with DCTERMS, RiC, CIDOC
note	An auxiliary variable to establish is heritage of/local to Slovakia, i.e., if the person’s compositions belong to the Slovak local quota.

Lived in Slovakia

variable name	has_lived_in_Slovakia
description	The person at some point lived in Slovakia
note	An auxiliary variable to establish is heritage of/local to Slovakia, i.e., if the person’s compositions belong to the Slovak local quota.

nationality

variable name	nationality

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

description	The person's nationality
ontological alignment	location omo:P73 - with qualifier, aligned with DCTERMS, RiC, CIDOC
note	An auxiliary variable to establish is heritage of/local to Slovakia or one of its regions, i.e., if the person's compositions belong to the Slovak local quota. Because the nationality connects the artist with a geopolitical entity, we use a special form of location; the ethnonym of "Slovaks" is reserved for the Slovak ethnolinguistic community.

Place of birth

variable name	place_of_birth
description	The person at some point lived in Slovakia
note	An auxiliary variable to establish is heritage of/local to Slovakia or one of its regions, i.e., if the person's compositions belong to the Slovak local quota or to a locally curated music collection.

Sex assigned at birth

variable name	sex_at_birth
description	The sex assigned at birth: female or male.
ontological alignment	sex or gender omo:P23 - aligned with Wikidata, SDMX.
note	SKCMDb can handle non-binary gender identities, but SOZA, following Slovak administrative practice, only has the judicially recognised sex available; the artist can ask the variable to be changed.

IPI name number

variable name	IPI_name_number
description	Interested Parties Information Name Code
ontological alignment	IPI name number omo:P42 - aligned the CisNet database's internal, globally harmonised proprietary identifier.
note	A unique identifying number assigned by the CISAC database to each Interested Party in collective rights management. An artist may have multiple IPI numbers for multiple pseudonyms or name variations.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

IPI name for

variable name	IPI_name_for
description	The name that the IPI number is attached to
ontological alignment	aligned the CisNet database's internal, globally harmonised proprietary identifier
note	A unique identifying string assigned by the CISAC database to each Interested Party in collective rights management. An artist may have multiple IPI numbers for multiple pseudonyms or name variations.

Pseudonym

variable name	pseudonym
description	a pseudonym used instead of the artist's name
ontological alignment	label or rdfs:label
note	If the person wishes to use a pseudonym, the person's record will be labelled with this name.

A.1.2 Repertoire Data

A.1.3 Sound recordings

A **sound recording** is the actual captured sound: a specific audio performance fixed in a sound file or master. Under the ISRC (International Standard Recording Code) standard, each distinct sound recording is identified by a unique ISRC. Different edits, remasters, or versions of the same performance typically receive different ISRCs, because they represent different sound recordings (ISO 2019a).

A.1.4 Recording Title

variable name	recording_title
description	Title of the sound recording as provided in metadata
ontology alignment	title
notes	May be multilingual; language tags are applied when known

Normalised Recording Title (Identifier-style Alias)

variable name	normalised_recording_title
description	Deterministically normalised alias of the recording title
ontology alignment	—
notes	Derived from recording_title by applying ASCII-only, uppercase normalization for identifier-style matching and reconciliation. This

	variable is an alias of the recording title, not a distinct work or title statement, and carries no ontological role. It is intended solely for technical matching, joining, and comparison, not for presentation or authoritative naming.

Artist Name (Display Artist)

variable name	artist_name
description	Presentation artist credited by Spotify
ontology alignment	creator
notes	Used only as a descriptive field; ontological roles must be assigned through curated agent statements

Normalised Artist Name (Identifier-style Alias)

variable name	normalised_artist_name
description	Deterministically normalised alias of the Spotify display artist name
ontology alignment	—
notes	Derived from artist_name by applying ASCII-only, uppercase normalization for identifier-style matching and reconciliation. This variable is an alias, not a separate artistic identity, and carries no ontological role. It is intended for technical matching, joining, and comparison only, not for presentation or authoritative naming.

Recording Lyrics Language

The recording lyrics language is only added to the Slovak, Livonian, and Karelian datasets.

variable name	recording_lyrics_language
description	Principal language of the lyrics in the recording
ontology alignment	language of work or name
notes	Empty or marked appropriately for instrumental recordings

Duration

variable name	duration
description	Duration of the sound recording
ontology alignment	duration
notes	Stored as milliseconds; semantically corresponds to ISO 8601 duration format

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

A.1.5 Identifiers

International Standard Recording Code

variable name	isrc
description	International Standard Recording Code
ontology alignment	ISRC, defined according to ISO 3901:2019 International Standard Recording Code
notes	Primary, globally unique identifier of the sound recording

Country of ISRC Registration

variable name	isrc_country_code
description	Country code of the national ISRC issuing agency that allocated the registrant code
ontology alignment	ISO 3166-1 alpha-2 location qualified with location applies to: country of ISRC registration
notes	This does not indicate the place of recording, artist nationality, or rights jurisdiction; and left NA (missing) if the code was not assigned by a national registration agency.

ISRC Registrant Code

For details on ISRC registration please see *International Standard Recording Code (ISRC} Handbook. 4th Edition.* (International ISRC Registration Authority 2021)

variable name	isrc_registrant_code
description	Registrant code assigned by the national ISRC agency identifying the organisation registering the recording
ontology alignment	identifier (string) qualified with identifier type: ISRC registrant code
notes	Distinguishes the label or distributor authorised to assign ISRCs

Year of ISRC Registration

variable name	isrc_reference_year
description	Two-digit year field from ISRC indicating the year in which the registrant assigned the code
ontology alignment	has time qualified with temporal role: date of ISRC registration, precision: year
notes	This is the ISRC assignment year, not the recording year or release year

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

ISRC Country Code

The ISRC country code does not refer to a country as a geopolitical entity, but to an organisation that registers sound recordings with a territorial jurisdiction. For example, *Association of Hungarian Record Companies* (Mahasz, omo:Q2035), the Hungarian national IPI Group member, handles the ISRC codes starting with HU. In this respect, the HU part of the ISRC code does not refer to Hungary but to Mahasz.

In most cases, HU and other ISRC country codes coincide with the country code of the jurisdiction, but several large countries ran out of their identifier range, and have multiple top-level codes, like the U.S., Brazil or Germany; and ISRC reserved some top-level registration codes for technical purposes, too.

variable name	isrc_country_code
description	Two-character ISRC prefix identifying the allocating authority
ontology alignment	identifier (string) omo:P37, qualified with identifier type omo:P39 as ISRC registrant code omo:Q2037.
notes	Extracted from the first two characters of the ISRC. Identifies the allocating authority according to ISO 3901 and IFPI practices. This does not necessarily correspond to the country of the artist, label, or release.

ISRC Registration Country

For example, *Association of Hungarian Record Companies* (Mahasz, omo:Q2035), handles the registrations on the territory of Hungary. Whenever the top-level registrar corresponds with a geopolitical entity, we placed the entity's name in this variable, including non-sovereign entities like Hong Kong or French Southern Territories. The non-territorial codes are marked with *No agency [exception]*.

Our aim was the study of Slovak and Hungarian registrations in and outside of their registrars, and the algorithms we used collected a lot of identifiers that fell out of the scope of these countries.

variable name	isrc_registration_country
description	Resolved allocating authority associated with the ISRC prefix
ontology alignment	—
notes	Indicates the national ISRC agency's location, or a non-standard or unassigned category. This field reflects allocation practice, not distributor, label, or rights ownership.

ISRC Designation Code

This range usually identifies a label or a rights management company that handles recording-side royalties. The reason why they cannot be directly mapped to labels is that certain designation codes were allocated to labels that were dissolved, or merged into newer entities, or have other

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

successors. They can be seen as a provenance variable for the particular registration. Unfortunately, we know that often the same recording receives new ISRC codes when the previous registering body or its administration recards are no longer available.

variable name	isrc_designation_code
description	Registrant-internal sequential identifier that distinguishes recordings assigned in the same year
ontology alignment	identifier (string) qualified with identifier type: ISRC designation code
notes	Unique within a registrant and year; not a globally unique identifier

Spotify Track ID

variable name	spotify_track_id
description	Platform-specific identifier for the recording on Spotify
ontology alignment	Spotify track ID
notes	Used as an access point; not a persistent identifier and not legally or semantically authoritative

A.1.6 Release

A **release** is a packaged publication of one or more recordings made available to the public. In DDEX terminology, a release groups recordings together in a specific commercial or distribution context, such as an album, single, EP, or digital compilation. The same recording can appear on multiple releases (for example, an album release, a single release, and a compilation), while retaining the same ISRC.

Unfortunately, as our research shows, new releases often receive a new ISRC code.

Release Year

variable name	release_year
description	Year in which the release became publicly available
ontology alignment	has time qualified with temporal role: date of release, precision: year
notes	Derived from release date for analysis and grouping

Release Name

variable name	release_title
description	Title of the release containing the recording
ontology alignment	title of the release entity

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

notes	Refers to the release, not the recording

Normalised Release Name (Identifier-style Alias)

variable name	normalised_release_title
description	Deterministically normalised alias of the release name
ontology alignment	—
notes	Derived from the release name by applying ASCII-only, uppercase normalization for identifier-style matching and reconciliation. This variable is an alias of the release name, not a distinct release entity or title statement, and carries no ontological role. It is intended exclusively for technical matching, joining, and comparison, not for presentation or authoritative naming.

Release Type

variable name	release_type
description	Classification of release as album, single, or compilation
ontology alignment	Maps to subclasses of release in the Annex (single, album, compilation)
notes	Follows the release categories used in Spotify metadata

UPC or EAN

The *Open Music Europe – ISRC-Sampled Sound Recording Metadata* does not contain release identifiers, and not in all cases in the smaller datasets.

variable name	upc
description	Universal Product Code or European Article Number identifying the release
ontology alignment	UPC or EAN
notes	Identifies the release entity, not the recording

Label Name

We provide this information when it is given. It is not provided in the Open Music Europe – ISRC-Sampled Sound Recording Metadata, and not in all cases in the smaller datasets.

variable name	label_name
description	Label or imprint text supplied via Spotify metadata
ontology alignment	record label

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

notes	This is a descriptive imprint only; rights-bearing organisations must be identified separately

Label Identifier (Curated)

This information is not given in the case of the *Open Music Europe – ISRC-Sampled Sound Recording Metadata*, and not in all cases in the smaller datasets.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

B Queries

1. Query: recordings that fit into the national quota

```
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd:  <http://www.w3.org/2001/XMLSchema#>
PREFIX wikibase: <http://wikiba.se/ontology#>

PREFIX wd:   <https://reprexbase.eu/skcmdb/entity/>
PREFIX wdt:  <https://reprexbase.eu/skcmdb/prop/direct/>
PREFIX s:    <https://reprexbase.eu/skcmdb/entity/statement/>
PREFIX p:    <https://reprexbase.eu/skcmdb/prop/>
PREFIX ps:   <https://reprexbase.eu/skcmdb/prop/statement/>
PREFIX pq:   <https://reprexbase.eu/skcmdb/prop/qualifier/>
PREFIX pr:   <https://reprexbase.eu/skcmdb/prop/reference/>

SELECT ?item ?label ?isrc ?language ?languageLabel
WHERE {
 ?item rdfs:label ?label .
 FILTER(LANG(?label) = "en")
 FILTER(CONTAINS(LCASE(?label), "sound recording"))

 # heritage of / local to Slovakia
 ?item wdt:P138 wd:Q36 .

 # ISRC (external-id)
 OPTIONAL { ?item wdt:P59 ?isrc . }

 # language (item-valued)
 OPTIONAL {
  ?item wdt:P37 ?language .
  ?language rdfs:label ?languageLabel .
  FILTER(LANG(?languageLabel) = "en")
 }
}
LIMIT 10
```

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

C Governance

C.1 Data Sharing Space Memorandum of Understanding

SOZA is a civic association of authors and publishers of musical works that has for more than a hundred years executed collective management of copyright within the territory of the Slovak Republic, including ensuring the process of registration of musical works. In registration and identification of musical compositions, SOZA plays a similar role as SNL.

Slovak Music Centre (MC) is a contributory organization established by the Ministry of Culture of the Slovak Republic, whose mission is to support Slovak music culture, including the documentation of Slovak music. MC has its own library and cooperates with music libraries as part of the Slovak National Group of IAML (International Association of Music Libraries, Archives and Documentation Centers) both at home and abroad. REPREX is a company specializing in knowledge and data management, cooperating with SOZA, MC, and MF and providing parties with services in the field of information science.

SNL is the Slovak national library, which plays a key role in the authoritative control of works and authors, and is part of the VIAF consortium, enabling international coordination of authoritative names, terms, and preferred names of various concepts and institutions.

MF is a public institution with its main mission to support creative activity in the field of music art. The Fund creates material conditions for the development of all music genres by providing financial resources to music creators and organizations. In its own publishing house, MF publishes and produces music carriers of Slovak composers and concert artists, and operates a brick-and-mortar shop and an e-shop Musica Slovaca with a selection of sheet music, books, and CDs by Slovak publishers presenting Slovak creation and performing art.

