

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

**Practical Steps Towards a Decentralised and Open European Music
Observatory**

Daniel Antal, CFA

2024-09-18

Table of contents

Introduction	4
Glossary	8
Music terms	8
Data terms	9
AI & Systems Terms	10
Data protection terms	12
Data curation and collection terms	12
Statistical terms	13
Registers, authorities, standards and identifiers	14
Organisations	16
Other abbreviations	17
1 Policy Context & Problem Map	18
1.1 Quest for Efficiency	20
1.2 Potential solutions	23
1.2.1 Fixing Metadata At Source	23
1.2.2 European Policy Baseline for Data Spaces	25
1.2.3 The Open Music Observatory as a Data Sharing Space	25
1.2.4 AI that Serves the Music Ecosystem	26
1.3 Case studies	27
2 Fixing Music Data at the Source	30
2.1 Discussion	31
2.1.1 Data fragmentation as a structural fact	31
2.1.2 Cost Barriers in Documentation and Claims	31
2.1.3 Why one grand collection model will not work	32
2.1.4 Legacy Metadata	33
2.1.5 Named-Entity Resolution, Attribution, and Privacy	35
2.2 Policy proposals	35
2.2.1 Reducing Redundancy	36
2.2.2 Reconciling Attribution and Privacy	36
2.2.3 Pragmatic Metadata Alignment (Enhanced with Cross-Domain Ex- amples)	37
3 Open Music Observatory: Building a Shared Music Data Space	40
3.1 Executive summary	41

3.2	Discussion	42
3.2.1	Why centralisation is a futile model	42
3.2.2	Open Data Directive: Right Without Means	44
3.2.3	Why Voluntary Workarounds Do Not Scale	45
3.2.4	Public cultural infrastructures bypass music’s real data flows	45
3.2.5	Alignment with the European Open Science Cloud (EOSC)	46
3.2.6	European Interoperability Framework	47
3.2.7	Subsidiarity in Platform Design	48
3.2.8	Economies of Scale in Metadata	49
3.2.9	The U.S. Mechanical Licensing Collective (MLC) as a Metadata Clearinghouse	49
3.2.10	Data Sharing Space	50
3.2.11	Wikibase as a Cultural Metadata Backbone	51
3.3	Policy Proposals	52
3.3.1	Federated Infrastructure as a Cost and Governance Solution	53
3.3.2	Make Open Data Truly Accessible	54
3.3.3	Connect to Europeana & ECCCH	55
3.3.4	Alignment with the European Open Science Cloud	55
4	AI that Works for Music, Not Against It	57
4.1	Discussion	59
4.1.1	Structural Problems for Music Businesses to Apply AI	59
4.1.2	European Regulation That Misses the Point	61
4.1.3	Policy Issues at the Intersection of AI, Copyright, and GDPR . . .	61
4.1.4	AI Design Without Awareness of Limits	62
4.1.5	Unfreezing Frozen Assets	63
4.1.6	AI Support for Investment Into New Repertoire Assets	63
4.2	Policy Proposals: Aligning AI with Governance and Value Creation	64
4.2.1	EU-Level Policy: Compass and Guardrails	64
4.2.2	Industry-Level Policy: Standards and Collaboration	65
4.2.3	Organisational-Level Policy: Playbooks for CMOs, Publishers, and Archives	66
4.2.4	Embedding AI in Creative Workflows Instead of Replacing Creatives	66
4.2.5	Closing Note	67
4.2.6	Curative AI as a Remediation Solution	67
4.2.7	Lowering Documentation Barriers	68
4.2.8	AI-Powered Claims and API Integration	68
5	What Europe Should Do Next for Music Data & AI	69
	Sources & Further Reading	71

Introduction



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

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

i Why this Green Paper matters for music professionals?

Streaming has centralised power in platforms, but left rights-holders with micro-royalties and huge admin burdens.

Metadata mistakes mean lost revenue — each unlinked ISRC or ISWC is money left on the table.

AI is already changing music — either it helps you fix documentation and get paid, or it floods the system with untracked works.

Europe needs federated, cooperative solutions so independents, CMOs, and archives can compete on fairer terms.

Unlike some other cultural or commercial data, there is rarely a single moment when music metadata can be considered complete. Metadata, like music itself, is open to reinterpretation. A name can be reconciled with an identifier; a work can be linked to a new

performance; a recording can be embedded in new file formats. Each act of documentation adds layers of meaning and makes the music informative in a new environment.

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

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

i Note

This Green Paper is an early-stage policy document, prepared in line with **Open Policy Analysis** and the **Horizon Europe Data Management Guidelines**. It has been released early to allow consultation, incorporate stakeholder input, and provide a transparent development process. It will serve as the basis for:

- *A Green Paper on AI, Data Governance, and Metadata Policies for Europe's Music Ecosystem — Practical Steps Towards a Decentralised and Open European Music Observatory* [current document, version 0.3]
- Deliverable D5.7 (Policy Brief) for the *Open Music Europe* consortium
- A White Paper to be discussed at LineCheck 2025 and at the final policy forum of *Open Music Europe* and its system projects in Brussels, December 2025

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

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.¹

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.² This is an early version (v0.3.)

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

- Big Data Value Association, Gaia-X: Dataweek² : Introducing a new European music dataspace³
- Echoes/ECCH:
- Hungarian stakeholders interested in replication of the Slovak pilot versions ⁴
- CISAC: Protecting Creators' Rights in the AI Era: OpenMusE at the European Committee Meeting, Vilnius, 29-30 April ⁵.
- IAMIC ⁶: The *International Association of Music Information Centres* and several key members of the organisation.
- IAML ⁷: The *International Association of Music Libraries, Archives and Documentation Centers* and several national chapters and key members.
- Polifonia
- Slovak national stakeholders interested in cultural data⁸
- Wikimedia community and developers⁹.

²This 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.

²Always 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.

³Jun 5, 2024, Dataweek², Leuven, Belgium.

⁴*Federation possibilities of the Slovak music data sharing space in Hungary* (Antal 2024a)

⁵[Protecting Creators' Rights in the AI Era: OpenMusE at the European Committee Meeting](#), our presentation (Mikš 2025)

⁶We 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).

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

⁸Based 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).

⁹Our work was presented in the Technology session of the Wikimedia CEE Meeting 2024 in Istanbul, and we have built relationships with various national chapters to help our technology development, data curation and dissemination efforts. (Antal 2024b)

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 2017c)

manifestation: physical embodiment of an expression (ISO 2017c)

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 general 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 or sound recording copyrights. 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 2017c)

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

Data terms

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

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

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

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

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

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

data extensive datasets. (ISO 2019b)

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

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

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

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).

AI & Systems 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

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

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 2017b)

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

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

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

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

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

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 2017c)

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

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 2017c)

thesaurus: controlled vocabulary and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms.

Note 1 to entry: The purpose of a thesaurus is to guide both the indexer and the searcher to select the same preferred term or combination of preferred terms to represent a given subject. For this reason a thesaurus is optimized for human navigability and terminological coverage of a domain. [SOURCE:ISO 25964-1:2011, definition 2.62] (ISO 2017c)

expert system: knowledge-based system that provides for solving problems in a particular domain or application area by drawing inferences from a knowledge base developed from human expertise

Note 1: The term “expert system” is sometimes used synonymously with “knowledge-based system”, but should be taken to emphasize expert knowledge.

Note 2. Some expert systems are able to improve their knowledge base and develop new inference rules based on their experience with previous problems. (ISO 2023b) Expert systems fall under the definition of the AI Act.

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

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

Data protection terms

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

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

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

Data curation and collection terms

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

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

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

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

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

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

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

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

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

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

registration: act of giving an entity a unique identifier on its entry into a system (ISO 2017c) 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 2017c); an entity wishing to assign an ISRC to an applicable recording (ISO 2019a); a **party** that requests an ISNI from the Registration Authority (ISNI 3.2 (ISO 2012, p15))

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

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

Statistical terms

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

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

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

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

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

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

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

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

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

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

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

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

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

Registers, authorities, standards and identifiers

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

OpenCorporates: a public corporation database which sources data from national business registries.

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

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.

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.

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

ISBN: the International Standard Book Number is an identification system for the publishing industry and its supply chains.

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.

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

Organisations

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

ALOADED: is a company which distributes and exploits recordings.

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

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

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

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

GESAC: GESAC comprises together 32 European authors' societies in music, audiovisual, visual arts, literature and drama.

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

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

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

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

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

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

Other abbreviations

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

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

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

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

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

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

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

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

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

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

1 Policy Context & 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 competitors. At the same time, the average transaction value collapsed from around €18 (in current prices) to less than €0.005. This shock hollowed out much of the traditional infrastructure — record stores, radios, and music television — and shifted value capture toward data-driven platforms able to control access through recommender algorithms.

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

EU-level studies and policy frameworks have recognised these dynamics and increasingly frame them as systemic challenges. The *Feasibility Study for the Establishment of a European Music Observatory* diagnosed the fragmented, scarce, and poorly harmonised nature of music data collection across Member States, calling it the fundamental reason for an EU-level observatory to ensure comparability, transparency, and regular monitoring. 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*, has echoed these concerns with explicit calls for reform.²

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

²*European Parliament Resolution on cultural diversity and the conditions for authors in the European music streaming market* (European Parliament 2024); it recognises streaming as the dominant global revenue source while leaving many authors with very low income (recitals F–H), stresses accurate metadata allocation at the time of creation using identifiers ISWC, ISRC, ISNI, IPI, and IPN (recital R), 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).

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

1. **Practical solutions**, grounded in interdisciplinary dialogue between research and industry, and inspired by concrete experiences with open, federated data-sharing approaches. These solutions can reduce duplication of work, lower costs, and improve interoperability across rights management, libraries, archives, and digital platforms.
2. **Potential pitfalls** in the implementation of certain policy proposals, particularly where well-meaning initiatives may clash with the realities of legacy systems, existing business practices, or contradictions in legislation. Our aim is to point out where promising ideas might fail without careful attention to operational details.
3. **Legal and operational conflicts**, such as the tension between GDPR’s data protection regime and the Berne Convention’s requirement of author attribution, or the challenge of creating a more open and complete metadata registration system while at the same time sustaining incentives for costly — and currently mostly private — investments in the maintenance of key registers like ISWC (works), ISRC (recordings), ISNI (creators), and IPN (performer numbers).
4. **Cooperation and workflow sharing**, recognising that no single actor in the music ecosystem can bear the full burden of metadata documentation. Federated registries and data spaces allow information to be captured once and reused many times, reducing duplication and improving quality for all.
5. **Technology** for documentation, including automation, entity recognition, reconciliation, and persistent identifiers embedded directly in files. These tools can ease manual burdens but also bring governance challenges that must be carefully managed.
6. **AI adaptation and cooperative infrastructures**, recognising that most stakeholders — micro-enterprises, NGOs, and even CMOs — cannot attract or retain scarce AI expertise. Instead of expecting each actor to develop in-house capacity, Europe must invest in shared AI infrastructures, open-source tools, and collaborative governance. This requires both educating the sector about what AI really is — agentic, generative, and curative — and building collective resources that allow smaller actors to benefit without dependence on global platforms.

By foregrounding these issues, the brief contributes to the broader direction set by the European Parliament and the Commission. It stresses that any effective response to the transformations brought by AI — both agentic and generative — must include a comprehensive overhaul of metadata practices, robust governance of identifiers, and integration with open science and cultural heritage infrastructures. In this sense, our approach complements the calls of the *Music Ecosystem 2025 study* and the feasibility study for a **European Music Observatory** for systemic, ecosystem-wide policies, while remaining attentive to the practical challenges of implementation across Europe’s diverse music and cultural landscapes.

We see a wide policy consensus on a wide range of issues: streaming has centralised value capture, AI has intensified disruption, and Europe needs systemic solutions. But consensus at this level remains vague. To design practical reforms, we must diagnose the precise pressures facing rights-holders, cultural institutions, and small industry actors.

Three structural pressures frame today’s metadata challenges:

1. **Extreme efficiency pressure.** Music is now monetised in micro-transactions worth a fraction of a cent. To equal the economic weight of a single CD, rightsholders must process and account for thousands of streams. 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.
2. **AI-driven disruption.** Agentic AI in streaming platforms has already displaced much of the traditional retail and promotion infrastructure. Generative AI now risks flooding platforms with derivative works and further destabilising discoverability and revenues. At the same time, AI tools could help with documentation, reconciliation, and translation — if governance frameworks can support them.
3. **Governance and incentive conflicts.** Key identifiers such as ISWC (works), ISRC (recordings), ISNI (creators), and IPN (performers) are essential for attribution and royalty distribution. Yet they are maintained under costly, largely private regimes. Public policy increasingly demands more open and complete metadata, but sustaining investment in these registers remains a challenge. This creates fundamental tensions between openness, compliance, and long-term viability.

These pressures mean that improving metadata is not only a matter of technical interoperability. It is a question of economic sustainability, legal coherence, and cultural policy. The following sections map how these pressures play out in practice and what directions of solution are possible.

1.1 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 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.

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. It showed how 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 [newglamservices].

[newglamservices:] See for example the Katona József Library’s adaptive strategies in (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).

The economic and technical pressures described above have made metadata efficiency a make-or-break issue for both industry and heritage institutions. To meet this challenge,

we identify five families of solutions, each of which combines insights from research with lessons from our own experiments.

Data, metadata, blockchain and AI

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**, far richer than anything Beethoven or his contemporaries could have imagined.

As these works move through their lifecycle — composition, registration, performance, recording, distribution, preservation — they accumulate further **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 it often remains fragmented, messy, or with gaps — the “shadows” that Karabinos has described.

Provenance (PROV) Data Model Applied for Music

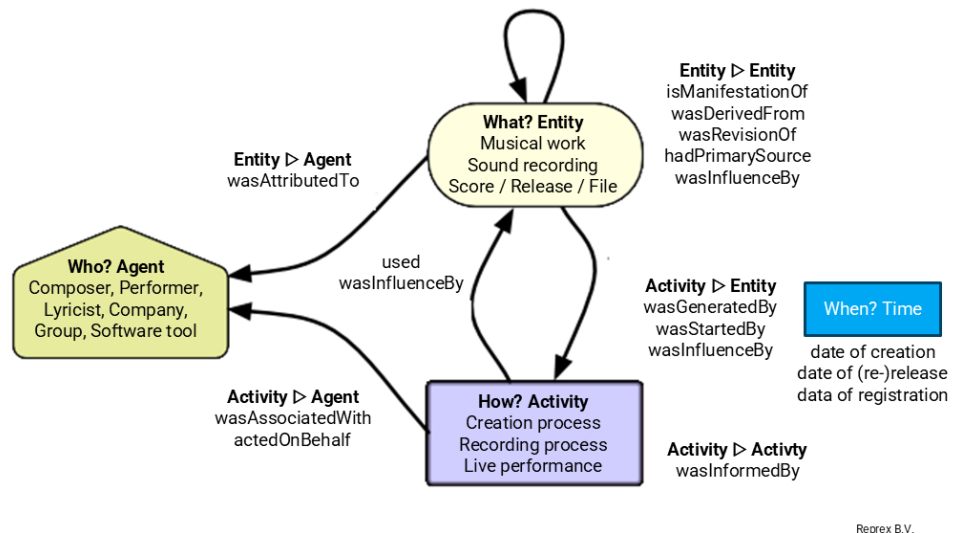


Figure 1.1: The PROV model helps us describe the lifecycle of music: who did what, when, and with what. A composer, performer, or software tool (agent) engages in an activity such as composing or recording, which results in a musical work or a sound recording (entity). Capturing these links over time makes provenance transparent, ensures correct attribution, and supports trustworthy data exchange across the music sector. Reuse: DOI: 10.6084/m9.figshare.30073210

Here, Pomerantz’s classic definition is a useful anchor: metadata is “data about data.” But in practice, this boundary is fluid. **Data by itself is inert** — a duration, a string of characters, a digital checksum is meaningless without context. Metadata transforms

it into a **knowledge statement** with potential truth value: “*This recording lasts 7:35.*” “*This file is identified as ISRC XY-ABC-23-00001.*” Crucially, what is “data” for one actor can be “metadata” for another. For a notation program, “7:35” is a descriptive property; for a rights manager, the same value may serve as an identifying attribute; for Spotify, it becomes one feature among many in an algorithmic profile. In a lifecycle-based data sharing space, data and metadata are not absolute categories — they are relative to role, workflow, and context.

This distributed, evolving record of provenance invites a comparison with **blockchains**, which are designed as tamper-evident, time-stamped ledgers of statements about digital objects. In music, we already have something similar in principle: every registration, every file embedding, every archival action adds another statement to the ledger of a work’s lifecycle. But unlike blockchains, this record is not unified, nor cryptographically secured. It is scattered across collective management organisations, distributors, streaming platforms, and archives. More seriously, the data is often inconsistent, disrupted, lost, or hidden due to conflicts of interest. Simply applying blockchain technology to such noisy data would not solve these problems; indeed, it could lock in their dysfunction. Curative AI or blockchain might one day help secure the chain of provenance, but only if the underlying parties work together rather than against one another.

Conceptually, however, the **born-digital lifecycle of music already resembles a distributed chain of provenance statements**: some verifiable, some contradictory, some lost in the shadows. The challenge is not to create a single immutable ledger, but to ensure that across this distributed chain, the statements that matter can be made reliable, reusable, and interoperable.

1.2 Potential solutions

The challenges described above call for **technical, organisational, regulatory, and governance responses**. To keep the structure manageable, we organise them into three layers that build on each other:

1. **Fixing Music Data at the Source** [Chapter 2](#)
2. **Open Music Observatory: Building a Shared Music Data Space** [Chapter 3](#)
3. **AI That Works For Music, Not Against It** [Chapter 4](#)

Standards, regulation, and public investment are horizontal topics that are mentioned in each layer where relevant, and recapped together in the conclusions.

1.2.1 Fixing Metadata At Source

No single actor in the music ecosystem can afford to duplicate the full burden of metadata documentation. Rights managers, distributors, libraries, and archives each collect and curate information, but they often do so in parallel, repeating the same effort. A cooperative approach can reduce these inefficiencies. Federated registries and data spaces

— such as those envisioned in the *European Interoperability Framework* — provide a model where each institution retains its role and mission, but contributes to a shared pipeline. In this way, data captured once can be reused many times across the lifecycle of a work or recording, lowering costs and improving quality for all.

Technology has a critical role to play in reducing manual burdens. Tools for automation — ranging from AI-supported entity recognition, transcription, and translation, to metadata extraction — already assist libraries and archives in coping with overwhelming digital volumes. Semantic reconciliation tools, such as those piloted in the MERA and Music Meta Ontology projects, show how databases can be linked pragmatically. Embedding persistent identifiers directly in files ensures that crucial information survives transformations and transfers. Yet, these technologies are not a panacea. Each efficiency gain brings new governance challenges, particularly around bias, explainability, and sustainability.

Industry-led initiatives, voluntary standards, and self-regulatory frameworks also have a vital role. History shows the value of lightweight, shared conventions: music notation itself has served for centuries as a simple but enduring metadata system. Today, the lesson is to avoid over-generalisation and monolithic ontologies. Instead, modular ontology patterns provide flexible bridges between systems, allowing interoperability without rigidity. Linking identifiers across domains — ISRC to ISWC, ISNI, VIAF, or Wikidata — can reduce friction in rights and documentation workflows. New forms of modular licensing and industry codes of conduct may also help reduce uncertainty without heavy-handed regulation.

Regulation has both driven and complicated the push for efficiency. European measures such as the *Collective Rights Management Directive* sought to increase transparency and accountability, reinforcing the need for better metadata. At the same time, the *General Data Protection Regulation (GDPR)* introduced new compliance burdens, particularly around personal names and performer information, which remain essential for attribution. Regulation should support, not hinder, the ability of cultural and commercial actors to cope with the enormous pressure of metadata management³.

Finally, public investment is indispensable. *The European Open Science Cloud (EOSC)*, the *European Collaborative Cloud for Cultural Heritage (ECCCH)*, national libraries and archives, and global platforms like Wikidata are all infrastructures that already serve scientific, cultural, and civic goals. With the right design, they can also support the music industry. Public investment into shared metadata infrastructures can lower the costs borne by small and medium-sized enterprises, collective management organisations, and self-releasing artists, ensuring that cultural diversity and economic sustainability are not sacrificed in the drive for efficiency.

The projects and drafts developed within our consortium — the *Slovak Comprehensive Music Database*, the *Open Music Observatory*, *MusicBase*, *Unlabel*, and *Open Music Reg-*

³*General Data Protection Regulation* Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (European Parliament and Council 2016); *Directive 2014/26/EU of the European Parliament and of the Council of 26 February 2014 on collective management of copyright and related rights and multi-territorial licensing of rights in musical works for online use in the internal market* (European Parliament and Council 2014).

isters — all attempt to cope with these pressures in complementary ways. They start from the insight that metadata cannot be managed by any single actor; and centralised data solutions anyways failed in the decentralised music industry riddled with conflicts of interests.

1.2.2 European Policy Baseline for Data Spaces

The European Commission has placed *Common European Data Spaces* at the heart of its data strategy, and it is also committed to this approach in the cultural domain. These are defined as federated infrastructures: participants keep control of their own data, but agree on common specifications for governance, interoperability, and trust. The *Data Spaces Support Centre (DSSC)* has published its *Blueprint v2.0*, which sets out key concepts (participants, rulebooks, services) and cross-data space interoperability patterns.

This provides the policy baseline: Europe is committed to federation, not centralisation. For music, this is encouraging but also challenging. Our sector is among the most structurally fragmented — with rights, metadata, and content dispersed across hundreds of micro-actors — which makes it a particularly demanding test case for the DSSC’s common specifications. As we discuss in Chapter 2, it is not only impractical, but theoretically impossible to build one schema that could simultaneously meet the needs of a CISAC-member author society, an IAML-member music library, an IAMIC-member music information centre, an IFPI-affiliated producer association, an AEPO-ARTIS performer society, and a commercial music distributor like ALOADED.

A **data sharing space** is designed precisely for such cases. It is an intelligent layer that enables near-instantaneous exchange, processing, sharing, and use of data on an “*as-needed*” or “*as-permitted*” basis — while each data holder retains complete control over who can access their data, under what conditions, and when (Curry 2020).

This does not happen automatically. It relies on **pragmatic, modular alignment** of metadata schemas and ontology patterns, recognising the differences between a library workflow and a collective management agency, and carefully designing who can press the “share” button under which conditions. In other words, a data sharing space is built through meticulous digital service design — exactly the approach promoted in the European Interoperability Framework.⁴

1.2.3 The Open Music Observatory as a Data Sharing Space

The fragmentation of today’s music metadata ecosystem forces every institution — rights managers, distributors, libraries, archives — to repeat the same documentation work in parallel. This wastes scarce resources and leaves smaller actors unable to keep up with

⁴For further reference see (Curry 2020), its application for cultural and creative sectors (EBU and Gaia-X 2022, 16), and (Nagel and Lycklama 2021). For the Blueprint 2.0 of the Data Spaces Support Centre (DSSC) see (Data Spaces Support Centre 2025b, 2025a). For the Commission’s recommendation on a common European data space for cultural heritage see (European Commission 2021).

the standards set by global platforms. Cooperation is therefore not only a question of efficiency, but of survival. Shared workflows can reduce duplication and, at the same time, create the **critical mass needed to negotiate with vertically integrated platforms** that dominate today’s music economy. Our proposal is to develop a **decentralised data sharing space**, where data captured once can be reused many times across the lifecycle of a work or recording. This is the principle underpinning the Open Music Observatory: not centralisation, but federation and reuse.

The evolution of **data sharing spaces** and our **Unlabel** experiment point to new business practices that integrate legal compliance, organisational coordination, and collaborative workflows. By combining public institutions (libraries, archives) with private distributors in a shared metadata infrastructure, supported by open source tools, we can reduce friction and align incentives. This is not theoretical: Unlabel demonstrates how librarians and rights managers can prepare metadata once for both distribution and long-term preservation.

As Norman Paskin warned two decades ago, standards such as ISRC, ISWC, ISMN, ISNI, and VIAF were developed in silos, without sufficient attention to interoperability. The result is inefficiency: named entity resolution today can consume up to **30% of back-office costs** in rights organisations and distributors. We do not reject standardisation — notation itself shows its long-term value — but we argue for a pragmatic, **pattern-based and modular approach** to standards, avoiding monolithic frameworks while improving crosswalks and reconciliation.

European regulation has both improved and complicated metadata practices. The *Collective Rights Management Directive* improved transparency, while the GDPR created new compliance burdens for names and personal data that remain essential for attribution. Regulation should be recalibrated to **support coping with metadata pressures, not add to them**. We will review these cases and suggest adjustments.

Europe has invested heavily in infrastructures such as the *European Open Science Cloud* (EOSC), the *European Collaborative Cloud for Cultural Heritage* (ECCCH), and *Europeana*. Yet these often remain disconnected from the highly digital private sector of the music industry. Better public-private alignment is needed. We argue for **bridges between public investment and industry use cases**, so that collective resources support both cultural preservation and competitive participation in digital markets.

The chapter **Open Music Observatory: Building a Shared Music Data Space** extends this topic with a discussion and policy recommendations for consultation. Chapter [3](#)

1.2.4 AI that Serves the Music Ecosystem

Advanced data documentation has so far been the preserve of large technology companies. They employ data engineers, ontologists, and machine-learning experts to automate the collection and management of music metadata — a capability that has directly increased their market share at the expense of rightsholders. To close this gap, **technological adaptation must reach small and micro companies** as well. Open-source software,

ontology design patterns, and libraries that package the latest advances in information science and AI can be embedded into the everyday workflows of labels, publishers, and cultural institutions. In this way, rightsholders and memory institutions alike can benefit from automation and reconciliation tools without the prohibitive costs of in-house R&D. Our pilots demonstrate how such adaptation can be made practical and affordable.