SOZA, MC, REPREX, and SNL entered into a Memorandum of Understanding regarding the Slovak Comprehensive Music Database in 2024 (“Original Memorandum”). MF has expressed interest to collaborate with the other parties in creating and ensuring the functionality of the Slovak Comprehensive Music Database and, for this purpose, to join the Original Memorandum. The Parties thus intend, following the expression of MF’s intent, to amend their contractual relationship accordingly. SOZA, MC, REPREX, and MF are collaborating to create the Slovak Music Dataspace. The dataspace is a coordination mechanism of data that enables libraries, other memory institutions, rights organizations, and other stakeholders to exchange trusted data in a legal, automated, and efficient way. This collaboration is expanded to SNL and other interested Slovak libraries possessing music collections or relevant items related to music, musical works, recordings, and musicians in a broad sense (composers, performers, lyricists, music theorists, etc.).

SOZA, REPREX, and MC are working on a research and development project supported by the EU under the Horizon Europe program (https://cordis.europa.eu/project/id/101095295) in accordance with a Memorandum of Understanding signed with the Ministry of Culture of the Slovak Republic (https://www.crz.gov.sk/zmluva/7645338).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

C.1.1 Visibility

The Parties wish to make Slovak music, musical works and recordings, including books about music, sheet music, CDs and other carriers and knowledge about music more visible to the public so that they are easier to access, search, read, and borrow.

The Parties collaborate on the Slovak Music Dataspace and a linked, publicly accessible database—the Slovak Comprehensive Music Database, accessible on https://hudobnadatabaza.sk —with the aim of increasing the precision of information about music and improving accessibility of sheet music, sound recordings, books, and other documents related to music in Slovak libraries and other public collections in the Slovak Republic.

The Parties aim to link information on the availability of Slovak music—sheet music, sound recordings, books, and other music documents available at home and abroad—using persistent global identifiers and other identifiers employing library authoritative control. The Parties seek ways to aggregate certain Slovak cultural artifacts related to music on the Europeana portal (https://www.europeana.eu/sk) and to aggregate data, when appropriate, to www.wikidata.org, an open knowledge graph used by many national libraries, libraries, and music service providers for authority control and data exchange.

C.1.2 Information Management

The Parties cooperate in informing Slovak libraries and librarians about new information and data techniques, and disseminate innovations and knowledge brought to Slovakia through international collaboration, especially innovative services that may attract new users to public libraries or help improve their services.

The Parties cooperate in public awareness campaigns about the availability of modern library services and in increasing the visibility and discoverability of physical and digitized objects related to music in Slovak music libraries, libraries with significant music collections, and the SNL itself.

C.1.3 Authority Control and Mutual Cooperation

Within the cooperation, the Parties undertake to harmonize identifiers of music, musical works, composers, performers, pedagogues, and musicologists, in such a way that it is possible to compare knowledge about their works and bibliographies.

In accordance with the MF's mission under Act No. 13/1993 Coll., Article 2, para. (2), MF purposefully and objectively supports, among other things, creative artistic activity in accordance with Slovakia's national and cultural interests. The MF Statute, created on the basis of Article 2, para. (4) of Act No. 13/1993 Coll., specifies the content and manner of MF's mission and activities, including publishing activity as a means of realizing its purposes.

To ensure and support MF's intentions through publishing activity, by which its main mission is fulfilled, MF has decided to participate in the creation of the Slovak Comprehensive Music Database by providing selected data and information collected and processed within its activity and mission.

According to the Establishment Deed of MC, Article I point 2, as a state contributory organization established by the Ministry of Culture of the Slovak Republic, the core activity of MC includes collecting, professionally processing, analyzing, and providing to the public information on professional music creativity and live professional music culture. Under Article I point 3 lit. c) of the

Establishment Deed of MC, MC ensures the systematic documentation of professional music art in all genres and coordinates building the central database of information about live professional music culture in Slovakia.

To ensure and support MC's primary activity and associated tasks through systematic documentation across all genres and the coordination of a central information database about live professional music, MC has decided to participate in the creation of the Slovak Comprehensive Music Database by providing selected data and information collected and processed within its activity.

According to Article 3 point 3.3 of the SOZA Statutes, SOZA's main activity includes, among other things, supporting the development of musical creativity in Slovakia and preserving cultural values in the field of music art. Under relevant provisions of the Contract for Representation of Rights Holders, SOZA as a collective management organization collects and processes personal data of right holders for the purpose of collective management, and gathers other data for the performance of collective management in selected areas where protected musical works are used. To ensure and support SOZA's activity in supporting the development of musical creativity and preserving cultural values as well as increasing the efficiency of collective management performance, SOZA has decided to participate in the creation of the Slovak Comprehensive Music Database by providing selected data and information collected and processed within its activity and mission.

MC, MF, SOZA, and REPREX will collaborate, among other things, on creating a special database of authoritative links and identifiers of persons related to music, including links to Wikidata, ISNI, VIAF, and identifiers from foreign national libraries. SNL will create VIAF records for Slovak physical and juridical persons (including SOZA members) appearing in Slovak music collections and holdings but so far unknown to SNL.

SNL and MC will encourage Slovak libraries to share their internal identifiers, and if they use such, their online catalogue URLs or URIs, or catalogue URLs for Slovakian works they have in their collection and to add VIAF or other international identifiers for catalogue and service interoperability.

The cooperation remains open to any other interested parties, especially for online music service providers to provide URLs or other identification keys to their services. However, if the cost such integration is significant, they may be asked to contribute to the costs of data integration or data maintenance. This Memorandum shall become valid and effective on the date of its signature by all parties and fully replaces the original memorandum.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

C.2 Proportionality test for accessing the controller’s legitimate interest

prepared in accordance with Article 6 (1) (f) of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

(hereinafter referred to as the "Regulation")

and in accordance with Section 13 (1) (f) of Act No. 18/2018 Coll. on the protection of personal data

(hereinafter referred to as the "Act")

(hereinafter referred to as the "Proportionality Test")

Controller:

Name: Slovenský ochranný zväz autorský pre práva k hudobným dielam

Address: Rastislavova 3, 821 08 Bratislava 2

ID No.: 00178454

(hereinafter referred to as the "Controller")

C.2.1 Introduction

Within the performance of its activities, which is the performance of collective management of rights to musical works within the meaning of Act No. 185/2015 Coll. on the Copyright Act, as amended (hereinafter referred to as the “Copyright Act”), the Controller has decided to participate in the creation and provision of functionality of the Slovak Comprehensive Music Database (hereinafter referred to as the “SKCMDB”) and the Slovak Music Dataspace (hereinafter referred to as the “SMDS”). SKCMDB and SMDS represent data coordination mechanisms that aim to enable selected organizations, institutions and other interested parties to exchange trustworthy data in a legal, automated and efficient manner.

The above-mentioned participation of the Controller is part of a research and development project supported by the EU under the Horizon Europe program (https://cordis.europa.eu/project/id/101095295) in accordance with the Memorandum of Understanding signed with the Ministry of Culture of the Slovak Republic (https://www.crz.gov.sk/zmluva/7645338/).

Under the current legal framework, any processing of personal data must be subject to one of the legal bases referred to in Article 6 (1) of the Regulation.

The Controller is interested in processing personal data on the basis of a legitimate interest pursuant to Article 6 (1) (f) of the Regulation.

In order to meet the conditions for the use of legitimate interest as a legal basis for the processing of personal data, the Controller has prepared this proportionality test.

The controller answered the following questions as part of the proportionality test:

1. What is the purpose of the personal data processed? What does the controller pursue by processing personal data and what is its interest?
2. Is the Controller's interest in processing personal data legitimate?
3. Is the processing of personal data necessary on the basis of a legitimate interest? Can the pursued objective not be achieved otherwise?

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

4. Could the data subject reasonably expect that the processing of their personal data might occur?
5. Has the Controller fulfilled its information obligation towards the data subject?

C.2.2 Identification of legitimate interest

Regarding question 1)
The purpose of processing personal data is:

- improving and streamlining the performance of collective rights management carried out by the Controller for the right holders represented by it,
- increasing the accuracy and availability of information about the musical creation of the right holders represented by the Controller,
- promoting Slovak music, musical works, musical recordings, knowledge about Slovak music in such a way that they are more easily accessible and searchable, and
- in particular promoting the authors of musical works and music publishers represented by the Controller.

The legitimate interest of the Controller lies in the economic interests of the rightsholders represented by it, namely through the collective management by the Controller of selected economic rights of these rightsholders, i.e. the provision of selected information about the represented rightsholders through SKCMDB and SMDS, to entities relevant to the economic interests of the represented rightsholders. The legitimate interest of the Controller is therefore implemented through the controlled disclosure and dissemination of selected information about the represented rightsholders, which, among other things, promotes them and supports the realization of their right to freely pursue entrepreneurial activity (which also includes the performance of authorial activities).

The controlled disclosure and dissemination of selected information about represented rightsholders is intended to bring:

- on the part of represented rightsholders (authors, heirs, publishers, etc.) an increase in public interest in their work through the availability of selected information and its promotion, which is expected to increase the use of their repertoire of musical works, and

- on the part of relevant entities (users of musical works) an increase in the conditions for timely and proper identification of the use of musical works, in other words, a decrease in the probability of erroneous identification or the impossibility of carrying out the identification of such use, which contributes to the growth of remuneration collected for the use made, to an increase in the accuracy of the accounting of the collected remuneration and the growth of the income of represented rightsholders. This minimizes situations where the use of musical works is identified incorrectly or not identified at all, which has a positive impact on the described growth in the collection of remuneration, more accurate identification of the use of works and the subsequent growth of the income via royalties of represented rightsholders.

Regarding question 2)
Legitimacy of the interest pursued by the Controller

According to the interpretation of the Office for Personal Data Protection of the Slovak Republic, a legitimate interest cannot be the processing of personal data that is unethical, illegitimate or contrary to good morals.

SOZA as the Controller publishes on its website, specifically in the section entitled "Membership Base", among other things, a list of living and deceased authors represented by it. The list in question contains information about living and deceased authors represented, namely the name and surname. By publishing the list of represented rightsholders in question on its website, SOZA fulfils its legal obligation pursuant to the provisions of Section 181, letter (m) of the Copyright Act.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

SOZA plans to provide SKCMDB and SMDS with and subsequently make available and disseminate through it selected information about the rightsholders represented by it in the scope of name, surname, pseudonym, date of birth, place of permanent residence and gender. Selected information in the scope of name and surname will be publicly accessible to any SKCMDB and SMDS user. Selected information in the scope of date of birth, place of permanent residence and gender will be accessible only to selected SKCMDB and SMDS users.

In the case of selected information in the scope of name, surname, date of birth, place of permanent residence and gender, the Controller derives the legitimacy of its interest in further processing primarily from the Authorization for the exercise of collective management No. MK-1247/2022-214/4460 from September 02 2022 issued by the Ministry of Culture of the Slovak Republic, valid at the time of preparation of the proportionality test, on the basis of which the Controller performs collective management in selected fields of collective management (hereinafter referred to as the "Authorization"). The selected information in the above scope is already collected by SOZA directly from the represented rightsholders themselves and processed exclusively for the purposes of collective management of their economic rights. The collection and processing in question is carried out based on the consent granted by the represented rightsholder in the representation agreement concluded with in the following wording:

"The personal data that the Rightsholder has provided and will provide in the future to SOZA in connection with the performance of the Agreement is SOZA authorized to collect, process and use exclusively for the purposes of collective management of their economic rights and fulfilling the purpose of this Agreement. SOZA is authorized to provide, make available and publish these personal data to the extent necessary for the collective management of economic rights to users, foreign collective management organizations with which it has concluded representation agreements, including international databases, and associations of collective management organizations with which it has a contractual or membership relationship. The Rightsholder declares that, prior to concluding this Agreement, they were informed by SOZA about the information pursuant to Article 13 of Regulation (EU) 2016/679 and Act No. 18/2018 Coll. on the protection of personal data. The information pursuant to the previous sentence is available to the Rightsholder on the website www.soza.sk."

In this respect, the legitimacy of the processing is given by the fact that the handling of the data is related to the exercise of collective rights management itself and the nature of this activity.

In view of the above, the Controller considers that the controlled disclosure and dissemination of selected information about the rightsholders represented by it to the extent specified within the SKCMDB and SMDS is legitimate. In addition, the disclosure and dissemination of selected information about the rightsholders represented by it to the extent of name and surname is already possible via the intermediary's website. The disclosure and dissemination of other selected information occur in a controlled manner, only to selected users of SKCMDB and SMDS. For the reasons stated above, the disclosure and dissemination of selected information represent only a minimal interference with rights of the represented rightsholders.

Regarding question 3)
Test of necessity of using legitimate interest as a legal basis for processing of personal data
Processing is necessary when the purpose of the Controller's processing cannot be achieved in another, less invasive way or when this purpose could be achieved in another way but would require a disproportionate effort.

As for the question of whether the Controller could subject the processing of personal data to another legal basis, the only one that would theoretically come into consideration would be consent under Article 6 (1) (a) of the Regulation. However, in practice, given the volume of information, the need for economic and administrative efficiency, obtaining consent is unfeasible, and so the use of this legal basis is practically impossible. Moreover, the Controller has tried in the past to subject the processing of personal data to the legal basis of consent under Article 6 (1) (a) of the Regulation, however, obtaining it in practice, given the number of rightsholders represented by SOZA, has actually proven to be unfeasible.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

The Controller therefore takes the liberty of primarily pointing out the inefficiency and administrative complexity in obtaining and registering consents.

In addition, the planned controlled disclosure and dissemination of selected information about represented rightsholders within the SKCMDB and SMDS will be carried out with the aim of increasing the efficiency and improving the performance of collective management by SOZA, so the legitimacy of the interest lies not only with SOZA but also with the represented rightsholders themselves.

Regarding question 4)
Balancing test
The legitimate interests of the Controller or a third party must not be overridden by the interests or fundamental rights and freedoms of the natural persons concerned. In order to determine whether the legitimate interests of the Controller are an overriding element compared to the interests of the persons concerned, the Controller has carried out a balancing test, which has focused on assessing in particular the following questions:

a) Could the data subject reasonably expect that their published personal data would be processed?

The mandatory publication of certain personal data is provided for by special regulations. It is an expression of the Slovak Republic's interest in transparency and informing public about essential facts that may have an impact on the rights and legally protected interests of individuals or society as a whole.

The activities of collective management organizations are subject to stricter rules to maintain transparency towards rightsholders, but also towards users. Users can verify whether SOZA represents the repertoire of a given rightsholder by searching for this information on the SOZA's website. According to the provision of Section 181 of the Copyright Act, SOZA is obliged to publish *"a list of rightsholders represented pursuant to Section 164 (1) in the territory of the Slovak Republic and their number"*. Each entity can thus find out who SOZA represents.

In the opinion of the Controller, the data subject may expect that this published information will be further processed for the purposes of collective management.

b) Is there a previous relationship between the Controller and the data subject?

A previous relationship exists between the Controller and the represented rightsholder. This relationship is given by a contractual relationship, the legal basis of which is the concluded representation agreement.

c) Does the original purpose of the processing correspond to the purpose of the processing defined by the Controller and its legitimate interest?

One of the legitimate interests of the Controller in this processing of personal data is the intention to increase the efficiency and improve the performance of collective management, as well as the promotion of represented rightsholders and their authorial activities. The purpose of the processing is closely related to the interest of the Controller in the most efficient performance of collective management.

d) What is the impact of the processing of personal data on the data subjects and on the Controller?

When assessing the impact of the intended processing of personal data of represented rightsholders, the Controller primarily relies on the fact that the processed selected information (personal data) in the scope of the name and surname has already been published once. The legal protection of the represented rightsholders has already been significantly weakened. The Controller believes that by further processing this selected information (personal data) about a natural person (represented rightsholders) it can no longer interfere more with their rights, as has already happened.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

As for the intended processing of selected information (personal data) of represented rightsholders in the scope of pseudonym, date of birth, place of permanent residence and gender, the Controller itself relies primarily on the fact that the processed selected information (personal data) in the scope mentioned has already been obtained by the Controller from the represented rightsholders on the basis of consent and processes it exclusively for the purposes of exercising collective management of economic rights. Moreover, most of this information (personal data) has already been made public, as in most cases it concerns publicly known and active persons. The Controller considers that the further processing of this selected information (personal data) about a natural person (represented rightsholders) interferes with their rights in a minor way.

The WP29 Working Party also considers the fact that the personal data of the data subjects have already been made public to be an important factor in carrying out the balancing test. The Controller does not carry out any operations that would treat the already published personal data of the represented rightsholders in a disparaging or inappropriate manner, nor does it in any way harass the data subjects in connection with the processed personal data. Although the misuse of this personal data by third parties cannot be completely excluded or eliminated, the Controller considers that the data subjects have already been exposed to this risk and could occur regardless of the activity of the Controller. Based on the above, the Controller considers that the processing of personal data as described above may only represent a minimal interference with the interests, fundamental rights and fundamental freedoms of the data subjects.

The aim of the balancing test is not to prevent any interference with the rights and freedoms of the data subjects – in other words, the balancing test may be in favour of the Controller even if a certain impact on the data subjects is established. The aim is only to prevent a disproportionate impact on the data subjects.

In the context of the proportionality test, it is also important to determine the scope of the processed personal data, in particular with a focus on whether sensitive personal data of a special category are processed, such as data revealing racial or ethnic origin, etc. Under the terms of the Controller, such personal data are not processed.

As for the duration of the publication of selected information about represented rightsholders:

Given the pursued legitimate interest of the Controller in connection with the establishment of SKCMDB and SMDS, which is to improve and streamline the performance of collective management and at the same time promote Slovak music, the duration of the publication of selected information about represented rightsholders is two-stage, of which each stage has its own basis constituting a legitimate interest together with the corresponding duration of the publication of the information in question.

For the legitimate interest of improving and streamlining the performance of collective management, the publication period represents the duration of the representation agreement concluded between the represented rightsholder and the Controller, but no more than the life of the represented rightsholder and 70 years after their death. The duration of the relevant period of publication of selected information about represented rightsholders for the aforementioned legitimate interest is derived from the duration of economic rights pursuant to the relevant provisions of the Copyright Act. The basis for the aforementioned legitimate interest is the Authorization.

For the legitimate interest of promoting Slovak music, the period of publication is a period based on the main subject of activity of SOZA, which, according to the relevant provision of its Statutes, is, among other things, assisting the development of music in the Slovak Republic and supporting the preservation of cultural values in the field of musical art. The duration of the relevant period of publication of selected information about represented rightsholders for the aforementioned legitimate interest is derived from the duration of the performance of the main subject of activity of SOZA. The basis for the aforementioned legitimate interest is the Statutes and the Authorization.

In accordance with the above, the duration of the period of publication of selected information about represented rightsholders within SKCMDB and SMDS is the duration of the performance of the main subject of activity based on the Authorization.

Regarding question 5)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Has the Controller fulfilled its information obligation towards the data subjects?
The Controller has informed the represented rightsholders (data subjects) about the conditions for processing their personal data on its website in the “Personal Data” section, by information sent in the form of a so-called “newsletter”, as well as by a notice sent to each represented rightsholder using the information system for authors called “ISpA”.

The Regulation grants each data subject the right to object to the processing of their personal data and the Controller has informed the data subjects about this right and the method of its implementation.

C.2.3 Has the Controller taken additional security measures in connection with the processing of personal data?

The Controller protects personal data by placing them in a protected area, to which only a limited group of persons has access.

Only persons authorized by the Controller who are duly informed of their rights and obligations in the area of personal data processing (hereinafter referred to as "Authorized persons") have access to selected information about represented right holders (personal data) in SKCMDB and SMDS, including name, surname, pseudonym, date of birth, place of permanent residence and gender. Authorized persons are also bound by the obligation of confidentiality regarding personal data that they have come into contact with in the performance of their activities, and this obligation continues even after the termination of their employment. The Controller also ensures that Authorized persons are provided with appropriate training in the area of personal data protection.

Selected entities other than Authorized persons have access to selected information about represented rightsholders (personal data) in SKCMDB and SMDS in the scope of name, surname, pseudonym, date of birth, place of permanent residence and gender. Such access is controlled and the selection of entities is subject to the decision of the Controller. These entities are also duly informed of their rights and obligations in the area of personal data processing.

The services of an intermediary selected by the Controller and meeting all technical and security requirements are used to process selected information about represented rightsholders.

In the event of a personal data breach, Authorized persons, selected entities other than Authorized persons and the intermediary are informed of the procedure for reporting a personal data breach, in order to promptly adopt preventive or corrective measures. All security incidents, regardless of their scope, are recorded in the Controller's records.

C.2.4 Evaluation of the proportionality test

As part of the proportionality test, the Controller:

a) Identified its legitimate interest and the legitimate interest of third parties in the processing of personal data as follows:

- **improving and streamlining the performance of collective rights management carried out by the Controller for the rightsholders represented by it,**
- **increasing the accuracy and availability of information on the music of the rightsholders represented by the Controller,**
- **promoting Slovak music, musical works, musical recordings, knowledge about Slovak music in such a way that they are more easily accessible and searchable, and**
- **special promotion of authors of musical works and publishers represented by the Controller.**

b) Assessed that the processing of personal data on the basis of legitimate interest is necessary, since the pursued objective could otherwise be achieved only with disproportionate effort, which would in no way provide greater legal protection for the published personal data.

c) Conducted a balancing test, assessing on the one hand the weight of the legitimate interests, fundamental rights and freedoms of the data subjects, for whom the greatest interference occurred without the activity of the Controller by the first mandatory disclosure of their personal data, as these are often publicly known persons, and on the other hand the weight of the legitimate interests of the Controller and third parties.

d) Has assessed that the rights and freedoms of the data subject do not outweigh the legitimate interests of the Controller.

Based on the above, in accordance with the applicable legal regulations and taking into account the supporting legal documents used, the Controller has made a decision that it has a legitimate interest in processing personal data as described above and, pursuant to Article 6 (1) (f) of the Regulation, is entitled to process selected personal data of the data subjects on the legal basis of legitimate interest.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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

D.1 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 offering a new reading of something that never becomes "final." The same is true of sound recordings: some are rediscovered decades later, remastered, and brought back into circulation for new audiences.

Music assets, in other words, have an unusually long lifecycle. This is equally 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 alongside 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.

This *Green Paper* builds on a substantial body of prior European policy analysis concerning copyright data, metadata interoperability, and the implications of new technologies for the cultural and creative sectors. In particular, it draws on the *Study on copyright and new technologies: copyright data management and artificial intelligence*, commissioned by the European Commission (DG CNECT) and published in 2022[37].

That study identified persistent structural weaknesses in rights metadata management across creative industries, including the music sector, despite long-standing identifier systems and standardisation efforts. These weaknesses include fragmentation between identifiers and registries (ISRC, ISWC, ISNI, VIAF, IPI), limited interoperability between sectoral systems, high administrative costs in rights management, and difficulties in ensuring authoritative, machine-readable information across the full lifecycle of creative works and recordings.

The study further noted that emerging technologies, including artificial intelligence, are increasingly deployed in environments where metadata quality and legal clarity are insufficient, amplifying

[37] *Study on copyright and new technologies: copyright data management and artificial intelligence* (SMART 2019/0038) (European Commission et al. 2022)

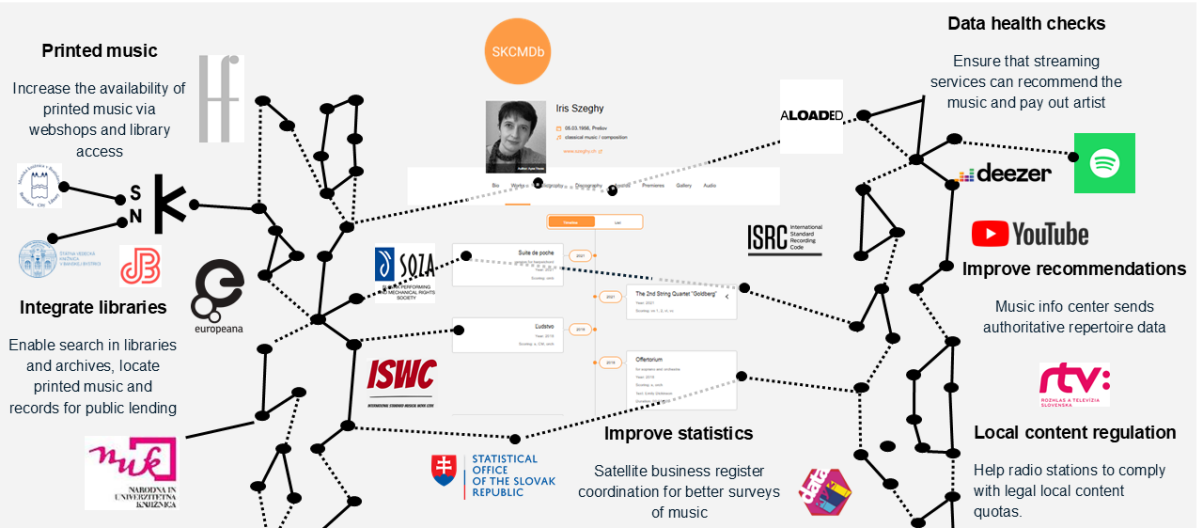
existing inefficiencies and legal uncertainties. It therefore emphasised the need for lifecycle-aware metadata governance, improved integration of existing standards and identifiers, and institutional arrangements that enable trust, transparency, and interoperability rather than further centralisation. These findings form an important analytical baseline for the policy considerations and implementation examples discussed in this Green Paper.

> ⓘ Why this Green Paper matters for music professionals?
>
> - Streaming has centralised power in platforms, while leaving rightsholders with micro-royalties and growing administrative burdens.
> - Metadata mistakes translate directly into lost revenue — each unlinked ISRC or ISWC is money left unclaimed.
> - AI is already reshaping music ecosystems: it can either support documentation and remuneration, or flood systems with untracked works.
> - Europe needs federated, cooperative solutions so independents, collective management organisations, and heritage institutions 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 may later be reconciled with an identifier; a work may be linked to a new performance; a recording may be embedded in new formats or platforms. Each act of documentation adds layers of meaning and makes music intelligible in new environments.

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

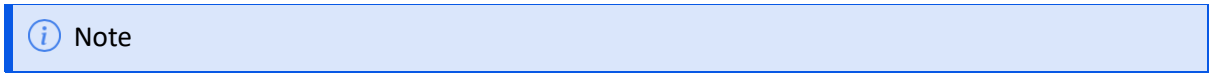
Data sharing and exchange in the Slovak music data sharing space



© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

In Slovakia, we implemented, using the European Interoperability Framework for shared digital services, public-private service integration to make music more visible and its handling more cost efficient. Our implementation aligns well with the requirements set by the Copyright Infrastructure Task Force.