The **AI adaptation and cooperative infrastructures** dimension recognises that most stakeholders — micro-enterprises, NGOs, and even CMOs — cannot attract or retain scarce AI expertise. Instead of expecting each actor to develop in-house capacity, Europe must invest in shared AI infrastructures, open-source tools, and collaborative governance. This requires both educating the sector about what AI really is — agentic, generative, and curative — and building collective resources that allow smaller actors to benefit without dependence on global platforms.

Shared AI services could operate as pooled utilities behind clear agreements: reconciliation-as-a-service, watchlists for plagiarism and near-duplicates, machine-assisted translations and summaries for archives and deposits, and automated assistance for identifier coverage. Governance must include explainability notes, audit logs, error budgets, and “green lists” of permissible automations per partner. Standards and licensing for models and data, GDPR and consent for personal data, and targeted public investment in cooperative AI infrastructures all form part of the solution.

The chapter **AI That Works For Music, Not Against It** extends this topic with a discussion and policy recommendations for consultation. Chapter 4

1.3 Case studies

Due to the complex governance of the Open Music Observatory, we piloted its technical, semantic, and organisational layers with the *Finno-Ugric Data Sharing Space*. With the curation of the *Finno-Ugric Heritage NGO*, we placed traditional and contemporary popular vocal music of several Finno-Ugric ethnic groups, sung in endangered or critically endangered European languages. The methodological challenges of this pilot were first presented at **Wikimedia CEE Summit 2024**, where we introduced the technical aspects of our work in Slovakia. We created a poster, a conference presentation, and a prototype for the **DHNB 2025 ‘Digital Dreams and Practices’ Conference** in March 5–7, 2025, Tartu, Estonia to receive feedback from data curators and semantic scholars.⁵

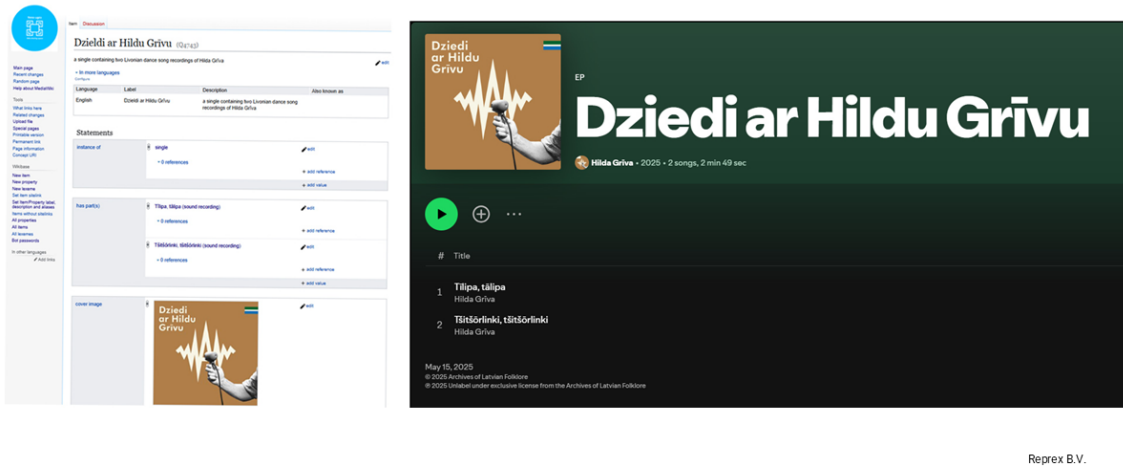
One of the case studies of our work is the first broad release of sound recordings of the now linguistically moribund European minority group, the Livs or Livonians. This Finno-Ugric

⁵We are particularly grateful for the feedback of the authors of the *8-Star Linked Open Data Model* — a key methodological upgrade beyond FAIR for multilingual, participatory contexts (Hyvönen and Tuominen 2024), and Alicia Fagerving, the author of *Wikidata for Authority Control: Sharing Museum Knowledge with the World*, a publication that demonstrates participatory use of Wikibase for museum knowledge (Fagerving 2023), received in a very productive pre-conference workshop. Our poster presentation can be found at (Daniel et al. 2025).

speaking group had already been critically endangered in the 1930s. After the Soviet occupation of Latvia, the remaining 1500–2000 Liv speakers were dispersed from the *Livonian Coast*, a congruous area of about 14 settlements where the language had been spoken. Following the re-establishment of Latvian independence, the country constitutionally pledged to preserve the culture of this small indigenous ethnic group.

Broad Interoperability: From Archive To Spotify

Our data model supports only “patterns” of important standard library, archive, museum, rights management conceptual models; functionality is not optimised for libraries but for cross-institutional use



Reprex B.V.

Figure 1.2: By repairing and harmonising metadata, archival recordings like *Dziedi ar Hildu Grīvu* can travel across systems — from museum or library catalogues into Spotify. This shows how the Open Music Observatory enables cultural heritage and commercial platforms to speak the same language.. DOI: [10.6084/m9.figshare.30075313](https://figshare.com/articles/figure/Broad_Interoperability)

Together with the *Latvian Archives of Folklore* and the Finno-Ugric NGO *Hõimulõimed*, we launched a data curation and repair programme to locate, document, and enrich Livonian music. This effort aims not only at scholarly preservation but also at making these materials available on popular commercial platforms, thus reconnecting heritage with contemporary cultural circuits. This exercise allowed us to reflect on some contemporary issues in data curation, and the need to go beyond the traditional 5-star FAIR model, expanding our focus beyond technical and semantic interoperability. Furthermore, it shaped our ideas about an ethical decentralisation that follows European subsidiarity principles.⁶

We will return to this example as a case study in the subsequent chapters, to demonstrate

⁶Our work draws on *Opening Archives: Respectful Repatriation* (Christen 2011), which critiques the circulation of digitised heritage without community authority, and *Metadata and Linked Open Data in Digital Heritage for Decolonization* (Dinler 2025), which highlights the importance of participatory and decolonial approaches in linked data infrastructures. Both perspectives informed our interpretation of subsidiarity and ethical decentralisation.

how community-based archival work can be aligned with European-scale data infrastructures.

2 Fixing Music Data at the Source

Scope: Bottom + middle layers combined: primary data capture/curation; technical & semantic improvement; identifier hygiene; practical exchange between institutions.

🔥 Executive summary (what we propose)

This is a placeholder

- Establish shared, low-friction pipelines for capturing and improving music meta-data at source (creation, release, registration, distribution, preservation), with simple mechanisms for reuse across stakeholders.
- Prioritise persistent identifier coverage and crosswalks (ISWC ISRC ISNI IPN VIAF Wikidata) to reduce reconciliation costs and leakage.
- Produce lightweight exchange patterns (profiles) that smaller actors can actually implement.

Data curation is the organisation and integration of data collected from various sources.

It involves annotation, publication and presentation of the data so that the value of the data is maintained over time, and the data remains available for reuse and preservation.



Reprex B.V.

Figure 2.1: Curating data from multiple sources ensures that music information stays accurate, visible, and reusable over time. Reuse (DOI): [10.6084/m9.figshare.30073888.v1](https://doi.org/10.6084/m9.figshare.30073888.v1)

2.1 Discussion

i These are pre-finalisation notes

- Streaming-era inefficiencies and micro-transaction reality; identifier silos and reconciliation overheads; feasibility/Observatory emphasis on harmonised, comparable data; *Music Ecosystem 2025*’s ecosystem framing.
- Industry practice: CISAC/IFPI/ISO identifier regimes; library/archives practice on authority files; Wikimedia/Wikidata as low-cost backbone.

2.1.1 Data fragmentation as a structural fact

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.

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).

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.

2.1.2 Cost Barriers in Documentation and Claims

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

poor discoverability on platforms, which in turn depresses revenues and leaves even less money for proper documentation.

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

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

These cost barriers are not isolated bookkeeping problems — they are structural features of music data curation. How a data sharing space can provide scale effects and relieve these constraints is discussed in Section 3.3.1.

2.1.3 Why one grand collection model will not work

Every actor in music — a library, an archive, a label, a rights society — has its own way of defining what is 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. These limits are well understood, but recognising them is not an excuse for inaction. It means we should work pragmatically: push systems closer to their theoretical limits, rather than accept today’s very imperfect state of affairs in music data.

i Why collections differ in databases

- **Libraries** collect under *legal deposit rules*: every book or score published in a country must be included, regardless of popularity.
- **Archives** follow *provenance*: they keep what an organisation or individual produced, not necessarily what is “important.”

- **Collective management organisations (CMOs)** must register *only what their members submit* — the collection reflects contracts and repertoire, not cultural completeness.
- **Distributors** take what their clients release: the “collection” is shaped by market demand and contracts.

Each of these logics is valid, but none can be reduced to the others. This is why a single “grand ontology” for all collections is not achievable. The pragmatic solution is to connect them through lightweight, modular patterns that allow data to flow across boundaries while respecting institutional differences¹.

This does not mean that we have to give up on cooperation and interoperability. Understanding the theoretical, absolute limits of interoperability allows us to go beyond the current state of the art and reach towards the limits.

This insight was already realised in the music metadata project Polifonia, which came to the conclusion that music is better served with modular ontologies (or data models) that share their design principles but do not try to solve every problem in one conceptual framework.

The solution we propose builds on these foundational results in information science, collections management, and music metadata research: the future European Music Observatory should work pragmatically — capture data once, reuse it many times, and connect systems through lightweight, modular patterns that respect institutional differences.

2.1.4 Legacy Metadata

The European Parliament has emphasised that accurate and standardised metadata is essential for ensuring fair remuneration and proper attribution in the music streaming market. It calls for identifiers such as ISWC, ISRC, ISNI, IPI and IPN to be allocated at the moment of creation, highlights the systemic problems caused by poor data quality,

¹As 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 “giga-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. Research on **Ontology Design Patterns (ODPs)** and the **Extreme Design (XD)** methodology underlines that interoperability can be achieved more sustainably by modelling recurring structures (e.g. Agent–Role–Activity, Work–Recording–Performance, Event–Time–Place) rather than enforcing schema unification ((Gangemi 2005); (Blomqvist, Hammar, and Presutti 2016)). The Polifonia project has already demonstrated this approach in practice, building a modular ontology network for musical heritage without attempting a monolithic schema (Berardinis et al. 2023). 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.

and warns that the flood of AI-generated tracks further exacerbates discoverability and revenue imbalances if metadata remains incomplete or inconsistent.²

While we fully agree with this goal, we have witnessed in the past years how difficult it is to achieve. This is partly due to data protection issues (introduced later), and partly because the registers that underpin music metadata are privately governed, require continuous investment, and cannot be rebuilt from scratch. Hundreds of millions of assets are in circulation, with billions of transactions handled annually, all of which rely on this existing infrastructure. Moreover, the very use of the word *metadata* is misleading in the music industry: while in libraries and IT it usually means descriptive information (title, genre, provenance), in music business practice it refers narrowly to administrative identifiers that drive royalty distribution. This terminological gap itself contributes to misunderstandings and misplaced expectations.

The *International Standard Recording Code* (**ISRC**) was introduced in 1986 as a 12-character identifier for individual sound and video recordings, formalised in ISO 3901 (ISO 2019a). Operational management was placed with the IFPI, which continues to publish practical guidance (International ISRC Registration Authority 2021). National agencies began implementation in the late 1980s and early 1990s, and by 1991–1992 IFPI and the RIAA had circulated operational recommendations. A revision of ISO 3901 in 2001 and later updates sought to address shortcomings in registrant roles and metadata practices as digital distribution became dominant. Despite these efforts, persistent problems include retroactive assignment, inconsistent embedding, and weak interoperability with other identifiers such as ISWC (Paskin 2006, p4). The ISRC system reflects the pre-internet era: much like the *International Standard Book Number* (**ISBN**, ISO 2108) which predated widespread online catalogues, ISRC relied on decentralised registrant ranges distributed by post and manually reported back (ISO 2017a).

The *International Standard Musical Work Code* (**ISWC**) was standardised later, as ISO 15707 (ISO 2022), and is managed centrally by CISAC through the ISWC Agency. It was designed to identify musical works (compositions, lyrics), complementing ISRC’s recording-level scope. Yet adoption has been hindered by data-quality problems: duplicate ISWCs for the same work, mismatches between ISWC and ISRC, and governance challenges within collective management organisations. As Paskin observed nearly twenty years ago, the fundamental issues of interoperability and reliability across identifier systems were already well recognised but unresolved (Paskin 2006, p7).