The invention of the computer, and later the internet, introduced new ways to document and transmit music. These innovations brought powerful efficiencies: identifiers such as the ISRC and ISWC, digital distribution pipelines, and networked catalogues enabled the global circulation of music at unprecedented scale. At the same time, they produced new fragmentation. Standards proliferated, identifiers failed to interconnect, and workflows designed for one purpose often broke down in another. What was intended as progress frequently resulted in overlapping, incompatible, or incomplete metadata — a legacy that now requires systematic repair.



© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

This *Green Paper* is a maturing policy document developed within the **Open Music Europe (OpenMusE)** Horizon Europe **Research and Innovation Action** (Grant Agreement No. 101095295). It deliberately combines policy research with implementation piloting, reflecting the project's emphasis on innovation, experimentation, and real-world validation rather than abstract policy design alone.

Prepared in line with the **Guidelines for Open Policy Analysis** (available at https://www.bitss.org/opa/community-standards/) and the **Horizon Europe Data Management Guidelines**, the document has been released early to support consultation, incorporate stakeholder input, and ensure transparency throughout its development. It extends the analysis developed in the first OpenMusE policy brief on music metadata mainstreaming and EU law (Deliverable D5.6), and its core findings are further condensed in the second policy brief (Deliverable D5.7), which integrates wider stakeholder consultations[38].

Transparency note: This *Green Paper* does not present a purely conceptual or speculative policy proposal. Substantial parts of the analysis synthesise lessons from policy-embedded implementation activities conducted using Open Policy Analysis methods, including openly documented data, code, workflows, and methodological decisions. These activities were carried out in cooperation with public authorities and sectoral institutions, allowing the paper to reflect both policy ambitions and real-world constraints.

In accordance with Open Policy Analysis principles[39], all related deliverables and technical documentation are publicly accessible to foster engagement and ensure a clear audit trail. 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.

This document situates the **Open Music Observatory** as a central reference point. The Observatory is a prototype of a modern European Music Observatory developed by the OpenMusE consortium, currently populated with data on economy, diversity, society, and innovation, and operating multiple federated modules. Technical documentation and versioned DOIs are available via Zenodo, with an overview at https://openmusicobservatory.eu/.

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[40].

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.[41] This is an early version (0.9.3.)

[38] 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,

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

- Big Data Value Association, Gaia-X: Dataweek[24]: Introducing a new European music dataspace[42]
- Echoes/ECCH: The ECHOES (European Cloud for Heritage Open Science) policy even and workshop, 12-13 December 2024 at the Royal Institute for Cultural Heritage (KIK-IRPA), Brussels,
- Hungarian stakeholders interested in replication of the Slovak pilot versions in various meetings throughout 2024 and 2025[43].
- CISAC: Protecting Creators' Rights in the AI Era: OpenMusE at the European Committee Meeting, Vilnius, 29-30 April [44].
- The Fair MusE - Prelude to a fairer music industry Fair MusE project[45]
- IAMIC [46]: The *International Association of Music Information Centres* and several key members of the organisation.

IAMIC, IAML, FairMusE, Music360, ECCCH forums, among others) and translates them into policy actions for EU institutions.

[39] The *Guidelines for Open Policy Analysis* form an actionable and practical set of directives that the OpenMusE consoritum was mandated to use under the Grant Agreement. It can be seen as good implementation framework for *Evidence-based policy making in the European Commission* and *The practice of reproducible research* [BITSS (2019); Open Music Europe (2023); (J 2015; Kitzes, Turek, and Deniz 2018).

[40] 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.

[41] 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.

[42] Jun 5, 2024, Dataweek[24], Leuven, Belgium.

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

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

[45] 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/

[46] 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 2024c, 2024d).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

- IAML: The *International Association of Music Libraries, Archives and Documentation Centers* and several national chapters and key members [47].
- 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. At this stage, we started to record user stories for the design of system competences of the Open Music Europe software ecosystem and the observatory's semantic architecture[48].
- Music Futures: the AHRC Creative Industries Cluster project MusicFutures in the United Kingdom.
- Slovak national stakeholders interested in cultural data.[49]
- Wikimedia community and developers[50].
- European music industry stakeholders on *LineCheck 2025* [51]

This *Green Paper* is aligned with the European Commission's *Culture Compass for Europe* and its ambition to strengthen shared European cultural data infrastructures. It proceeds from the premise that cultural data challenges are largely shared across domains, but that effective implementation requires domain-specific, federated units where sectoral complexity exceeds what generic frameworks can capture. Music is one such domain.

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.

---

[47] 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 2025c, 2025d).

[48] The user stories were placed in an OPA compliant folder on GitHub. This work was not carried on after the initiated grant agreement change, when the WP4 did no longer participate in building open-source software for this purpose, but we did not abandon the competency requirements (Open Music Europe Consortium 2023).

[49] 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).

[50] 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 coordination of our open source technology, data curation and dissemination efforts. (Antal 2024b, 2025e; Antal, Pigozne, and Federico 2025).

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

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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[52].

[52] 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

D.2 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.[53]

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.[54]

ⓘ Transparency Note

[53] *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.

[54] *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).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

This Green Paper was developed as part of the *Open Music Europe* Horizon Europe Research and Innovation Action and reflects a deliberate methodological choice: to integrate policy analysis with implementation piloting rather than treating them as separate phases. The project combined research, stakeholder consultation, and hands-on experimentation to test how European music data infrastructures can be built incrementally, under real institutional constraints, and across multiple governance levels.

The Open Music Observatory (OMO) is both a research output and an operational prototype. During the project, the consortium implemented:

- a **national music data sharing space** (the Slovak Comprehensive Music Database), available at https://hudobnadatabaza.sk/en/
- a **partial national replication** in another Member State (Hungary),
- a **sub-national and cultural-identity–based module** (the Finno-Ugric Data Sharing Space), available at https://finnougric.net/en/, which also shows that like the EU Culture Data Hub, it is not practical to work with music as an isolated domain, and
- a **pan-European core module** aggregating datasets, indicators, and studies, available at https://openmusicobservatory.eu/

Together, these pilots demonstrate how a federated, decentralised model can function across supranational, national, regional, and community levels without forcing centralisation. They show how research concepts such as interoperability, provenance, FAIR data, and subsidiarity translate into governance agreements, workflows, and technical interfaces.

This implementation-driven approach responds directly to long-standing European policy calls for a European Music Observatory, including the *European Parliament Resolution on Cultural Diversity and the Conditions for Authors in the European Music Streaming Market* (P9_TA(2024)0020), and aligns with emerging frameworks such as the European Interoperability Framework, the EU Data Strategy, the European Open Science Cloud, and the European Collaborative Cloud for Cultural Heritage. The Green Paper therefore serves both as a policy synthesis and as a documented reference implementation that can inform future European action.

The European Commission's feasibility study for the establishment of a European Music Observatory explicitly highlighted CEEMID as a best-practice example of decentralised, open, and reproducible music data integration, recommending that its approach be further explored during the start-up phase of a future Observatory[55]. The *Open Music Observatory* can therefore be understood as a direct response to this recommendation, updated to reflect subsequent European data space, copyright infrastructure, and AI policy developments.

[55] *Measuring and Reporting Regional Economic Value Added, National Income and Employment by the Music Industry in a Creative Industries Perspective. Memorandum of Understanding to Create a Regional Music Database to Support Professional National Reporting, Economic Valuation and a Regional Music Study* (Artisjus et al. 2014) and *Central And Eastern European Music Industry Report (2020)* (Antal 2020).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Beyond metadata coordination, a key ambition of the Slovak pilot was to address a deeper policy problem: the absence of music-specific economic evidence produced in line with best statistical practice. The pilot was explicitly designed around established methods for business satellite accounts, with the aim of harmonising authoritative collective rights management data with administrative registers, industry datasets, and statistical surveys. This approach sought to generate music-specific statistical indicators that Slovakia's existing cultural and creative satellite accounting system cannot provide, and which are unavailable in most EU Member States due to the absence of dedicated cultural or creative satellite accounts altogether.

In practical terms, this meant treating rights management data—often the most complete and economically meaningful source of information on music activity—not merely as sectoral metadata, but as a statistical input capable of feeding official economic indicators when properly harmonised and governed. By linking these data sources while preserving methodological transparency, reproducibility, and comparability, the pilot anticipated a model in which music could become visible in official economic statistics without requiring centralised data extraction. Although this experiment was not implemented within the OpenMusE project due to resource constraints and changing institutional conditions, its design remains methodologically sound and operationally feasible, and is directly aligned with current efforts to develop the *EU Cultural Data Hub* as a foundation for evidence-based cultural policy[56].

In line with the European Commission's *Culture Compass for Europe*, this Green Paper argues that Europe's cultural data future must be built on shared, interoperable infrastructures. At the same time, it demonstrates that certain cultural domains — notably music — require specialised, federated implementation layers due to the density of rights, frequency of reuse, and exposure to algorithmic systems. Music therefore provides a strategically valuable reference case for implementing the Culture Compass's data and AI ambitions in practice.

From the outset of this work, we have taken the view that Europe should avoid building siloed cultural data infrastructures by domain. Wherever possible, cultural data should be shared across domains through common principles, identifiers, and governance frameworks. Domain-specific solutions should emerge only where generic infrastructures reach their limits.

This brings us to a third major contribution to this landscape in the form of the *Copyright Infrastructure Task Force*, a voluntary cooperation to solve one aspect of the puzzle with willing member states and stakeholders. We see the *Copyright Infrastructure Task Force* as a possible vehicle to carry on some of our findings, and also a role model to find music's specialist space in the European cultural infrastructure, including the Cultural Compass, the EU Cultural Data Hub, Europeana and its new data space and the European Collaborative Culture Cloud.

Music is one of the domains where these limits are reached earliest and most visibly, because of its very complex rights management infrastructure, and the extremely high number of events and digital artefacts related to it.

[56] *Pilot Program for Novel Music Industry Statistical Indicators in the Slovak Republic* (Antal 2023). The pilot was not implemented within the OpenMusE project due to resource constraints and institutional changes, but its methodological design remains valid and suitable for continuation in the context of EU Cultural Data Hub preparatory work.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

The feasibility study of the *Common European Data Space for Cultural Heritage* (ECCCH) focuses primarily on digitisation, discovery, preservation, and reuse across cultural heritage domains. This cross-domain orientation is both necessary and appropriate at the foundational level. However, the study does not substantially address music as a domain, nor the specific challenges of music rights, reuse intensity, or lifecycle-driven metadata mutation.

This omission is not accidental, but illustrative: the generic cultural heritage layer cannot realistically encode the sector-specific legal, economic, and technical logic required for music. At this point, neither Europeana or the ECCCH's first draft ontology can handle music's copyright rights management and attribution system. Music needs to find its own platform to bring to the cross-domain infrastructure its own data needs.

In 2025 the *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[57].

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

2. **Practical solutions**, grounded in dialogue between research and industry, and inspired by concrete experiences with open, federated data-sharing approaches.
3. **Potential pitfalls** where well-meaning initiatives may clash with legacy systems, existing business practices, or contradictions in legislation.
4. **Legal and operational conflicts**, such as the tension between GDPR's data protection regime and the Berne Convention's requirement of author attribution.
5. **Cooperation and workflow sharing**, recognising that no single actor can bear the full burden of metadata documentation.
6. **Technology**, including automation, entity recognition, reconciliation, and persistent identifiers.
7. **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.

D.2.1 Three structural pressures

Three structural pressures frame today's metadata challenges:

[57] *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)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

8. **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.

9. **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.

10. **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[58], 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. A recent policy paper of *Open Future* and *Europeana* made the same diagnosis in a wider, not music domain specific context: heritage institutions are being forced to publish data into AI-driven reuse environments before governance, provenance, and rights frameworks are ready[59]. The music domain digitised early and has plenty of experience with this increasing problem.

D.2.2 National and European pilots as anchors

The analytical conclusions of the *Study on copyright and new technologies* subsequently informed further work at Member State and European levels[60], including the establishment of the **Copyright Infrastructure Task Force** (**CITF**). The first CITF project report translates the study's findings into a set of concrete infrastructure-oriented requirements, focusing on interoperable identifiers, standardised metadata schemas, and governance arrangements capable of supporting copyright-relevant data in the digital and AI era[61].

In this sense, the CITF process represents a continuation and operationalisation of earlier Commission-initiated analysis rather than a departure from it. The national and European pilots discussed below should be understood within this broader policy trajectory, as practical efforts to test, implement, and refine the types of solutions identified through these preceding analytical and consultative processes.

[58] 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).
[59] *Publishing Cultural Heritage Data in the Age of AI* (Keller 2025)
[60] *Study on copyright and new technologies: copyright data management and artificial intelligence* (SMART 2019/0038) (European Commission et al. 2022)
[61] *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)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Several of the implementation pilots referenced in this Green Paper were conducted within formal policy cooperation frameworks rather than as isolated technical experiments. In the Slovak Republic, this work was guided by a *Memorandum of Understanding concluded between the Ministry of Culture of the Slovak Republic and members of the OpenMusE consortium*[62]. The Memorandum established a framework for the reuse of open policy analysis outputs in cultural and creative sector policymaking and explicitly supported transparent, evidence-based experimentation with data governance and metadata infrastructures.