Often overlooked in this debate is the *International Standard Music Number* (**ISMN**, ISO 10957), which identifies printed music publications (ISO 2021). ISMN plays a bridging role: it links the abstract work level of ISWC with the concrete manifestations of print editions, enabling continuity between bibliographic and rights-management practices. Despite its potential, ISMN remains underused in digital workflows and is rarely considered in discussions of streaming-era metadata.

Together, ISRC, ISWC, and ISMN form the backbone of music identification. Their coexistence illustrates both the strength and fragility of the current system: robust in theory,

²European Parliament resolution of 17 January 2024 on cultural diversity and the conditions for authors in the European music streaming market, recital 32.

but in practice plagued by legacy design, inconsistent uptake, and high maintenance costs. This makes the European Parliament’s ambitions difficult to realise without new layers of interoperability, observability, and shared responsibility.

2.1.5 Named-Entity Resolution, Attribution, and Privacy

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

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 3.2.3, where we show how data sharing spaces provide a way forward.

2.2 Policy proposals

- Pilots: Slovak CMD + MusicBase for capture and reconciliation; Open Music Registers for cross-domain linking; Unlabel for “document once, reuse many” (distribution + preservation).
- Deliverables: (a) “Capture once, reuse many” profile pack; (b) minimal identifier coverage checklist; (c) recommended crosswalk templates and validation scripts.
- Standards/regulation/public investment: call out per subsection (e.g., GDPR for name exposure; DCMI/METS/MODS/DCAT/JSON-LD profiles; EOSC/ECCCH/Europeana alignment).

2.2.1 Reducing Redundancy

The European Parliament has rightly highlighted that fragmented and unreliable metadata remains a major obstacle in the music sector. We agree with this diagnosis, but stress that the root cause lies partly in the need for backward compatibility with several hundred million 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.

Our policy solution is to encourage the parallelisation of registration processes and the harmonisation of the data that enters costly, manually curated registers. We demonstrate this approach with our *Open Music Registers* pilot: a federated infrastructure designed to enable harmonised, parallel, and redundancy-free registration. Instead of duplicating efforts, the registers interconnect persistent identifiers (ISWC, ISRC, ISNI, VIAF) and link them where relevant to business and statistical identifiers (OpenCorporates, NACE, ISCO). This allows music creators and organisations to benefit from smoother workflows, while downstream users gain more reliable data for royalty distribution, cultural visibility, and AI-driven discovery.

The Open Music Registers deliberately avoid centralisation. Each registrar—collective management organisations, libraries, archives, or statistical offices—retains ownership of its data, but contributes to a shared semantic framework built on the European Interoperability Framework and the 8-Star FAIR model. By connecting rather than merging registers, we reduce redundancy while safeguarding subsidiarity, accountability, and trust across public and private actors. This distributed model directly answers Parliament’s call for metadata systems that are reliable, inclusive, and supportive of creators.³

As the DSSC’s *Blueprint v2.0* notes, identifiers and rulebooks are the foundation of any data space (Data Spaces Support Centre 2025b). For music, however, attribution identifiers themselves are caught in the GDPR contradiction (see Section 2.1.5).

2.2.2 Reconciling Attribution and Privacy

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

³For a full methodological discussion and technical architecture, see (Antal and Mester 2025).

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

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

2.2.3 Pragmatic Metadata Alignment (Enhanced with Cross-Domain Examples)

Attempts to build one comprehensive, harmonised schema for music metadata have repeatedly failed. The sector is too heterogeneous: collective management organisations, libraries, archives, distributors, and platforms all operate with different standards and governance models. Pursuing a single “giga-ontology” would be prohibitively costly, brittle, and ultimately futile.

Our policy solution is to promote *pattern-based modular alignment*, where small, reusable ontology design patterns (ODPs) are used to model recurring structures such as **Agent–Role–Activity**, **Work–Recording–Performance**, or **Event–Time–Place**. This method, formalised in the *eXtreme Design* (XD) approach to be used together with pattern-based development, allows interoperability to be achieved incrementally and pragmatically, without forcing any actor to abandon its systems. The Polifonia project has already demonstrated how XD and ODPs can underpin modular music ontologies at European scale [^polifonia].

[^polifonia:] Working with ontological patterns can be reviewed in (Gangemi 2005; Blomqvist, Hammar, and Presutti 2016; Carriero et al. 2021) as method; the Polifonia project applied it to create a new family of modular music ontologies (Berardinis et al. 2023).

Other domains provide useful analogies for this approach. For instance, **ORCID** identifiers have been increasingly reconciled with **VIAF** authority files using reconciliation services such as OpenRefine, helping to unify researcher identity across research and library systems without enforcing a single authority file Jegan et al. (2023). Similarly, the **DataCite to Dublin Core mapping**, as formalised in DataCite’s v4.4 schema, shows how two widely used metadata data repository and library standards can be aligned systematically and bidirectionally, while retaining their distinct schema scopes (DataCite 2021).

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

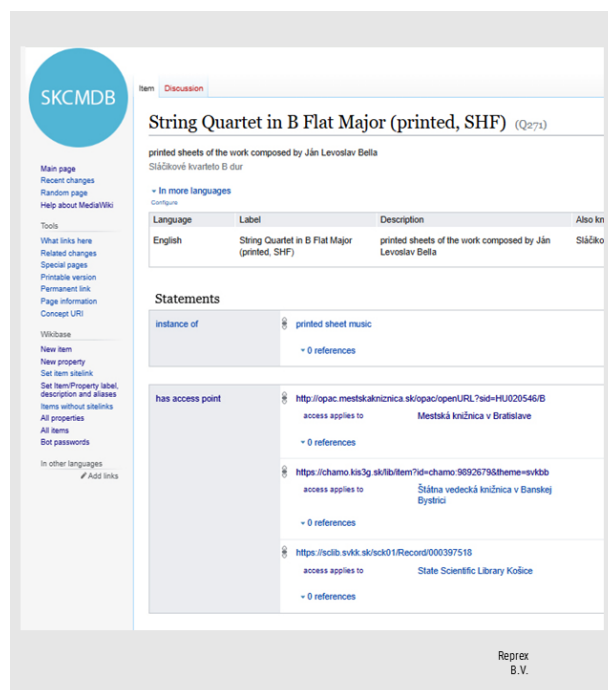


Figure 2.2: Pragmatic metadata alignment relies on modular patterns, not “giga-schemas.” The example shown here from our Wikibase pilot encodes roles, events, and provenance using reusable ontology design patterns. This allowed identifiers from rights management (ISWC, ISRC) to be reconciled with library 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)

In music, similar periodic reconciliation is necessary across identifier systems. As Paskin highlighted, **ISRC** (for recordings), **ISWC** (for works), and **ISMN** (for printed music) were designed separately and risk divergence if not actively maintained (Paskin 2006). The same applies to personal and organisational identifiers like **ISNI**, **VIAF**, and **IPI**. Without regular cross-checking, authority records can fragment, causing duplication and inconsistency.

In our pilots, this modular alignment approach was tested effectively. The *Slovak Comprehensive Music Database* used modular ODPs to reconcile rights identifiers (ISWC, ISRC) with library authorities (VIAF, ISNI) without schema unification. *MusicBase* implemented these patterns in Wikibase, encoding roles, events, and provenance via property–qualifier bundles and enabling corrections in one register to propagate to others. The *Unlabel* pathway streamlined metadata capture for self-releasing artists and libraries by allowing a once-only documentation process, reused across distribution and preservation systems. This builds directly on our *Open Music Registers* proposal (Antal and Mester 2025), which argued for federated, redundancy-free metadata workflows, now extended into the broader governance framework of this Green Paper.

By focusing on modular patterns instead of monolithic schemas, we lower entry costs, reduce reconciliation errors, and support federated alignment flexible enough to accom-

modate European diversity. This pragmatic approach answers Parliament's demand for interoperable metadata systems (European Parliament 2024, 19), while avoiding the shortcomings of over-centralisation and standardisation.

3 Open Music Observatory: Building a Shared Music Data Space

i Open Music Observatory

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

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

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

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

¹Always 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.

3.1 Executive summary

Caution

This will be finalised after consultation

- Build a federated data sharing space that lets rights orgs, distributors, platforms, libraries/archives share *just enough* metadata with clear roles, auditability, and lifecycle continuity.
- Use simple, modular agreements and conformance tests rather than a heavy centralised repository.

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.

Open Music Observatory Timeline

From CEEMID via Open Music Europe towards a decentralised music observatory data sharing space

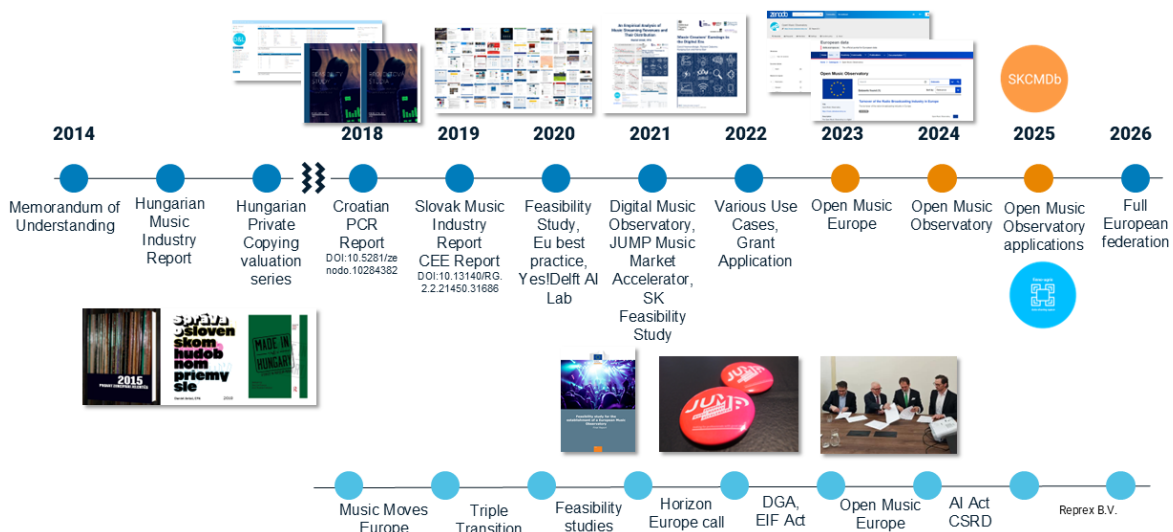


Figure 3.1: Over the past decade, feasibility studies, national reports, and EU pilot projects have laid the foundation for the Open Music Observatory. The roadmap (2014–2026) shows a gradual build-up: from local experiments, through cross-border collaborations, to a European-wide federation aligned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

3.2 Discussion

Caution

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

3.2.1 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.

EU infrastructure initiatives have already moved beyond this logic. Since the 2000s, projects such as *Europeana*, the *European Open Science Cloud (EOSC)*, the *European Collaborative Cloud for Cultural Heritage (ECCCH)*, and *DARIAH* have all adopted **federated architectures**, linking distributed collections through shared standards, profiles, 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 have adopted authority files (e.g. VIAF) and collaborative platforms (e.g. *Wikidata* and *Wikibase*) to enable interoperability while preserving institutional autonomy (Bianchini, Bargioni, and Pellizzari di San Girolamo 2021, p210; Sardo and Bianchini 2022, p297). The emphasis here is not only on efficiency, but also on *openness* and the ability to reuse metadata across cultural domains without restrictive licensing.

The two most important industry bodies supporting the commercial circulation of hundreds of millions of music assets, CISAC and IFPI, already operate on principles of federalisation. The ISRC system, managed by IFPI as the International Registration Authority, is inherently decentralised: 58 national agencies allocate prefixes, and codes are assigned by rights owners or their representatives (International ISRC Registration Authority 2021, p5). This model resembles ISBN for books or ISMN for sheet music: globally unique identifiers achieved through subsidiarity and distributed governance.

CISAC operates *CIS-Net*, a federated database of works and rights information maintained by its national member societies. Although users access CIS-Net through a centralised interface, the underlying data remains under the stewardship of each member. Likewise, 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 global repository (CISAC/SUISA/SESAC 2017).

These federated arrangements are not temporary fixes but structural features of global identifier systems. The ISRC system functions only because allocation is distributed among national agencies (International ISRC Registration Authority 2021, p5), while CISAC’s CIS-Net provides access to rights data without centralising ownership (CISAC/SUISA/SESAC 2017). The Observatory builds on this same principle: decentralised stewardship combined with shared interfaces and conformance rules.

Even official governmental statistics, thought to be highly centralised, have always relied on decentralised structures, particularly in the field of culture. The **ESSnet-Culture** project, coordinated under Eurostat, produced the first comprehensive framework for cultural statistics in 2012. Its final report is described as a “*basic reference*” for European culture statistics, elaborated by thematic task forces across Member States and grounded in both EU and **national data sources** (Commission et al. 2020, p9). The framework, adapted from the UNESCO model, spans ten domains (heritage, archives, libraries, performing arts, etc.) and six functions (creation, trade, preservation, etc.), demonstrating that official cultural indicators are constructed through *distributed cooperation among national agencies*.

This decentralised model is not an exception but a structural feature of European statistics. National statistical offices, labour force surveys, and administrative registers each collect partial data, which are subsequently harmonised at EU level for comparability. Public policy indicators are thus inherently the result of federated cooperation, rather than centralised data lakes. Recent research extends this point further. In the age of abundant data, cultural and creative indicators increasingly rely on the *reuse of private-sector information*. Surveys and administrative datasets are being complemented by data flows from platforms, rights management organisations, and other industry actors. By extending statistical practice into these domains, European statistics are entering a new phase where privately held datasets become part of official evidence frameworks, building on earlier decentralisation while adding new layers of governance and interoperability. This represents an increasing decentralisation of evidence creation, where official statistics and policy indicators emerge from hybrid constellations of public and private data sources².

This decentralised model is not an exception but a structural feature of European statistics. National statistical offices, labour force surveys, and administrative registers each collect partial data, which are subsequently harmonised at EU level for comparability. Public policy indicators are thus inherently the result of federated cooperation, rather than centralised data lakes.

This trajectory mirrors the broader evolution of European data infrastructures, where federation has become the dominant organising principle. The Music Ecosystem 2025 study framed the music sector as an ecosystem distributed across many small actors (Music Moves Europe 2024, pp6–7), while the Feasibility Study for a European Music Observatory highlighted fragmentation as the primary obstacle to evidence-based policy (Commission

²We discussed the possibility of creation music statistics in a public-private cooperation in (Antal 2023).

et al. 2020, p9). Both perspectives underline why federated architectures are more feasible than centralised repositories.

3.2.2 Open Data Directive: Right Without Means

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

Studies consistently show that open data remains more of a promise than a reality³. In practice, much open data is poorly documented, lacks common identifiers, and is released in unstandardised formats. While it may be free of charge or available at marginal cost, making it interoperable and trustworthy for cross-border use requires significant additional effort. The burden of curation, harmonisation, and enrichment falls on downstream users, which can be prohibitively expensive for smaller organisations. As the CEDAR project put it, “*Public authorities are only required to make existing data available, not to create new data or improve existing systems. This leads to significant disparities in usability and accessibility*” (Project 2023). A recent EU-wide usability study adds that “*many open data portals remain difficult to navigate, poorly documented, and inconsistent in their metadata quality, limiting actual reuse*” (Jachimczyk and Nowak 2024).

This gap is directly relevant to the music ecosystem. Simply publishing datasets under the Open Data Directive does not make them fit for royalty distribution, cultural statistics, or AI-driven discovery. The Open Music Observatory seeks to fill this gap by turning “open data in principle” into **reusable data in practice**, through shared identifiers, metadata profiles, and federation rules.

³Early analyses underlined that “*the economic potential of open data is significant, but realising this potential depends on overcoming barriers to data availability, quality and usability*” (Carrara et al. 2015, 7). Later modelling confirmed the vast estimated value of open data, but stressed that “*in practice, the impact of open data depends heavily on the extent to which it is discoverable, accessible, and usable*” (Huyer and van Knippenberg 2020, 14). This has direct consequences for domains with complex metadata like music, where rights and attribution depend on accurate, standardised, and machine-readable records. Empirical studies across sectors show why legal openness alone rarely yields machine-actionable reuse. This is not only a music-specific problem, though the need for high-quality, standardised data in music is unusually high. Analyses of EU public-sector openness find that, absent prescriptive technical mandates, agencies publish heterogeneous formats and divergent practices, which fragment reuse across countries and sectors (Buttow and Meijer 2024, 12). Regional benchmarking of high-value geospatial data similarly reports variability in machine-readable provision and metadata completeness, with practical usability suffering even where datasets exist (Kević, Kuveždić Divjak, and Welle Donker 2023, 3). Cross-holder case studies (e.g., mineral intelligence) continue to call for harmonised structures and output formats, underscoring the need for shared profiles beyond the baseline legal right to reuse (Simoni, Aasly, and Schjøth 2021, 5). Beyond formats, evidence shows that preparing legacy administrative data for actual reuse (cleansing, standardising, describing) imposes non-trivial costs on data holders and re-users—zero price does not eliminate acquisition, transformation, and integration costs (EuroSDR 2021, 9; Schnurr 2021, 14; Nakos and Tsoulos 2022, 6).

3.2.3 Why Voluntary Workarounds Do Not Scale

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

For observatories and data spaces, this is untenable. Interoperability requires clarity and legal certainty across borders and institutions. Without guidance from a Data Protection Authority or the European Commission, every national or sectoral initiative risks being challenged. The consequence 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. These unresolved attribution issues ultimately undermine not only observatories but also AI fairness and governance (see Section 4.1.3).

3.2.4 Public cultural infrastructures bypass music’s real data flows

Europe has invested heavily in Europeana, ECCCH and other cultural data infrastructures, but their data models and workflows remain weakly aligned with how music metadata is generated and maintained in practice — largely by private actors (labels, distributors, CMOs) who already finished most digitisation, including heritage repertoire. Unlike in the case of old libraries, documentary records in archives, artefacts in museums, where digitisation is mainly funded with public money, in music and film most of the digitisation took place (also for high-value classical, early or folk music) in the private sector.

The *Europeana Data Model* was based mainly on public collection practices and library models, and it cannot truly handle the fact that music that can be listened to and viewed has at least three creator groups that need to be assigned: authors, producers and performers. We have been proposing solutions for this to Europeana, and we hope that during the remaining period of OpenMusE we will be able to demonstrate our solution (see later.)

⁴

The *Report on a European collaborative cloud for cultural heritage* (ECCCH report) does not mention music at all. Its scope was conceived around heritage institutions (archives,

⁴The Europeana data model largely builds on DCTERMS, which is suitable for handling printed music sheets; they are regularly stored in library systems. Printed music needs a reference for the author and the publisher. However, sound recordings (including video sound recordings) must identify the neighboring rightsholders, the producer and the performer. See our suggestions later in this chapter. (Europeana 2017)

museums, libraries, monuments), and it overlooked the specific needs of the music sector.⁵ The first ECCCH projects focus on digitisation and advanced data infrastructures for cultural heritage, but none address music directly. **AUTOMATA** develops low-cost, AI-assisted digitisation workflows for archaeological ceramics and lithics, integrating enriched 3D data into the cloud. **TEXTaiLES** tackles the specific challenges of textile heritage, creating AI- and sensor-based tools for capturing, preserving, and restoring fabrics. **HERITALISE** advances digitisation across diverse heritage assets, using machine learning, CT scanning, HBIM, and XR/VR services for immersive representation and conservation. **ECHOES** establishes the core European Cultural Heritage Cloud, building a shared platform with interoperable tools, metadata vocabularies, and training for heritage professionals. We tried to establish contact with ECHOES, and we also suggested in their cascading grant program to fit in large music datasets that handle all aspects of music metadata; one of them was eliminated in the first stage, the other, related to a conceptual exploitation pathway of our Open Music Observatory, the Finno-Ugric Data Sharing Space, may still be able to try to insert music data into the ECCCH.

Even the **Polifonia project**, which we cited earlier as a good example and whose methodology we could use well, was “blind” to the private data infrastructure of music. While it created a set of modular ontologies, it did not cover rights management, and even its central classes for musical works and sound recordings were not fully aligned with the ISWC and ISRC standards used by industry to identify and register musical works and sound recordings.

This misalignment is structural, not incidental. The result is duplication and gaps: public knowledge graphs and registries don’t interoperate smoothly with industry identifiers (ISRC/ISWC/ISNI/IPN), reconciliation is costly, and culturally significant catalogues stay under-used across public and commercial systems. This is a missed opportunity for both sectors.

3.2.5 Alignment with the European Open Science Cloud (EOSC)

The **European Open Science Cloud (EOSC)** is intended as Europe’s flagship infrastructure for sharing and reusing research data across domains. By “EOSC data model” we mean the conceptual and technical framework that enables EOSC to function as a federated, distributed, and interoperable ecosystem for research data, tools, and services.

In practice, EOSC remains highly complex and fragmented, and its connections to music are limited. Music has been touched only tangentially, mainly through the work of **DARIAH-EU** (*Digital Research Infrastructure for the Arts and Humanities*), which has supported humanities-driven approaches to digital musicology but not the broader music ecosystem. Yet, just like in ECCCH, we see music underrepresented in the DARIAH-EU system.

⁵*Report on a European collaborative cloud for cultural heritage – Ex-ante impact assessment* (Commission et al. 2022)

In our understanding, the **ECHOES** project aims to build interoperability with cultural data providers as well as EOSC.

Our projects already interact with EOSC through **Zenodo** and **OpenAIRE**. We deposit music-related datasets on Zenodo, ensuring visibility, persistence (via DOIs), and EOSC integration. This creates a foothold for music data within EOSC, and we see that some of our deposits are highly visited. However, so far we do not see a clear curatorial pathway for music-related research being deposited in large quantities to EOSC, or industry being able to easily utilise music data from EOSC.

From a policy perspective, EOSC should not be overlooked: it is the default European infrastructure for open science, and cultural heritage and creative industry data are increasingly expected to align with it. For music, this means recognising where EOSC provides real value — long-term preservation, persistent identifiers, cross-domain discovery — while also acknowledging its gaps in rights-aware, industry-aligned metadata. Our strategy is to use EOSC as a **public backbone for open deposits**, while federating more specialised infrastructures (such as the Open Music Observatory) to ensure that music’s specific needs for identifiers, attribution, and rights management are respected.

3.2.6 European Interoperability Framework

The **European Interoperability Framework (EIF)** is the European Commission’s reference model for building interoperable digital public services across Member States. It defines interoperability not only in technical terms, but also across legal, organisational, and semantic layers, ensuring that administrations can exchange data and services smoothly while respecting national sovereignty.

Although the EIF formally applies to public administrations, its layered approach is highly relevant for the music ecosystem. Cultural data and rights metadata circulate between public institutions (such as libraries, archives, or ministries), private actors (platforms, CMOs, publishers), and hybrid organisations (NGOs, research infrastructures). By adopting the EIF’s multi-layered model, the Open Music Observatory ensures that federation is not limited to technical standards but also extends to governance, semantics, and legal compliance.

The European Commission’s *European Strategy for Data* already frames interoperability as a cornerstone of Common European Data Spaces, which are to be built as **federated infrastructures** underpinned by European laws and standards (European Commission 2020). BDVA and DAIRO echo this point, emphasising that interoperability frameworks such as the EIF are necessary to operationalise federation in practice (BDVA/DAIRO 2023, 18; BDVA/DAIRO Federation Working Group 2023, 4). For participants in the music ecosystem, this means lower integration costs and greater predictability: institutions already adapting to EIF-based rules in research or e-government can reuse those frameworks in the Observatory.

3.2.7 Subsidiarity in Platform Design

The European principle of subsidiarity requires that decisions be taken as closely as possible to the citizens they affect. In cultural policy, this translates into responsibilities distributed across multiple levels: in some Member States, culture is managed regionally or provincially; in others, at the national level. This creates a patchwork where data about music and cultural heritage is controlled by different authorities, depending on the administrative tradition of each country.

This diversity extends beyond public administrations. Many important datasets are held by private actors, including collective management organisations, commercial platforms, and non-profit archives. As a result, any attempt to centralise music data governance would risk losing both legitimacy and local relevance.

Designing the Observatory around subsidiarity means recognising these structural differences and embedding them into its architecture. The EIF helps here by providing a layered model for reconciling data governance (permissions, licences) at the legal level and organisational rules at the institutional level. This ensures that local, regional, national, and private actors can contribute data under their own governance, while outputs remain reusable across Europe.

The **Data Governance Act (DGA)** reinforces this approach. It establishes EU-wide rules for trusted data sharing and data intermediation services while leaving implementation and control at the Member State level (European Parliament and Council 2022). In other words, it codifies subsidiarity: national authorities retain stewardship over sensitive datasets, but EU-level standards guarantee that these datasets can circulate securely and comparably across borders. The Observatory applies this same principle to music: rights organisations, archives, and platforms remain in charge of their data, while interoperability, accountability, and observability are ensured through a shared European layer.

The European Data Strategy explicitly recognises subsidiarity by promoting data spaces that respect *sectoral and national diversity while enabling interoperability across borders* (European Commission 2020). The BDVA Federation Working Group goes further, arguing that federation is the only model compatible with Europe’s fragmented governance landscape (BDVA/DAIRO Federation Working Group 2023, 4). The Observatory takes these insights into account by embedding subsidiarity into its platform design: respecting local authority, while enabling European-scale reuse.

i Note

Our pilot with the Finno-Ugric Data Sharing Space illustrates how subsidiarity works in practice. By collaborating with regional NGOs and national archives, we were able to curate and repair 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.