This institutional context is significant because it situates the Slovak pilot activities, its replication in Hungary and in the Baltic region within existing public policy processes and legal competences, involving national authorities, collective management organisations, universities, and memory institutions. As a result, the *Open Music Observatory* (pan-European central module, available at https://openmusicobservatory.eu/), its national federated *Slovak Comprehensive Music Database* (available at https://hudobnadatabaza.sk/), the national federated *Hungarian Music Database*, and the music module of the sub-national *Finno-Ugric Data Sharing Space* (available at https://finnougric.net/) should be seen not only as a technical pilot, but as an example of policy-embedded implementation aligned with both national cultural policy objectives and emerging European discussions on copyright infrastructure.

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), the proposed European Music Observatory on the basis of the *Open Music Observatory* (Chapter 3), and AI (Chapter 4).

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.

This MoU formalises a federated governance model where:

[62] *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](Ministerstvo kultúry SR and Open Music Europe 2023)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

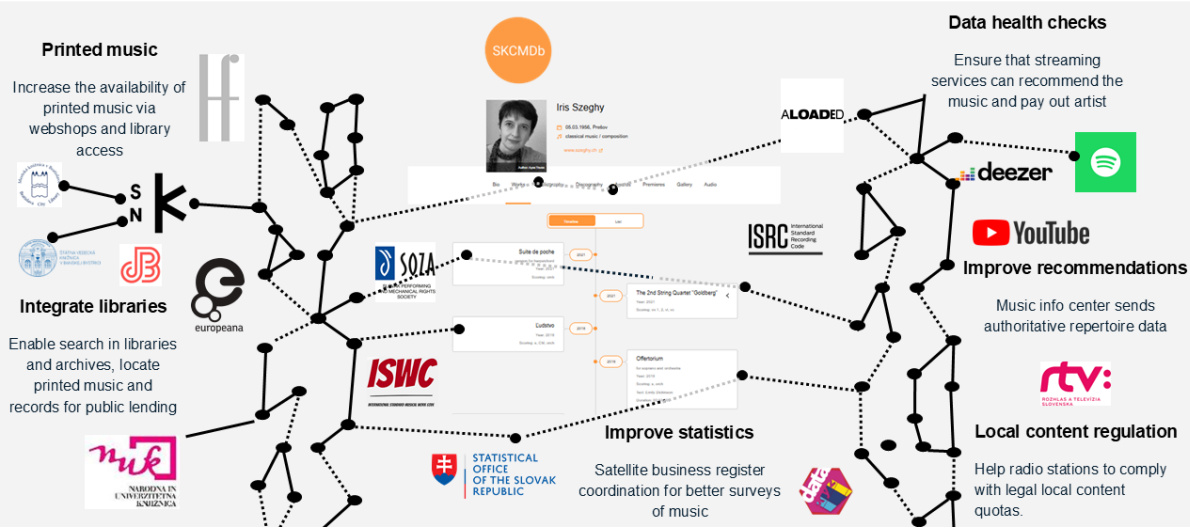
- **Attribution** (names of authors, performers, composers) is preserved as legally mandatory under copyright law.
- **Privacy** is safeguarded by layered access: public data (names, works, identifiers) circulate broadly, while sensitive data (e.g., addresses, birth dates) remain restricted.
- **Interoperability**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

The SKCMDb in action

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

Data sharing and exchange in the Slovak music data sharing space



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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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).[63]

ⓘ Building the Diversity Pillar of the European Music Observatory with Federated Knowledge Graphs

[63] 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

The diversity dimension of the Open Music Observatory is developed most fully *Music, Heritage, and Policy in the Age of AI — Building the Diversity Pillar of the European Music Observatory with Federated Knowledge Graphs* DOI: 10.5281/zenodo.17922390[64], which operationalises the policy principles and recommendations of this Green Paper into a concrete implementation blueprint, workflows, and indicators, and serves as the primary technical reference for Deliverable D2.3. Unlike many policy analyses that treat diversity primarily as a monitoring objective, the Open Music Observatory Diversity Pillar operationalises diversity as a data governance and infrastructure problem.

Prepared within the Open Music Europe Horizon Europe Research and Innovation Action, it demonstrates how diversity-related policy goals — including cultural diversity, linguistic plurality, regional representation, and the visibility of marginalised repertoires — can be embedded directly into metadata models, identifiers, workflows, and observability rules. Rather than relying on downstream correction or reporting alone, the approach implements diversity by design at the point where data is created, linked, and reused.

The deliverable draws on implemented pilots, including the *Slovak Comprehensive Music Database*, the *Finno-Ugric Data Sharing Space*, and cross-border replication experiments, to show how federated infrastructures can support diversity-sensitive indicators without centralisation. It addresses issues such as unequal documentation capacity, under-representation of minority repertoires, language bias in identifiers and authority files, and the amplification of existing inequalities by algorithmic systems.

In policy terms, the document is relevant to current European debates on AI, cultural diversity, as an extension of this *Green Paper*. It shows how trustworthy, provenance-aware metadata infrastructures can make diversity measurable, actionable, and auditable — rather than symbolic — and how these mechanisms can feed both cultural policy and AI governance frameworks.

Music, Heritage, and Policy in the Age of AI — Building the Diversity Pillar of the European Music Observatory with Federated Knowledge Graphs therefore complements the present *Green Paper* by providing a fully implemented reference for the diversity pillar, demonstrating how the Open Music Observatory can support European cultural diversity objectives in practice, not only in principle.

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.

[64] *Music, Heritage, and Policy in the Age of AI* (Antal 2025a)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.

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.

D.2.3 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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[65].

By contrast, in most industries administrative overhead is modest:

- Retail/distribution: ~2–5% of net sales
- 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.[66]

Metadata as provenance

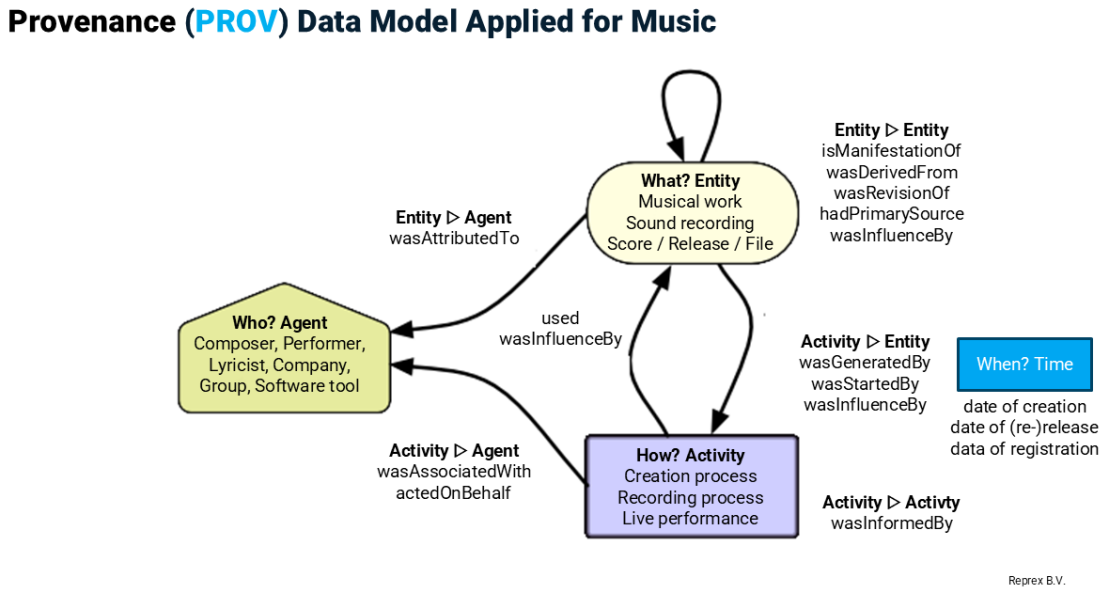
In today's music ecosystem, almost every asset is born digital. A modern composer's score is produced in notation software; a performer's recording originates as a digital file; even printing, distribution, and promotion leave their own digital traces. From the very start, each musical work and each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

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

[66] 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).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295



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).

The earlier introduced *Open Future* and *Europeana* policy brief on the publication of cultural heritage data in the age of artificial intelligence highlights why provenance failures can no longer be treated as a tolerable archival imperfection. When incomplete, ambiguous, or degraded provenance metadata is incorporated into automated and AI-driven systems, those systems tend to treat it as authoritative, propagating errors, omissions, and misattributions at scale rather than correcting them. In this context, broken provenance is not merely a documentation problem but a structural risk: once embedded in algorithmic pipelines, it shapes discovery, attribution, reuse, and economic outcomes in ways that are difficult to reverse. This reinforces the need for lifecycle-aware metadata governance that treats provenance as a core infrastructural concern in the AI era, rather than as a secondary administrative afterthought.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

D.2.4 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

The analysis and examples presented in this Green Paper are informed by a set of closely related analytical, legal, and implementation documents, including:

- *Study on copyright and new technologies: copyright data management and artificial intelligence*
- *Interoperable, trustworthy, and machine-readable copyright data in the AI era* (CITF First Project Report)
- Memorandum of Understanding on the reuse of Open Policy Analysis outputs in Slovak cultural policy
- OpenMusE technical deliverables documenting implemented data infrastructures and workflows

These materials provide additional detail on the legal, technical, and institutional assumptions underlying the policy synthesis offered here.

CITF

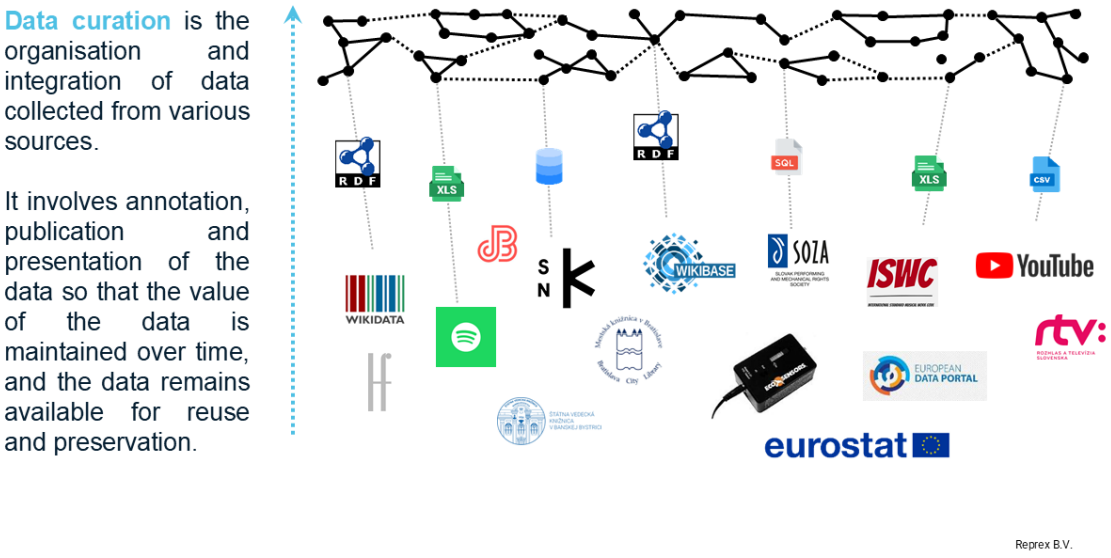


You can download this report at <https://julkaisut.valtioneuvosto.fi/server/api/core/bitstreams/3fb3f9ed-e93a-48fb-92e8-93c61a999e48/content>

The CITF report, based on the *Study on copyright and new technologies: copyright data management and artificial intelligence*, 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

D.3 Fixing Music Data at the Source



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)

D.3.1 Discussion

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).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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 2025b).

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[67].