3.2.8 Economies of Scale in Metadata

Large platforms and major labels manage to document millions of assets in parallel, achieving economies of scale that smaller actors cannot match. By contrast, for SMEs, non-profits, or community archives, the cost of documentation per asset is disproportionately high, often exceeding the commercial value of the repertoire. This gap explains why so many “frozen” assets remain unregistered and invisible in the digital ecosystem.

Agentic AI, if deployed within a shared knowledge base and aligned with modular ontologies in a federated data sharing space, can reduce operational expenditure (OPEX) by automating repetitive documentation tasks. This would allow smaller players to benefit from the same scale effects as the global platforms, without compromising quality or compliance. The frozen asset case is only the most visible example: AI-enabled economies of scale would lower costs across the entire long tail of music assets.

3.2.9 The U.S. Mechanical Licensing Collective (MLC) as a Metadata Clearinghouse

The *Mechanical Licensing Collective* (MLC) was created under the U.S. *Music Modernization Act* (2018) and began operations in January 2021. It administers a **blanket mechanical license** for digital streaming and download services, replacing fragmented song-by-song licensing. Its mandate is to ensure that songwriters, lyricists, composers, and publishers receive timely mechanical royalties, while providing transparency through a public works database and member portal (Mechanical Licensing Collective 2021).

From its inception, the MLC inherited more than **\$424 million in unmatched royalties** from digital service providers (DSPs)—funds that could not be allocated because works were not properly registered or matched. These unmatched sums must eventually be distributed, and if they cannot be claimed, they are paid out to publishers on a market-share basis. For independent songwriters, this creates both an opportunity and a risk: registration with the MLC is free, but failure to register means royalties may be permanently lost.

The MLC highlights a systemic lesson: identifiers like ISWC and ISRC must be correctly captured and maintained at source, or royalties remain trapped in the “black box.” To address this, the MLC has developed **matching and reconciliation routines** and offers search tools so creators can check whether their works are properly registered. Education remains a critical challenge, since many independent songwriters incorrectly assume that affiliation with ASCAP, BMI, or SESAC (performance rights organisations) covers mechanical royalties as well.

By late 2022, the MLC had already distributed nearly **\$700 million in royalties**, but its operations also illustrate the fragility of metadata-dependent systems. In 2025, it

filed—and lost—a high-profile lawsuit against Spotify, which had sought to classify audiobooks within its music service to reduce royalty obligations (Varghese 2024). This underscores that even with centralised licensing, **governance and enforcement remain contested**.

For European debates, the MLC demonstrates how a large-scale, rights-compliant clearinghouse can consolidate reporting, improve metadata quality, and distribute royalties more transparently. Yet it also shows the **limits of centralisation**: creators still need to actively claim and maintain their records, education gaps persist, and disputes between collective agencies and global platforms remain unresolved.

3.2.10 Data Sharing Space

A **data sharing space** is both a governance framework and a technical architecture that allows independent organisations to share, access, and reuse data under agreed rules, without relinquishing control over their assets. Instead of pooling everything into a central repository, each participant maintains stewardship of its own datasets while aligning with interoperability profiles, identifier schemes, and contractual safeguards. This arrangement lowers duplication costs, strengthens legal certainty, and builds trust among actors with diverse incentives but overlapping needs. In practice, data sharing spaces operate through a common semantic and governance layer — for example, shared identifier systems, metadata crosswalks, and access policies — while leaving local databases and workflows intact.

The European Commission’s communication *A European Strategy for Data* defines *Common European Data Spaces* as **federated infrastructures**, designed to enable secure and accountable data flows across Member States and sectors, underpinned by European laws, standards, and governance frameworks (European Commission 2020). Industry and research communities reinforce this interpretation. The BDVA/DAIRO discussion paper specifies that data spaces are “*federated data ecosystems within a certain application domain and based on shared policies and rules*” (BDVA/DAIRO 2023, 18), while the Federation Working Group stresses that federation is the key enabler of cross-sector interoperability in Europe (BDVA/DAIRO Federation Working Group 2023, 4). This convergence of policy and practice underscores why federation — not centralisation — is the preferred model for future observatories.

For the music ecosystem, a data sharing space provides three immediate benefits. First, it reduces reconciliation costs by promoting “*capture once, reuse many*” practices: identifiers and metadata entered in one system can propagate to others. Second, it safeguards subsidiarity by ensuring that collective management organisations, archives, and platforms remain responsible for their own registries. Third, it enhances observability: policymakers, creators, and researchers can monitor flows across the sector without requiring full access to proprietary data.

The *Feasibility Study for a European Music Observatory* already recognised that fragmentation and duplication are structural features of the music sector (Commission et al. 2020, p9). A federated data sharing space addresses this by acknowledging decentralisation as a fact and turning it into an advantage. Instead of aspiring to a monolithic central repository,

the Open Music Observatory is conceived as a convening, conformance, and observability layer that enables decentralised contributions to function as a coherent whole.

As we mentioned in the problem discussion, our Slovak pilot shows that voluntary workarounds for attribution under GDPR are possible, but they do not scale.

Partial solutions exist. The **Data Space Support Centre (DSSC)** led by KU Leuven, in collaboration with GAIA-X and BDVA, has developed blueprints and building blocks to scale data sharing under current legal frameworks. These mechanisms can help stabilise the “tinkering” approach and provide a technical foundation for federated governance.

But the real solution cannot come from technical fixes alone. The **European Parliament’s resolution on music streaming** clearly calls for systematic identifier usage and stronger attribution. Meeting this goal requires action from the Commission: either by clarifying GDPR for attribution data, requesting guidance from DPAs, or ultimately pursuing a legislative or jurisprudential solution. Only then can data spaces for music scale legally and sustainably across Europe.

3.2.11 Wikibase as a Cultural Metadata Backbone

Wikibase, the software underpinning *Wikidata*, has become a proven solution for collaborative metadata management. Originally developed by Wikimedia Deutschland, it is now widely used in cultural heritage contexts, including libraries, museums, and music-related projects. Studies highlight that Wikibase provides both a **flexible but structured data model** and an **accessible user interface**, making it suitable for non-expert users to maintain knowledge graphs and produce CIDOC-CRM compliant RDF (Kesäniemi, Koho, and Hyvönen 2022, p542).

Comparative analyses show that Wikidata, powered by Wikibase, is already functioning as a **complementary tool for authority control**, alongside traditional infrastructures like VIAF, enabling interoperability across bibliographic and heritage domains (Bianchini, Bargioni, and Pellizzari di San Girolamo 2021, p210; Sardo and Bianchini 2022, p297). Other work had already emphasised its role six years ago as an emerging authority hub, where Wikidata identifiers can act as universal references, lowering barriers to linked open data integration (Veen 2019, p75).

Within the European Union itself, Wikibase has been adopted as the infrastructure behind the **EU Knowledge Graph**, used by the European Commission to integrate projects funded under EU programmes and make them accessible to citizens (Diefenbach, Wilde, and Alipio 2021). The *SEMIC 2020 initiative* also highlighted Wikibase and Wikidata as sustainable data services for public-sector interoperability (2020 2020), while recent work demonstrates their potential as research data management services in the cultural heritage field (Rossenova, Duchesne, and Blümel 2022).

Within Europe’s public-sector linked-data ecosystem, **SEMIC** has identified *Wikidata* and Wikibase as the primary tools to enable collaborative semantic modeling and ensure interoperability among public data spaces (SEMIC Support Centre 2023). At the Belgian national level, the *MetaBelgica* project—a collaboration among federal scientific heritage

institutions—explicitly uses Wikibase to host FAIR-compliant entity data and to support long-term, multi-institutional curation workflows (Stallmann et al. 2023, p1). This demonstrates Wikibase’s viability as an infrastructure component for federated metadata systems in Europe, and makes its adoption in the Flemish OSLO-linked framework (e.g., for *cultuurparticipatie* via the Uitwisselingsplatform) both plausible and justified.

At the sectoral level, the Flemish performing arts field has already piloted such an approach. Since 2017, PACKED (now meemoo) and *Kunstenpunt* (Flanders Arts Institute) have experimented with publishing performing arts data on Wikimedia platforms. In 2021, Kunstenpunt uploaded production data from its database (since 1993) into Wikidata using Wikibase as the curation and linking environment. As Magnus and Van D’huynslager explain: “*We wanted to test how data becomes richer from the outset by publishing it as linked open data on Wikidata, and to see how much enrichment would arise later through the community*” (Magnus and Van D’huynslager 2021). The upload produced immediate enrichment: “*About 1700 of the uploaded venues have specific geo-coordinates on Wikidata ... more than 2000 persons and more than 750 venues ... now have a photo on Wikidata*” (Magnus and Van D’huynslager 2021).

Together, these examples show that Wikibase/Wikidata is not only technically mature, but already embedded in Flemish and Belgian cultural infrastructures: from EU-level interoperability frameworks to federal scientific heritage institutions and practical performing arts pilots. This makes it a pragmatic reference point for designing federated, decentralised metadata infrastructures in the music sector.

This makes Wikibase/Wikidata a pragmatic starting point for federated metadata infrastructures: institutions can host their own Wikibase instances while linking into the global Wikidata ecosystem, ensuring decentralised control with global interoperability.

3.3 Policy Proposals

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

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).

Public-Private Partnership for Trustworthy Music Data

We create data (sharing) spaces that not only follow the models of the [European Interoperability Framework](#) and [EOSC](#) but extend to interoperability with private partners

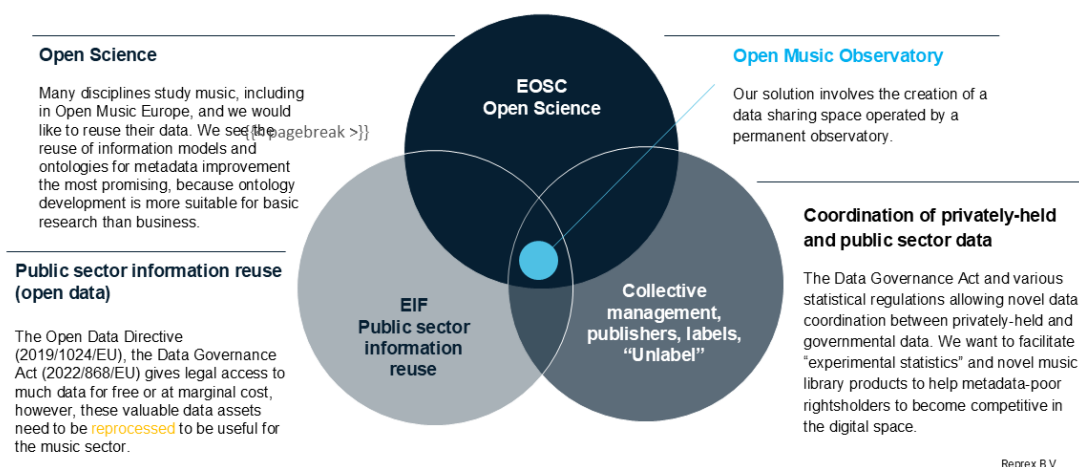


Figure 3.2: The Open Music Observatory sits where open science, public sector information reuse, and music industry workflows overlap. By aligning with the European Interoperability Framework, it creates a shared space where libraries, rights managers, publishers, and researchers can collaborate. This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

3.3.1 Federated Infrastructure as a Cost and Governance Solution

As introduced in Section 1.2.2, a **data sharing space** allows organisations to exchange and reuse data on an "*as-needed*" or "*as-permitted*" basis, while keeping full control of their own assets. This avoids the impossible task of forcing everyone into a single metadata schema or legal agreement in advance. For music — where rights, identifiers, and content are dispersed across hundreds of micro-actors and institutions — such federation is the only realistic option.

The challenge is that music also stress-tests the model. Attribution is entangled with privacy law, rights data are inconsistently standardised, and most enterprises are too small to build complex compliance systems. In this sense, music is both a "problem case" and a laboratory: if federation can work here, it can work anywhere.

The lesson is clear: the **European Music Observatory cannot be a central database**. It must be designed as a federated data sharing space — a convening and observability layer that enables decentralised contributions to function as a coherent whole. Only in this

form can it reduce duplication, lower per-unit costs for small actors, ensure attribution, and provide the governance substrate needed for trustworthy AI in the music ecosystem.⁶

3.3.2 Make Open Data Truly Accessible

The gaps of the *Open Data Directive* are not unique to the music sector. Across Europe, datasets are frequently released in heterogeneous formats, with incomplete metadata, and little harmonisation between Member States. A well-designed **data sharing space** addresses these weaknesses by providing the semantic, technical, and governance building blocks that turn legal openness into practical interoperability.

The *Open Music Observatory* can take on this role in three complementary ways. First, it should provide *reference blueprints* and mappings to existing standards, ensuring that high-value datasets are consistently published in machine-readable formats rather than ad-hoc encodings (Noardo et al. 2024). Second, it should integrate *semantic vocabularies and metadata profiles*, reducing interpretation friction across Member States and sectors (Atzori et al. 2023). Third, it should supply *shared transformation pipelines and validation tooling*, so that cleaning and integration costs are not borne by every user independently but distributed across the ecosystem (Klímek et al. 2023).

Equally important, data spaces embed governance rules and certification processes into their architecture: conformance tests, trust frameworks, and legal interoperability mechanisms ensure that data use is both technically reliable and legally safe (Terzis 2024). In this way, data sharing spaces do not simply *publish* data but actively lower transaction costs, prevent fragmentation, and ensure compliance across borders.

For the music sector, a data sharing space anchored in the *Open Music Observatory* would guarantee that datasets — whether from CMOs, platforms, or cultural institutions — are published with agreed profiles, identifiers, and governance safeguards. This creates an infrastructure where decentralised actors can interoperate at scale without losing autonomy.⁷

⁶**Policy and practice foundations for data spaces in music.**

- **EU baseline.** The *European Strategy for Data* (2020) defines Common European Data Spaces as federated ecosystems; the *Data Governance Act* (2022) and *Data Act* (2023) provide the governance tools and access rights.

- **Technical frameworks.** The *Data Spaces Support Centre (DSSC)* in its *Blueprint v2.0* sets out participants, services, rulebooks, and cross-data space interoperability profiles (Data Spaces Support Centre 2025b, 2025a).

- **Governance framing.** The **Big Data Value Association (BDVA)** highlights trusted frameworks, lifecycle integration, and governance pillars with trust at the centre (Big Data Value Association 2019; Loutas et al. 2022; BDVA/DAIRO 2023). The **Federation Working Group** identifies federation, not centralisation, as the only sustainable model (BDVA/DAIRO Federation Working Group 2023).

- **Music mandate.** The *Feasibility Study for a European Music Observatory* and the *Music Ecosystem 2025* study both document fragmentation and duplication in music markets, calling for federated solutions. The **European Parliament’s Resolution on the music streaming market** reinforces this by urging systematic use of ISNI, IPI, IPN alongside ISWC/ISRC, and stronger attribution (Commission et al. 2020; Music Moves Europe 2024; European Parliament 2024).

⁷Comparative research shows that without prescriptive standards, high-value datasets are released in divergent formats and with incomplete metadata, undermining reuse (Klímek et al. 2023, 184). Reference models and blueprint components can mitigate this by aligning producers around canonical

Yet, data sharing spaces are not a silver bullet: they can prevent the creation of unusable data in the future, but legacy releases require a complementary strategy — namely the application of **curative AI**. This is the focus of the next section (Chapter 4).

3.3.3 Connect to Europeana & ECCCH

We are not only pointing out gaps — we are also building bridges. In 2024, we formally requested to become a **Europeana data provider**, with the intention of delivering a dataset larger than Europeana’s current music holdings. Our pilot shows that even the existing EDM can be used to expose a much richer set of music for listening and viewing, though some modernisation will be inevitable. This would demonstrate that music can be integrated into Europeana without waiting for a full re-engineering of its model.

We also submitted **two cascading grant proposals** to the ECHOES consortium, seeking to demonstrate how music metadata can flow across both public and private infrastructures. One proposal — based on Livonian folk music from the Latvian Archives of Folklore — was unsuccessful, but another, linked to our **Finno-Ugric Data Sharing Space** (a conceptual pathway of the Open Music Observatory), may still enable insertion of music data into ECCCH. Regardless of funding outcomes, we see value in dialogue between the ECHOES/ECCCH data model and the Open Music Observatory’s model.

Our choice of **Wikibase/Wikidata** as the technical backbone is deliberate. Wikidata has proven its ability to connect public and private infrastructures at global scale, and it is formally recommended by the EU Publications Office as a semantic authority. Wikibase offers a pragmatic way to map elements of public and private ontologies, enabling cooperation without forcing either side to abandon its systems. With Google among its early supporters, Wikidata is today the largest open graph in the world — an infrastructure already woven into the backbone of the web.

Our approach is one of **ontological pragmatism**: using tested, open technologies to bridge gaps, while focusing on the harder legal and organisational aspects of interoperability in a public-private setting. This makes the Open Music Observatory not just another cultural repository, but a realistic path towards aligning music’s real data flows with Europe’s wider cultural data infrastructures.

3.3.4 Alignment with the European Open Science Cloud

3.3.4.1 Other

Bridge cultural clouds and market workflows via a federated Music Data Sharing Space.

schemas (Noardo et al. 2024). Data spaces also address semantic and legal interoperability together — shared vocabularies, certification processes, and trust frameworks reduce uncertainty and lower entry costs for SMEs and public institutions alike (Atzori et al. 2023; Terzis 2024).

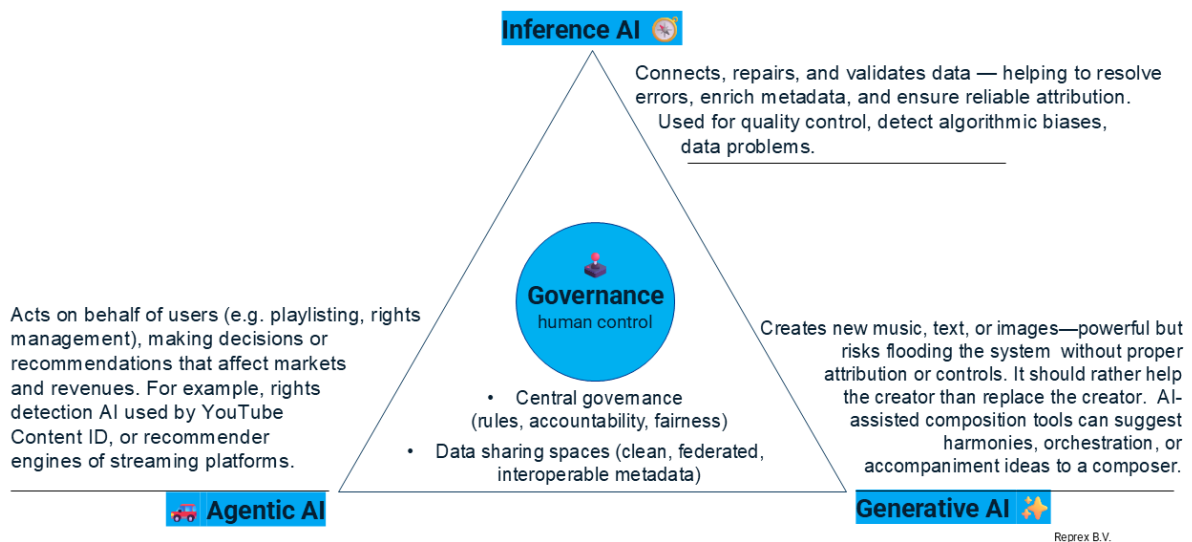
Position the Open Music Observatory as the *convening + conformance + observability* layer that connects ECCCH/Europeana and GLAM authority files with industry pipelines. Concretely: (1) **capture once, reuse many** across creation→registration→distribution→preservation; (2) require **minimal profiles** that smaller actors can actually implement; (3) prioritise **identifier crosswalks** (ISRC ISWC ISNI VIAF/Wikidata) and change-propagation; (4) use **Wikibase/Wikidata** as a low-friction backbone where appropriate; (5) govern with EIF/FAIR-aligned rules, auditability, and PPP participation so rights-holders and memory institutions keep stewardship while interoperating.

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

4 AI that Works for Music, Not Against It

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).

Trustworthy AI That Works For Music



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

European policy provides guidance for this balancing act. The *Ethics Guidelines for Trustworthy AI* underline that AI must be lawful, ethical, and robust throughout its lifecycle

(Commission, Directorate-General for Communications Networks, and Technology 2019). The *Getting the Future Right* report by the Fundamental Rights Agency stresses the need to align AI with fundamental rights, especially where vulnerable groups and cultural participation are concerned (European Union Agency for Fundamental Rights 2020). Most recently, the **AI Act** enshrines a risk-based regulatory framework, defining obligations for providers and deployers of AI systems while reaffirming the principles of subsidiarity and proportionality in EU digital policy (European Parliament and Council 2024).

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

i Case Study: Lessons from the Spotify API

Our first encounter with AI governance in music came during the *Listen Local* feasibility study (2020). By experimenting with the Spotify API, we 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.

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.

Executive summary

Caution

This is a pre-releasing editing reminder and placeholder for comments

Scope: Top layer: applied services that ride on the data space (curative, agentic, and generative uses); shared AI capacity so SMEs/NGOs aren't forced to hire scarce talent. - Create *shared AI services* (not bespoke in-house teams) for entity recognition, duplicate detection, cross-catalogue matching, multilingual enrichment, and compliance monitoring.

- Educate stakeholders using a simple three-part model of AI (agentic/generative/curative) tied to concrete tasks and guardrails.
- Reduce operational and capital costs by embedding AI into metadata, claims, and documentation workflows.
- Support creators by embedding metadata from the point of creation, extending the Parliament's "metadata from birth" principle.

4.1 Discussion

Pre-release reminder and editing space

- – EP resolution on AI-generated flooding and discoverability/remuneration risks (European Parliament 2024); ecosystem report on AI as disruption *and* opportunity.
- Practical gaps: SMEs/CMOs lack AI engineers; asymmetry with platform capabilities.

The following subsections highlight lessons from international practice and concrete cases where AI can reduce costs, improve data quality, and create more equitable access to metadata services.

4.1.1 Structural Problems for Music Businesses to Apply AI

1. AI benefits are unevenly distributed.

Music businesses operate in value chains where platforms and large intermediaries already use agentic, generative, and even inference AI. 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 rightsholders 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 concern¹.

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

¹Surveys 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 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).

2 people in full-time positions, this AI and data talent shortage cannot be solved on an individual business level².

4.1.2 European Regulation That Misses the Point

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

Music was classified as “low-risk” on the assumption that nobody is harmed by being offered a bad song. This framing ignores how **agentic AI governs the marketplace itself**. If recommendation systems consistently fail to show music 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.

4.1.3 Policy Issues at the Intersection of AI, Copyright, and GDPR

AI in music does not operate in a legal vacuum. It interacts with existing European law on intellectual property, author’s rights, moral rights, and data protection. In practice, this creates tensions and unresolved policy gaps that directly undermine cultural policy goals. This governance problem builds directly on the interoperability failures described in Section 3.2.3.

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:

²According 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.

- 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.

2. Local content protection gaps.

In broadcasting, local content quotas were established in line with WTO rules to safeguard cultural diversity (e.g. Slovak radios playing at least 20% Slovak music). Similar obligations now exist in audiovisual streaming. But in music streaming there are no binding European diversity or local content rules. This creates two problems:

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

3. Voluntary compliance is impractical.

Current practice relies on voluntary measures by radio editors, festival curators, or platform users to include local or diverse content. But without accessible data, this becomes unworkable. Our own experiments with GDPR balancing tests and opt-ins show the futility of this approach. Fewer than 1% of artists responded to requests to consent to attribution data — even prominent Slovak artists, puzzled at being asked to consent to rights they already legally hold.

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

4.1.4 AI Design Without Awareness of Limits

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

- **Agentic AI** systems (recommenders, playlist builders, rights-management bots) operate without recognising the biases or incompleteness of the datasets they learn from. Because European legislation deems the agentic use of AI in music “low risk”, currently there are no real expectations to address this problem.
- **Generative AI** produces synthetic material without constraints, and its training processes seldom acknowledge gaps or skew in the underlying data. This problem touches upon various issues that we discussed earlier in this paper: author’s rights

and performer rights are assigned to natural persons (and their heirs), as well as sometimes producer’s rights, too. GDPR currently appears to be at conflict with both designing safer AI system, and generally to provide proper attribution without legal risk to creators of protected work. We could technically guardrail generative AI to not produce plagiarism, but not without giving it access to whose work is forbidden.

- Even **Inference AI**, which is supposed to reason from formal rules, can miss the point: ontological relativity and incompleteness are structural limits and not optional refinements. We discussed in the Chapter 2, and just as well as database designers must be aware that that no ontology or schema is ever complete, AI engineers must realise that they train algorithms that cannot capture all perspectives. Without this awareness, AI will silently reproduce exclusions — whether of women, minorities, or smaller repertoires — while appearing “intelligent.”

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

4.1.5 Unfreezing Frozen Assets

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

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

4.1.6 AI Support for Investment Into New Repertoire Assets

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

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

4.2 Policy Proposals: Aligning AI with Governance and Value Creation

Generative AI and related technologies 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 therefore needed on three levels: **EU, industry, and organisational**.

Our policy brief proposal is focused on the data needs and metadata improvements of the music industry, i.e., from labels via distributors, promoters, talent managers, and not on the creative process of making or generating new music.