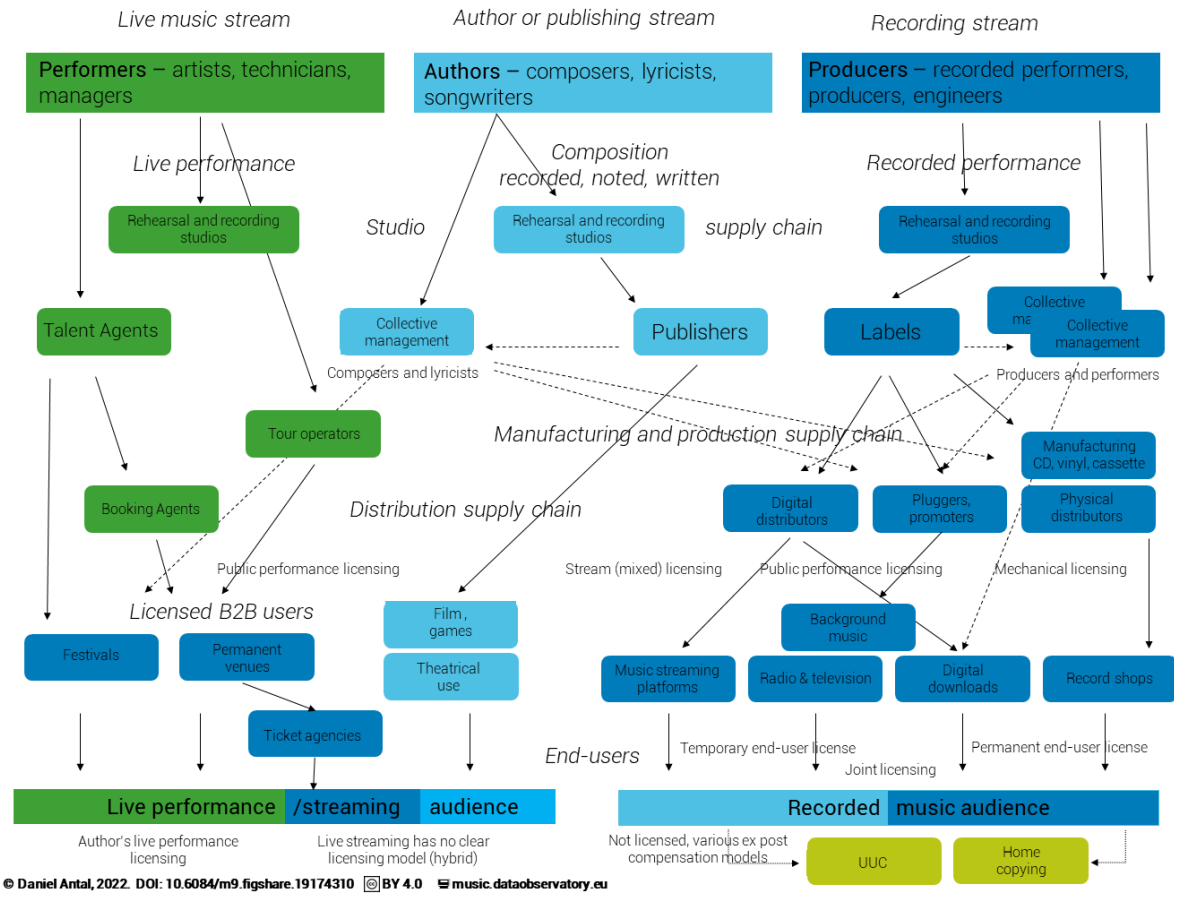
Recognising this scattered landscape is essential. It explains why reconciliation overheads are high, why identifier coverage is incomplete, and why “capture once, reuse many” pipelines are necessary. It also provides the foundation for the next chapter: explaining why *attempts at centralisation are 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 — that still structure industry practice.[68] 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.

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

[68] 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295



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[69]. This approach is especially useful for designing data collection that measures 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.

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

[69] 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).

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.

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.[70] 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.

ⓘ Why collections differ in databases

[70] 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].

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

- **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.

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 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.[71]

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.

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

[71] 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).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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)[72]. This scale of legacy under-documentation cannot 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

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

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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[73].

ⓘ Case Study: Metadata Repair — Heritage and Repertoire

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

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.
- 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

- 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.

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 proportionality 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[74].

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.

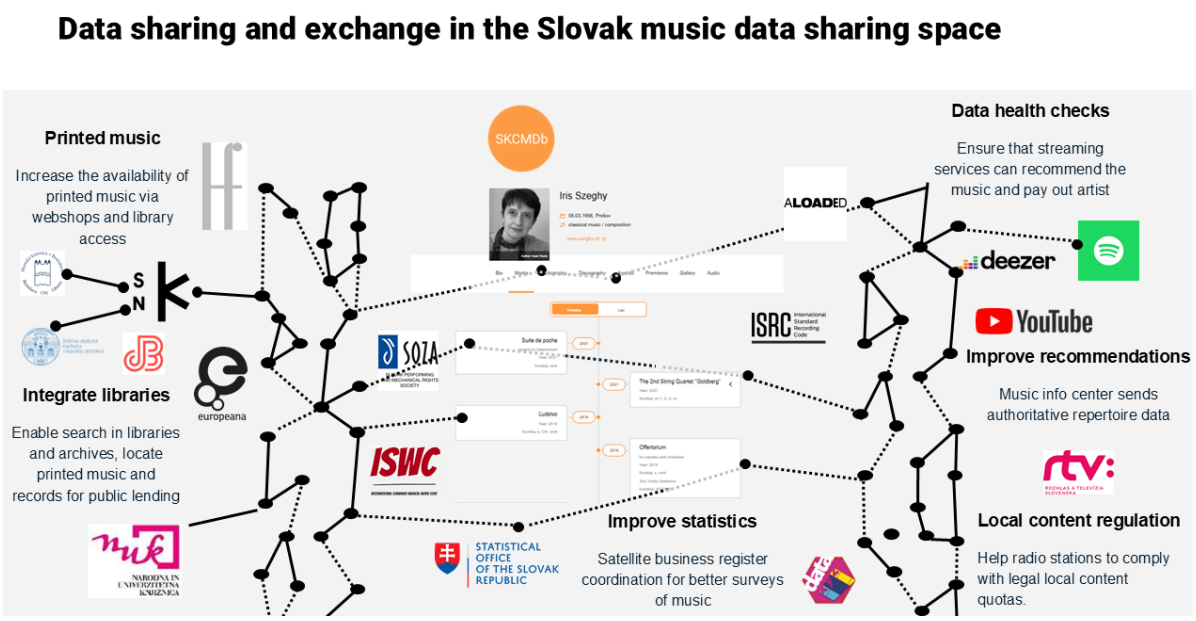
Policy communities in Europe recognise these issues. The **Big Data Value Association (BDVA)** has long argued that trust frameworks and governance pillars are essential for data sharing, while the **Federation Working Group** stresses that federation — not centralisation — is the only realistic model 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.

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

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

D.3.2 Policy proposals



Explanation

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[75].

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 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

[75] 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].

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

to a shared semantic framework.[76] 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.[77]

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.

> ⓘ Interoperability is a means, not a goal

[76] 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).

[77] 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.
- **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.

Proportionality 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 Directive* (2019/1024) provide the horizontal framework for treating music metadata as part of Europe's emerging common data spaces.[78]

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.[79]

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.[80]

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.[81]

---

[78] 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.

[79] 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.

[80] 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.

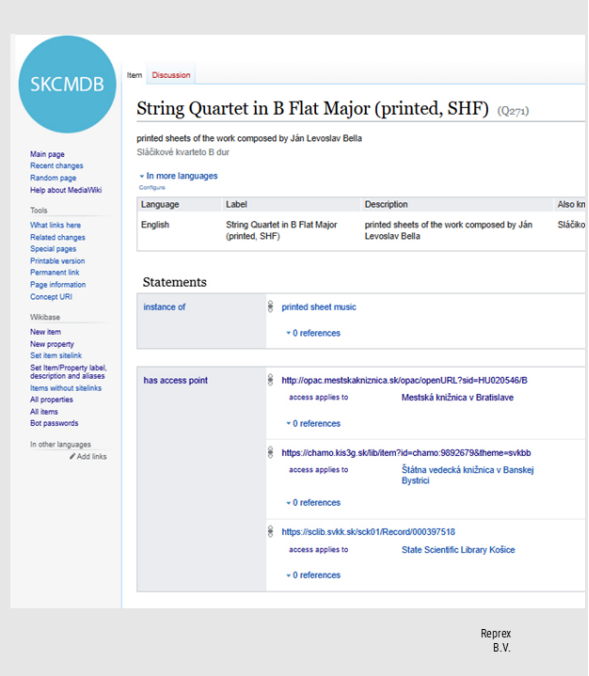
[81] For ORCID–VIAF reconciliation via OpenRefine, see (OpenRefine Community 2021; Jegan et al. 2023). For systematic mappings between DataCite and Dublin Core, see (DataCite 2021).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

DCTERMS patterns

Allow sufficient mapping to/from DCTERMS and interoperability with other industry and domain standards like DDEX, Wikidata, Polifonia, RiC and CIDOC

- No new ontology development
- Use eXtreme Design for designing system competencies
- Ontological patterns to pragmatically work with music distributors and rights managers who work with DDEX
- For cultural sector, with CIDOC museums, old ISAD(G) and RiC archives
- DCTERMS / DataCite for EOSC – Dariah data
- DCTERMS / Europeana



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.[82]

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[83].

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

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

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

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.[84]

[84] 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).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

D.4 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** https://openmusicobservatory.eu/ 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.

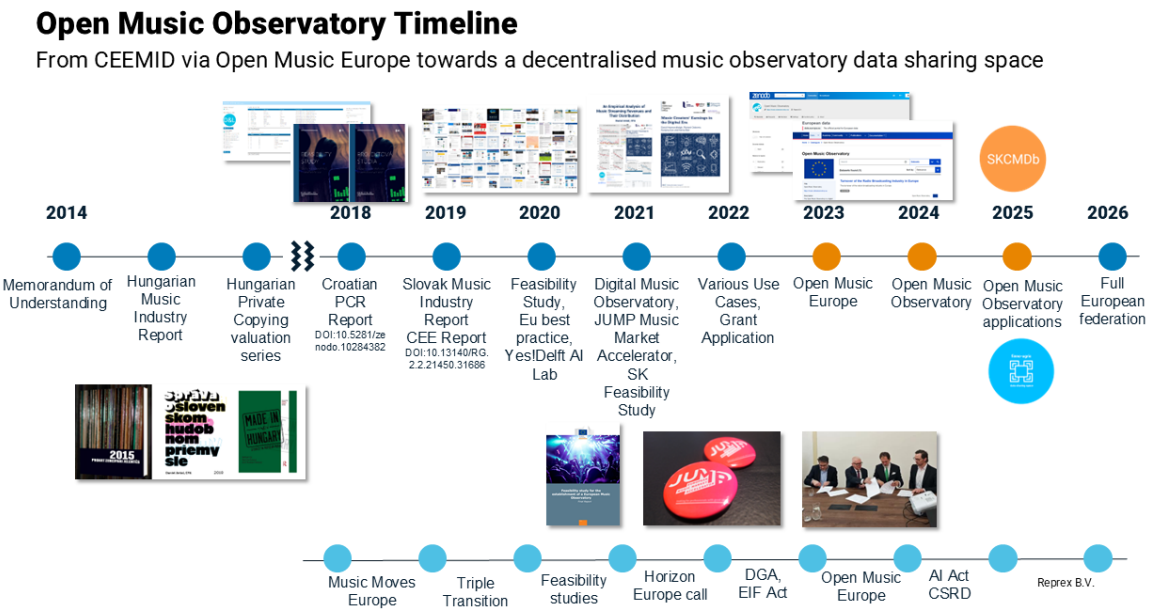
Transparency note: Following the principles of **Open Policy Analysis**, we have made all key deliverables (including versions 0.99, 1.01, and 1.1 of the *Open Music Observatory* document) publicly accessible to foster broad stakeholder engagement and to provide a clear audit trail. These versions are 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.[85]

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.

[85] 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295



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)

D.4.1 Discussion

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.

> ⓘ 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[86].

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[87].

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

[86] See for example Goldenfein and Hunter (n.d.); Milosic (2015).
[87] See (Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula 2025, pp15–18).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.[88]

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.[89]

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*

[88] 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.

[89] 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 into coherent cultural indicators.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.[90]

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 proportionality 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 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)[91].

Public infrastructures bypass music's real data flows

The European Union has made significant investments in cross-domain cultural and research data infrastructures, including Europeana, the European Collaborative Cloud for Cultural Heritage (ECCCH), the European Open Science Cloud (EOSC), and, more recently, the planned EU Culture Data

[90] 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).

[91] The Slovak pilot demonstrated that even with careful communication and GDPR proportionality 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).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Hub announced in the Cultural Compass for Europe. These initiatives establish essential foundations for digitisation, discovery, preservation, and evidence-based cultural policy.

However, by design, such infrastructures operate at a generic, cross-domain level. They are not equipped to encode the sector-specific lifecycle logic, rights complexity, and attribution requirements that characterise recorded music. Music is not marginal in these systems because of lack of relevance, but because its data structures—multiple overlapping rights per asset, high-frequency reuse, and dense interconnection between public and private registries—exceed what cross-domain cultural infrastructures can reasonably handle without domain-specific implementation layers.

The EU Culture Data Hub reflects this same architectural logic. It is intended to aggregate and harmonise cultural evidence across domains, not to replace or replicate sectoral infrastructures. As such, it depends on the existence of interoperable, rights-aware domain-level data spaces capable of supplying structured, comparable inputs. Music illustrates particularly clearly why such domain-level implementation is necessary.

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.[92] The *ECCCH report* likewise overlooked music entirely, focusing instead on monuments, archaeology, textiles, and museums.[93] 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 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

[92] 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).
[93] *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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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[94].

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.[95]
This challenge directly motivates the Observatory's bridging role with EOSC, Europeana, and ECCCH, elaborated in Section 3.2[96].

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[97].

---

[94] 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].
[95] 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).
[96] 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).
[97] 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

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.[98]

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.

> ⓘ 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 [99].

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.[100]

> ⓘ The U.S. Mechanical Licensing Collective (MLC)

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).

[98] 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).

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

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

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).

Economies of scale in metadata

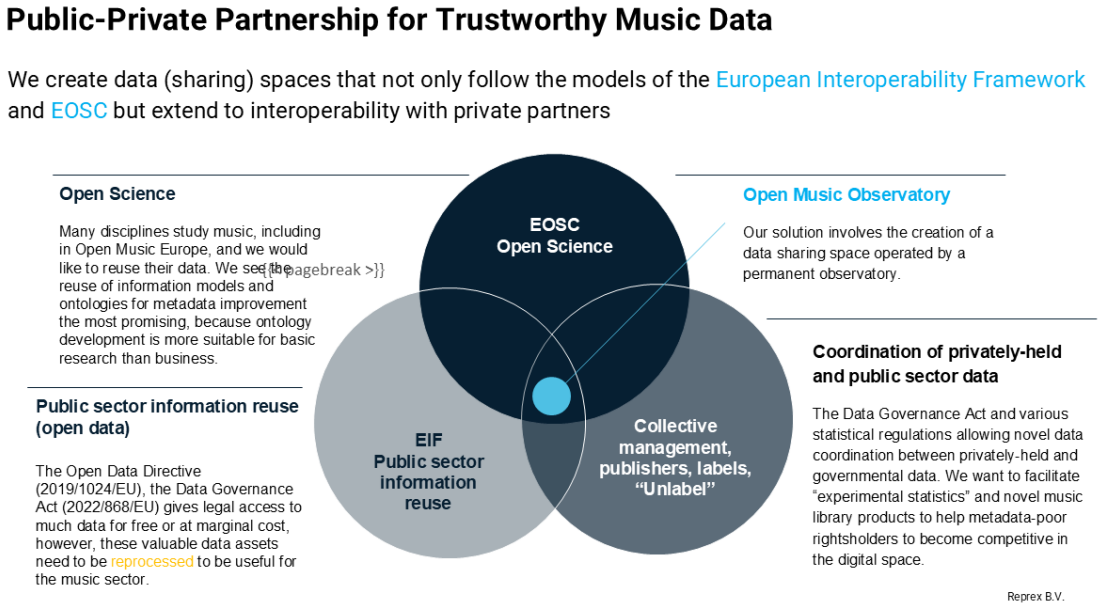
Large platforms and major labels can document millions of tracks at very low per-unit cost, because they manage everything in bulk. Smaller actors — independent labels, non-profits, or community archives — face the opposite situation: the cost of registering and maintaining each track is often higher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

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.[101]

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).

[101] 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).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295



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](https://figshare.com/articles/dataset/The_Open_Music_Observatory_at_the_Intersection_of_Open_Science_Open_Data_and_Music_Industry_Workflows/30073267/1?file=57754399)

D.4.2 Policy Proposals

The policy proposals set out in this section are directly grounded in, and fully consistent with, the analytical and institutional trajectory established at European level over the past years. The *Study on copyright and new technologies: copyright data management and artificial intelligence* identified persistent structural failures in rights metadata management, interoperability, authority, and provenance across creative sectors, and highlighted that emerging AI uses amplify these weaknesses rather than resolve them. The first project report of the Copyright Infrastructure Task Force (CITF) subsequently translated these analytical findings into concrete, infrastructure-oriented requirements, including interoperable identifiers, standardised metadata schemas, trustworthy provenance, and governance arrangements capable of supporting copyright-relevant data throughout the lifecycle of works and recordings.

The *Open Music Observatory* proposals presented here should be read as an operational continuation of this policy line. They do not introduce a parallel conceptual framework, but instead demonstrate how the requirements articulated in the Commission-initiated study and the CITF process can be implemented in practice through federated, non-centralised infrastructures. The national and regional pilots referenced throughout this section — including the *Slovak Comprehensive Music Database*, its replication in Hungary, the Baltic and Latvian pilots, and the *Finno-Ugric Data Sharing Space* — function as policy-embedded test cases that validate, refine, and

contextualise these requirements within existing legal competences, institutional mandates, and public–private cooperation frameworks.

In this sense, the Open Music Observatory is positioned neither as a standalone technical solution nor as a new central authority, but as a convening, conformance, and observability layer that operationalises the findings of the *Study on copyright and new technologies* and the first CITF report. It provides a practical mechanism for translating earlier Commission analysis and task-force recommendations into reusable governance patterns, interoperable workflows, and auditable data practices suitable for the digital and AI era.

> ⓘ 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, demonstrating how public heritage metadata can be reconciled with private music supply chains.
>
> - Archival recordings (including Hilda Griva's songs and Latvian/Latgalian midsummer songs) were identified in the Latvian Archives of Folklore.
> - Metadata was translated, enriched, and aligned with international authority files.
> - ALOADED extended this material with DDEX-compliant catalogue transfers and ingested it into Spotify and other platforms.
>
> This pilot demonstrates that reconciliation between public infrastructures (archives) and private infrastructures (distributors and platforms) is both technically and institutionally feasible. It reconnects suppressed or marginalised repertoires with contemporary audiences without requiring centralisation or loss of institutional control.
>
> A more technical description of the workflow is available here: https://downloads.reprex.nl/2025/open-music-observatory/coordination.html#sec-future-proofing

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 alignment ensures compatibility with wider European data-space initiatives and reduces integration costs for institutions already adapting to these standards (Commission et al. 2020, p9).

Workflow playbooks and provenance trails

The Observatory should not only harmonise data formats, but also document workflow playbooks that describe how metadata moves across the music lifecycle:

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

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Each step should define change-propagation rules: when a correction is made in one register, it should propagate to dependent systems. Provenance trails must survive system boundaries, using standards such as PROV-O to document who did what, when, and under what authority. This enables auditability, cross-border comparability, and long-term trust, and prevents “data death” when assets move between systems.

Federated infrastructure as a cost and governance solution

The Cultural Compass for Europe establishes a clear strategic direction for European cultural policy: shared, interoperable, and federated data infrastructures that support evidence-based policymaking, cultural diversity, and trustworthy use of artificial intelligence. The planned EU Cultural Data Hub reflects this ambition at cross-domain level. At the same time, the Compass does not prescribe how sector-specific complexities should be handled in practice.

This section argues that music represents a domain where such complexity exceeds what generic cultural data infrastructures can reasonably encode, and therefore requires a specialised, federated implementation layer that remains fully compatible with the broader European framework. The *Open Music Observatory* should be understood precisely in this role: not as an alternative to the *EU Culture Data Hub*, but as a music- and music-rights–specific implementation layer that operationalises the Hub’s objectives for one of Europe’s most complex cultural sectors

The imbalance described in Section 3.1.6 makes one point clear: small and medium-sized actors cannot compete on metadata quality without shared infrastructure. In a sector characterised by extreme fragmentation, federation — rather than centralisation — is the only viable path forward.

A data sharing space provides the appropriate governance framework for such federation. Instead of forcing participants into a single schema, database, or legal agreement, it enables organisations to share and reuse data on an as-needed or as-permitted basis while retaining stewardship over their own assets. For music — where rights, identifiers, and content are distributed across hundreds of micro-actors and institutions — this approach reduces duplication without creating dependency, while remaining compatible with cross-domain cultural data infrastructures such as the **EU Cultural Data Hub**.

Crucially, federation avoids the risks inherent in centralisation: technical fragility, governance lock-in, and the emergence of a single gatekeeper capable of restricting access or imposing unilateral conditions on others[102]. These risks are particularly acute in copyright-relevant environments, where errors in attribution or provenance have direct legal and economic consequences.

Music represents one of the most demanding test cases for European data governance. Attribution interacts directly with privacy law, identifiers are applied unevenly across the sector, and most music enterprises lack the resources to maintain their own compliance, documentation, and audit infrastructures. If a federated model can function in this environment, it can function elsewhere. At

[102] This definition paraphrases (Curry 2020) and reflects the view that a data sharing space is an ecosystem of exchange, processing, sharing and provision of data between trusted partners (EBU and Gaia-X 2022, p16). The CITF further emphasises that trustworthy provenance, machine-readable rights metadata, and auditable rights management information are essential components of such federated copyright infrastructures, particularly in the AI era [Partanen, Rixhon, Bandere, Ziediņš, Dutt, Bolšteins, Frosterus, Lehtinen, Miklūna-Žukeviča, Ozerskis, Pihlaja, Sauka, Sornova, and Uzula (2025), p31; pp101–102].

the same time, decentralisation also creates incentives for data hoarding. Effective governance must therefore make participation more attractive than isolation — through lower administrative costs, legal clarity, shared compliance benefits, and increased visibility.

In practice, this requires combining hard alignment (minimal metadata profiles and baseline identifiers) with soft alignment (mappings, crosswalks, and shared workflow playbooks). This combination allows interoperability to be achieved without forcing all actors into a single organisational or technical model, and ensures that music-specific solutions remain interoperable with the *EU Cultural Data Hub* and other cross-domain infrastructures.

For these reasons, a future European Music Observatory, as part of the emerging EU Culture Data Hub ecosystem, cannot be designed as a single central database. Instead, it must function as a federated, domain-level observability and coordination layer. In this role, the Observatory complements the *EU Culture Data Hub* by providing validated, rights-aware, provenance-rich music data that can be aggregated at European level without centralising ownership or governance.

The Open Music Observatory is conceived as a music- and music-rights–specialist federated unit that connects shared cultural data infrastructures—including the *European Collaborative Cloud for Cultural Heritage*, *Europeana* and its emerging data sharing space, and the *EU Culture Data Hub*—with sector-specific workflows in music creation, distribution, rights management, and preservation. Rather than duplicating cross-domain functions, it supplies the missing domain logic required for music to be meaningfully represented within European cultural data infrastructures.

Architecturally, the Open Music Observatory inherits CEEMID's core principles: federation rather than centralisation, reproducible indicators rather than static datasets, and collaboration across public, private, and academic actors. What changes is not the logic, but the scale and governance context. OMO aligns these principles with the European Interoperability Framework, FAIR data principles, the common European data space for cultural heritage, and the work of the *Copyright Infrastructure Task Force*, ensuring coherence across cultural, data, and copyright policy domains.

Earlier experimentation, including the CEEMID initiative and the Open Music Observatory's Slovak cultural statistics pilot within the Open Music Europe project, already anticipated this architecture by demonstrating how music-specific data coordination could feed policy-relevant indicators without centralising data—an approach that aligns closely with the design principles now articulated for the EU Culture Data Hub.

> ⓘ Diversity & Circulation Pillar Implementation
>
> The Diversity & Circulation Pillar of the Open Music Observatory is implemented and documented in *Music, Heritage, and Policy in the Age of AI*[103]. While this Green Paper defines the policy rationale and governance principles for diversity-aware data infrastructures, the diversity pillar's implementation paper provides the concrete data models, workflows, and pilot results used to operationalise these principles.

A central objective of the Slovak pilot was to demonstrate how music-specific statistical indicators could be produced in line with best statistical practice, where existing cultural and creative satellite

[103] *Music, Heritage, and Policy in the Age of AI* (Antal 2025a), inthe OpenMusE project it is Deliverable D2.3.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

accounts fall short, which are themselves rarely implemented in EU member states[104].Drawing on established methods for business satellite accounts, the pilot sought to harmonise authoritative collective rights management data with administrative registers, industry datasets, and statistical surveys in order to generate indicators that are not available through Slovakia's current cultural and creative satellite accounting system. In practice, this involved linking rights management information—often the most complete and economically meaningful source for music activity—with official statistical frameworks, while preserving methodological transparency, reproducibility, and comparability.

This approach responds to a broader structural gap at European level: most EU Member States do not operate dedicated cultural or creative satellite accounts at all, and where such systems exist, they rarely capture the distinctive revenue structures, cross-border flows, and attribution dynamics of music. The pilot therefore addressed not only a national shortcoming, but a systemic limitation of European cultural statistics, anticipating the need for domain-specific data pipelines capable of producing robust, music-aware indicators. While the OpenMusE project was ultimately unable to carry out this experiment due to resource constraints and changing institutional conditions, the design remains methodologically sound and operationally feasible, and could be taken forward as part of the preparatory piloting of the EU Cultural Data Hub.

Recent analysis of cultural heritage data governance in the context of artificial intelligence clarifies that the limitations of existing heritage infrastructures should not be interpreted as a reason to exclude libraries and memory institutions from future data architectures. On the contrary, these institutions occupy a distinctive position in the AI era precisely because they combine long-term custodianship with public accountability, authority control, and responsibility for provenance and attribution[105]. From the perspective of the Cultural Compass, integrating these institutions more directly into copyright-relevant data flows is essential for ensuring that attribution, provenance, and accountability are preserved across the full lifecycle of music data.

Europe already has the necessary policy and technical foundations for this approach. The *European Strategy for Data*, the *Data Governance Act*, and the *Data Act* define data spaces as federated by design[106]. The *Data Spaces Support Centre* (DSSC) has translated these principles into operational blueprints that can be applied directly to the music sector[107]. Comparable logics already exist in music and adjacent domains: ISRC governance through national agencies, CISAC's CIS-Net, and European statistical systems based on subsidiarity rather than central repositories.

Concerns about digital sovereignty make this approach urgent. In the absence of a European solution, metadata infrastructures risk drifting toward US-style centralisation, such as the Mechanical Licensing Collective (MLC), where legislative mandates and market power converge in a single hub.

---

[104] Pilot Program for Novel Music Industry Statistical Indicators in the Slovak Republic (Antal 2023). The pilot was not implemented within the OpenMusE project due to resource constraints and institutional changes, but its methodological design remains valid and suitable for continuation in the context of EU Cultural Data Hub preparatory work.

[105] *Publishing Cultural Heritage Data in the Age of AI* (Keller 2025)

[106] The European Strategy for Data (2020), the Data Governance Act (2022), and the Data Act (2023) establish federated data spaces supported by trust frameworks and shared services (European Commission 2020; European Parliament and Council 2022).

[107] The Data Spaces Support Centre (DSSC) provides blueprints and building blocks for implementing federated data spaces across sectors (Data Spaces Support Centre 2025b, 2025a).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

OpenMusE Deliverable D5.6 explicitly warns against this outcome[108]. This Green Paper complements that legal-institutional analysis by demonstrating how a federated, culture-led infrastructure can translate EU law into practice.

In practical terms, this requires capture-once, reuse-many pipelines across the music lifecycle: from registration of works and recordings, through distribution and royalty attribution, to preservation and cultural statistics. The Observatory should therefore function as a redundancy-minimising registration and coordination space, aligned with the European Interoperability Framework and provenance-oriented models such as PROV-O[109].

When implemented effectively, this approach lowers entry barriers for smaller actors, improves interoperability across institutions, and grounds Europe's cultural and economic policies in reliable evidence rather than fragmented silos.

Broader debates on AI governance caution that technological systems can reproduce existing hierarchies when governance structures remain unequal. Guest, Suarez, and van Rooij warn that AI may extend "projects of domination, of hierarchies, of extractivism of cognitive labour"[110]. Federated, transparent, and participatory infrastructures therefore play a corrective role by preventing the concentration of informational and cultural power.

This consideration is particularly salient in the *Finno-Ugric Data Sharing Space* music module (https://finnougric.net/en/), which addresses decolonisation and subsidiarity by consolidating dispersed knowledge about the musical heritage of Livonians, Sámis, Maris, and Setos. The posthumous publication of Hilda Griva's recordings illustrates how federated infrastructures can enable cultural self-representation rather than algorithmic erasure.

3.2.3 Legal, standards, and funding levers

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

- Legal: explicit GDPR legal bases per data flow (for example, legitimate interest for attribution; research and cultural-heritage exemptions for archives).
- Standards: codes of conduct and minimal conformance profiles that remain achievable for micro-enterprises.
- Funding: targeted support through ECCCH pilots, national ministries, and EU programmes for metadata fitness and data-quality improvements as public-interest infrastructure.

3.2.4 Alignment with the European Open Science Cloud

Position the Open Music Observatory as the convening, conformance, and observability layer connecting ECCCH, Europeana, and GLAM authority files with industry workflows. Concretely:

[108] See Policy Brief 1: Music Metadata Mainstreaming and EU Law (Senftleben et al. 2024).
[109] The European Interoperability Framework defines interoperability across legal, organisational, semantic, and technical layers (Commission and Digital Services 2017). The W3C PROV model and PROV-O ontology enable standardised provenance chains linking actors, activities, and entities (W3C 2013b, 2013a).
[110] Towards Critical Artificial Intelligence Literacies (Guest, Suarez, and Rooij 2025, p3)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

1. Capture once, reuse many across creation, registration, distribution, and preservation.
2. Require minimal profiles that smaller actors can realistically implement.
3. Prioritise identifier crosswalks (ISRC–ISWC–ISNI–VIAF/Wikidata) and change propagation.
4. Use Wikibase and Wikidata as a low-friction backbone where appropriate.
5. Govern through EIF- and FAIR-aligned rules with auditability and public–private participation.

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

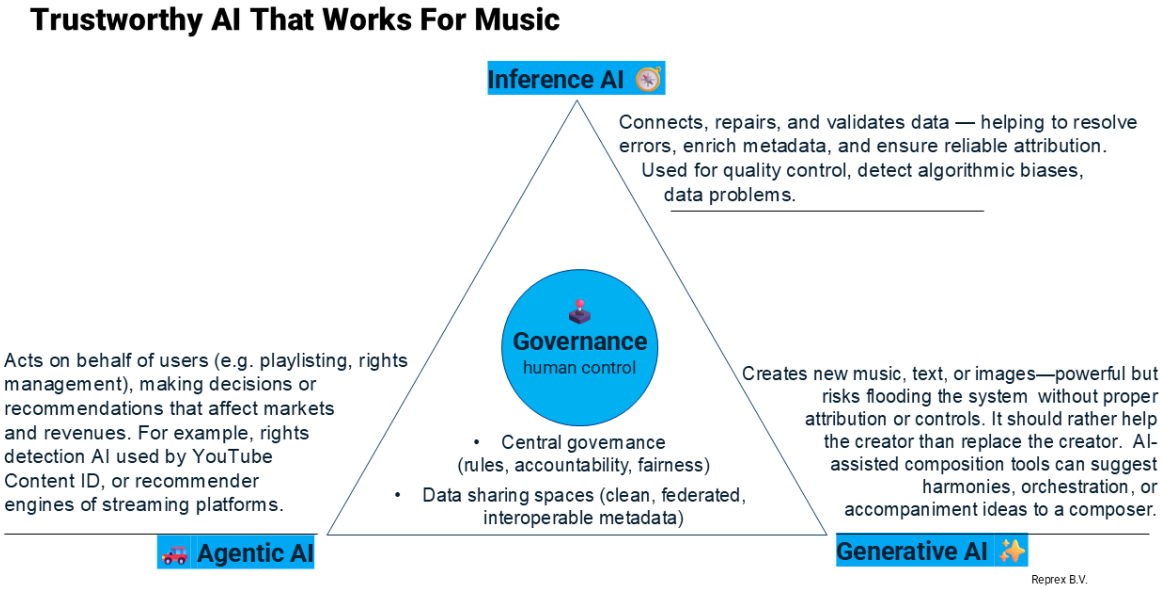
© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

D.5 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295



The Culture Compass for Europe recognises both the opportunities and risks of AI for cultural and creative sectors.

The *Culture Compass for Europe* recognises both the opportunities and risks of AI for cultural and creative sectors. Music illustrates these risks earlier and more sharply than most domains, because AI systems rely heavily on large-scale music datasets for training, recommendation, and generation. In this context, trustworthy provenance and rights-aware metadata are not optional safeguards, but prerequisites for lawful and ethical AI deployment.

The use, development and governance of **artificial intelligenc**e (AI) systems should foster human creativity through a fair human-centric and rights-based approach. It should: respect cultural rights, accessibility, inclusivity and cultural diversity; develop and promote discoverability of European, national and local content; foster competitiveness; counter digital divides; and foster digital inclusion. We commit to:
• Promoting human creation and European cultural and linguistic digital sovereignty, and addressing ethical risks of biases and cultural homogenisation.
• Protecting intellectual property rights, by addressing the impacts of AI on creators' remuneration, while embracing innovation.
• Monitoring and mitigating the impact of AI on jobs, as well as supporting the cultural and creative sectors and industries adapt to technological change and acquire digital skills.
• Fostering the use of AI as a tool to support cultural and creative professionals, and enabling the cultural and creative sectors and industries to harness the opportunities offered by these technologies.

Draft Joint Declaration "Europe for Culture — Culture for Europe" (European Parliament, Council of the European Union, and European Commission 2025)

European policy provides guidance and evolving regarding to this proportionality 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[111].

> ⓘ 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[112]. 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.

---

[111] 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 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].

[112] 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

D.5.1 Discussion

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[113].

[113] 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 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

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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[114].

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.

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[115].

---

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).

[114] 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.

[115] *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

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

ⓘ 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[116].

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

(Berardinis et al. 2025). For more information on the *Human Artistry Campaign*, see https://www.humanartistrycampaign.com/.

[116] 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 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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[117], 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 proportionality tests and opt-ins show the futility of this approach. 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.

[117] 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).

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.

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

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.

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.

D.5.2 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.

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.

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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

- **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.[118]

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."

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

[118] 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Curative AI as regeneration, not replacement



伊勢神宮 (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.

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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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[119].

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.

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

[119] 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)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

guardrails — improving attribution and interoperability while reducing the risk of generative AI misuse.

> ⓘ 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.

D.6 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 evidence presented in this *Green Paper* suggests that the primary challenge for European music data policy is no longer the absence of conceptual frameworks or technical standards, but the limited scaling and coordination of solutions that already exist and have been partially implemented. Future European action should therefore prioritise the consolidation, federation, and long-term governance

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

of proven approaches, rather than the repeated creation of new pilot structures detached from policy and institutional contexts.

In this respect, the Open Music Observatory and related national implementations should be understood as components of an evolving policy infrastructure, capable of informing regulatory discussion, institutional coordination, and future standardisation efforts through documented practice rather than abstract design alone.

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.

This *Green Paper* shows that the strategic ambitions of the *Culture Compass for Europe* — particularly regarding cultural data, AI, and evidence-based policy — cannot be realised through generic infrastructures alone[120]. They require federated, domain-specialised implementation units where cultural complexity demands it. Music, as one of the most demanding cultural domains, provides a robust reference case. The Open Music Observatory demonstrates how such a unit can operate in practice while remaining fully interoperable with Europe's shared cultural data spaces.

By tracing its origins to CEEMID, the Open Music Observatory demonstrates how long-standing, practitioner-led solutions can be scaled and embedded within European policy frameworks. This continuity underscores that Europe's cultural data challenges do not require abstract reinvention, but the systematic extension of proven, decentralised approaches to new domains such as copyright infrastructure and AI governance.

An important lesson from the national pilots discussed in this paper is that music data infrastructures should be designed not only for interoperability and rights management, but also to support the production of policy-relevant evidence. Experiments such as the Slovak pilot showed how

[120] "Pursuing the ambitions of this declaration and fully leveraging culture's transversal impacts means promoting culture in the EU's internal and external policies, investing in the cultural and creative sectors and industries, and monitoring progress. We commit to: …. Advancing initiatives aimed at improving the availability of sound and comparable data on culture to support the development of evidence-based cultural policies."(European Parliament, Council of the European Union, and European Commission 2025, p4)

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

authoritative rights management data can be harmonised with administrative and statistical sources to generate music-specific indicators aligned with best statistical practice — an area where most Member States currently lack dedicated cultural or creative satellite accounts. Embedding this capability into future European data spaces is essential if music is to be visible in evidence-based cultural and economic policymaking.

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.

The use, development and governance of artificial intelligence (AI) systems should foster human creativity through a fair human-centric and rights-based approach.

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.

D.7 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*. September 23, 2021. https://doi.org/10.5281/zenodo.5249010.

———. 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. *SKCMBDb: Building the Slovak Comprehensive Music Database in the Slovak Music Centre*. Open Music Observatory. https://doi.org/10.5281/zenodo.18036465.

———. 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. *Music, Heritage, and Policy in the Age of AI*. Version 1.0. Open Music Observatory. https://doi.org/10.5281/zenodo.18076940.

———. 2025b. *Open Music Observatory*. Version 1.43. Open Music Observatory. https://doi.org/10.5281/zenodo.17897724.

———. 2025c. *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.

———. 2025d. *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.

———. 2025e. *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.

Artisjus, HDS, SOZA, and Candole Partners. 2014. *Measuring and Reporting Regional Economic Value Added, National Income and Employment by the Music Industry in a Creative Industries Perspective. Memorandum of Understanding to Create a Regional Music Database to Support Professional National Reporting, Economic Valuation and a Regional Music Study.*

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.

BITSS. 2019. *Guidelines for Open Policy Analysis*. Berkeley Initiative for Transparency in the Social Sciences. http://www.bitss.org/wp-content/uploads/2019/03/OPA-Guidelines.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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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 Commission, Directorate-General for Communications Networks, Content and Technology, Crowell&Moring, IMC University of Applied Sciences Krems, Philippe Rixhon Associates, Technopolis Group, and UCLouvain. 2022. *Study on Copyright and New Technologies – Copyright Data Management and Artificial Intelligence*. Publications Office of the European Union. https://doi.org/doi/10.2759/570559.

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 Parliament, Council of the European Union, and European Commission. 2025. *Draft Joint Declaration “Europe for Culture – Culture for Europe”*. Brussels: European Commission. https://culture.ec.europa.eu.

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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.

Guest, Olivia, Marcela Suarez, and Iris van Rooij. 2025. *Towards Critical Artificial Intelligence Literacies*. Zenodo. https://doi.org/10.5281/zenodo.17786243.

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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

———. 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.

J, Wilson. 2015. *Evidence-Based Policy Making in the European Commission*. Edited by Elisabeth Lannoo. Oslo (Norway): CICERO Centre for International Climate; Environmental Research. http://www.cicero.uio.no/en/posts/news/report-from-science-to-policy-how-to-improve-the-dialogue/.

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.

Keller, Paul. 2025. *Publishing Cultural Heritage Data in the Age of AI*. Policy report. Open Future; Europeana. https://openfuture.eu/wp-content/uploads/2025/12/251202PublishingCulturalHeritageDataInTheAgeOfAI.pdf.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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.

Kitzes, Justin, Daniel Turek, and Fatma Deniz, eds. 2018. *The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences*. 1st ed. University of California Press. http://www.practicereproducibleresearch.org/.

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 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). 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://doi.org/10.5281/zenodo.17944096.

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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

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. 2023. *Open Music Europe (OpenMusE) – An Open, Scalable, Data-to-Policy Pipeline for European Music Ecosystems*. https://doi.org/10.3030/101095295.

Open Music Europe Consortium. 2023. *Open Music Europe User Stories: OPA-Compliant User Stories for System Competences and Semantic Architecture*. Https://github.com/dataobservatory-eu/open-music-europe-user-stories/. https://github.com/dataobservatory-eu/open-music-europe-user-stories/.

———. 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

———. 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.

© 2023-2025 The Authors | OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295