4.2.1 EU-Level Policy: Compass and Guardrails

- **Embed cultural sector in the EU AI Act & Data Spaces**
Ensure that the implementation of the EU AI Act and European Data Spaces addresses the needs of music, cultural heritage, and creative industries.
- **Subsidise “shared AI utilities”**
Fund pan-European AI services for identifier reconciliation, metadata repair, and plagiarism/fraud detection. These should be open to CMOs, archives, publishers, and SMEs.
- **Adopt “metadata from birth” principles**
Require EU-funded projects and subsidised AI tools to embed ISNI/ISWC/ISRC identifiers at the point of creation, enabling downstream interoperability.
- **Tax incentives for onboarding frozen assets**
Provide financial support for digitising and enriching under-documented back catalogues, unlocking dormant cultural and economic value.

4.2.1.1 Towards a Real Solution

For AI systems, voluntary GDPR workarounds are not enough. As already discussed in Section 2.1.5, attribution is indispensable but legally uncertain under GDPR. If attribution data cannot be processed with legal certainty, then recommender systems cannot be tested for fairness, quotas cannot be monitored, and generative AI cannot be constrained against plagiarism.

The real solution must come from EU-level action:

- The European Commission could initiate a technical amendment to GDPR clarifying how attribution and identifier data in the cultural and creative sectors should be treated.
- The Commission or Member States could bring a test case to the Court of Justice of the EU to resolve the contradiction between copyright attribution and GDPR.

- Industry bodies could formally request guidance from Data Protection Authorities, giving operators a clear baseline for processing attribution data.

Without one of these interventions, AI governance in music will remain fundamentally misaligned with Europe’s own copyright and cultural diversity objectives.

4.2.2 Industry-Level Policy: Standards and Collaboration

- **Codes of conduct for AI in music**
Following the GDPR “codes of conduct” model, CISAC, IFPI, IAML, IAMIC and others should develop voluntary but enforceable standards for trustworthy AI in music documentation, licensing, and claims.
- **Identifier crosswalks**
Extend interoperability across existing identifier systems (ISRC ISWC ISNI VIAF), ensuring AI tools can reconcile and enrich catalogues across domains.
- **Federated AI services**
Build shared AI services at the sector level (claims automation, multilingual reconciliation, watchlists), to avoid duplication and ensure critical mass.
- **Sectoral training and reskilling**
Coordinate industry-wide training initiatives to address the talent gap in data stewardship and AI governance within the creative economy.

4.2.2.1 AI for Working Capital Optimisation

For many rightsholders, especially smaller ones, cash flow is delayed because of slow or inefficient links between their systems and platforms, or between authors, publishers, and collective management organisations. Uploads, claims, and distributions are often handled through manual, fragmented processes that increase transaction costs and tie up working capital.

AI can support working capital optimisation by enabling cheaper API-based liaisons with platforms and CMOs, automating the management of routine claims, and improving the matching of works and recordings. This reduces delays and transaction friction, allowing creators and small organisations to access revenue more quickly.

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

4.2.4 Embedding AI in Creative Workflows Instead of Replacing Creatives

We propose to embed AI services directly into music creation and distribution tools, so that metadata is generated and validated as part of the creative process. These services should respect human authorship and repertoires, while reducing costs for creators by automating documentation and identifier allocation. Public support could incentivise software vendors and open-source projects to integrate such features, ensuring that SMEs, independent artists, and non-commercial creators also benefit. This would extend the Parliament’s “metadata from birth” principle into a broader vision of *creation with embedded metadata*, lowering barriers for creators and strengthening the metadata foundations of the European music ecosystem.

4.2.4.1 Avoid redundancy

- **Adopt “capture once, reuse many” workflows**

Ensure metadata captured at creation (song, recording, performance, archive item) can flow across internal and external systems.

We propose to establish AI-enabled metadata services within European data sharing spaces, governed by shared ontologies and ethical guardrails. These services should automate routine documentation tasks (identifier reconciliation, crosswalks, enrichment from external knowledge bases) and distribute the resulting efficiencies across all actors, not just the majors. Public investment should support the creation of open, agentic AI modules that plug into observatory infrastructures and enable SMEs, non-profits, and independent creators to achieve platform-level economies of scale. This would not only unfreeze frozen assets but systematically reduce the OPEX of music metadata management across Europe.

4.2.4.2 CAPEX

- **Invest in knowledge capital, not just IT CAPEX**

Prioritise ontologies, controlled vocabularies, and multilingual metadata enrichment that can feed into AI tools.

We propose a funding and governance model that prioritises investment in *knowledge capital*—shared ontologies, pattern libraries, and AI modules—rather than continuous replacement of IT assets. By deploying AI to extend the functionality of legacy systems, CMOs and cultural institutions can redirect scarce capital away from hardware/software churn and into data quality, interoperability, and human expertise. This reallocation of CAPEX would create more durable value: instead of buying new IT every five years, institutions would invest in collective intelligence that strengthens the entire music ecosystem.

4.2.4.3 Use shared AI utilities

Subscribe to industry-level services for fraud detection, multilingual enrichment, and claims reconciliation, rather than building isolated tools in-house.

We propose to develop AI-assisted services for claims management and API integration across the European music ecosystem. By lowering the cost of system-to-system communication, these tools would give European SMEs and rightsholders the same efficiencies that global platforms already enjoy. The US **Mechanical Licensing Collective (MLC)** has demonstrated how coordinated metadata and streamlined claiming processes can accelerate payouts and reduce disputes (Varghese 2024). Europe should build on this model by investing in shared, AI-enabled liaison services that optimise working capital for creators, CMOs, and independent labels alike.

4.2.4.4 Develop internal AI governance

Appoint a Chief Data Officer (or equivalent) responsible for AI adoption and metadata governance. Even SMEs and CMOs can designate an “AI steward” role to avoid shadow AI dependence.

4.2.5 Closing Note

AI will only create sustainable value for the creative industries when governance, interoperability, and human capital are aligned. The EU sets the **compass and guardrails**, industry bodies provide **shared standards and utilities**, and organisations themselves must adopt **governance and playbooks**. Without this three-level alignment, AI risks becoming the digital equivalent of *lemons on a tree* — showy but unfit for purpose.

4.2.6 Curative AI as a Remediation Solution

While data sharing spaces establish the rules and structures for *new* data flows, they do not address the **legacy backlog** of poorly formatted or incomplete open data already in circulation. Here, **curative AI** provides a complementary solution. For the music ecosystem — and for public-sector bodies obliged under the *Open Data Directive* to release high-value datasets — such AI applications should be prioritised.

AI-assisted services can automatically detect duplicates, infer missing identifiers, and reconcile heterogeneous formats. They can enrich metadata with multilingual descriptions, cross-link authority files (e.g. ISNI, VIAF, ORCID), and repair legacy exports from outdated IT systems. In effect, they transform datasets that are legally open but practically unusable into resources that can circulate across the ecosystem.

In practical terms, curative AI reduces the hidden costs of “free” data. Even when datasets are available at zero monetary price, acquisition, transformation, and integration impose

substantial burdens (Schnurr 2021). By pooling remediation services at the level of the Observatory — rather than leaving each SME or archive to hire scarce AI engineers — the sector can benefit from economies of scale and achieve more consistent outcomes.

Thus, governance and remediation are two sides of the same coin:

- **Data sharing spaces** ensure that *new* data is created and exchanged 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.³

4.2.7 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 be directed to subsidise initial onboarding costs, create standardised transfer pipelines, and incentivise low-friction reuse of metadata across systems. This would extend the benefits of metadata interoperability to cultural assets currently excluded from the digital market.

4.2.8 AI-Powered Claims and API Integration

³Empirical studies underline that legal openness alone rarely produces machine-actionable reuse. Without prescriptive standards, high-value datasets are published in divergent formats and with inconsistent metadata (Klímek et al. 2023, 184). Reference models and standards can mitigate this going forward (Noardo et al. 2024), but existing datasets often require intensive remediation. Curative AI provides this remediation by automating anomaly detection, schema transformations, and semantic enrichment. For example, Ni et al. demonstrate that machine-learning frameworks can flag outliers, propose corrected values, and normalise structures at scale, significantly reducing manual repair costs (Ni et al. 2023, 3). In cultural and company register contexts, pilots such as STIRData have shown how AI-driven pipelines transform heterogeneous exports into interoperable, HVD-compliant formats (Klímek et al. 2023, 185). These examples illustrate why curative AI, deployed as a shared Observatory service, is indispensable for turning open data from a *legal right* into a *practical resource*.

5 What Europe Should Do Next for Music Data & AI

Europe’s music ecosystem is under pressure. Streaming pays in micro-royalties, metadata mistakes cost real money, and AI threatens to overwhelm platforms with untracked content. But solutions are within reach. This Green Paper sets out a path forward, built on three pillars: better metadata, shared data spaces, and AI that works for everyone. (See Chapter 1 for the background and policy context.)

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 issue — it’s the foundation for fair royalties and cultural visibility.

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.

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

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

Taken together, these steps close the gap between the *right of reuse* granted by the Open Data Directive and the *means of reuse* that the music industry actually needs. Europe should:

- Support metadata capture and cross-domain identifiers.

- Invest in federated data sharing spaces like the Open Music Observatory.
- Provide pooled AI services that SMEs, CMOs, and archives can all use.

This is how we make Europe’s music ecosystem more fair, efficient, and future-proof — for creators, for industry, and for audiences alike.

Sources & Further Reading

- 2020, SEMIC. 2020. “Providing Sustainable Data Services Through Wikibase and Wikidata.” October 15. https://joinup.ec.europa.eu/sites/default/files/custom-page/attachment/2020-11/Parallel-track-4_B-Fischer_J-Thill_A-Angeli%20final%20ppt.pdf.
- Antal, Daniel. 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. *Trustworthy AI and Data-Sharing Spaces for the Slovak Music Centre. Poster Presentation at the IAMIC Conference 2024 on November 21, 2024, at Music Austria, Vienna*. Open Music Observatory. <https://doi.org/10.5281/zenodo.16540605>.
- . 2025a. *SKCMDb. Interoperability of Music Libraries and Archives with Public and Private Music Services. Presentation at the IAML 2025 Conference in Salzburg, Austria Held on the 7th of July 2025*. Open Music Observatory. <https://doi.org/10.5281/zenodo.16634558>.
- . 2025b. *Slovak Music Data Sharing Space. Poster Presentation at the IAML 2025 Conference in Salzburg, Austria Held on the 7th of July 2025*. Open Music Observatory. <https://doi.org/10.5281/zenodo.15814286>.
- Antal, Daniel, and Anna Márta Mester. 2025. *Open Music Registers*. <https://doi.org/10.5281/zenodo.14767717>.
- Atzori, Maurizio, Angelo Ciarabella, Claudia Diamantini, Beniamino Di Martino, Salvatore Distefano, Tullio Facchinetti, Fabrizio Montecchiani, Antonino Nocera, Giancarlo Ruffo, and Roberto Trasarti. 2023. “Dataspaces: Concepts, Architectures and Initia-

- tives.” *ITADATA 2023: The 2nd Italian Conference on Big Data and Data Science* (Naples, Italy), 2023.
- 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.
- Bianchini, Carlo, Stefano Bargioni, and Camillo Carlo Pellizzari di San Girolamo. 2021. “Beyond VIAF Wikidata as a Complementary Tool for Authority Control in Libraries.” *Information Technology and Libraries* 40 (2). <https://doi.org/10.6017/ital.v40i2.12959>.
- Big Data Value Association. 2019. *Towards a European Data Sharing Space: Enabling Data Exchange and Unlocking AI Potential*. BDVA. <https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V1.pdf>.
- Blomqvist, Eva, Karl Hammar, and Valentina Presutti. 2016. “Engineering Ontologies with Patterns – the eXtreme Design Methodology.” In *Ontology Engineering with Ontology Design Patterns*, 23–50. IOS Press. <https://doi.org/10.3233/978-1-61499-676-7-23>.
- Bloomberg, and Boston Consulting Group. 2025. *CEO Radar: AI and Machine Learning References in European Earnings Calls*. Bloomberg L.P.; Boston Consulting Group. <https://www.bcg.com/publications/2025/ai-machine-learning-mentions-earnings-calls>.
- Buttow, Carolina V., and Albert Meijer. 2024. “Public Sector Data Openness in the Crafting of the Data-Driven Society: The Co-Constitutive Role of Regulation and Innovation.” *Swiss Yearbook of Administrative Sciences* 15 (1): 12–27. https://serval.unil.ch/resource/serval:BIB_118FD2F827C6.P002/REF.
- Carrara, Wendy, Wae San Chan, Sandra Fisher, and Eva van Steenberg. 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>.
- Christen, Kimberly. 2011. “Opening Archives: Respectful Repatriation.” *The American Archivist* 74 (1): 185–210. <https://doi.org/10.17723/aarc.74.1.4233nv6nv6428521>.
- 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, 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>.
- 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.
- Daniel, Antal, Kata Gábor, Ieva Pigozne, and Ieva Vivere. 2025. *A Finno-Ugric Data Sharing Space*. Finno-Ugric Dataspace. <https://doi.org/10.5281/zenodo.15382553>.
- Data Spaces Support Centre. 2025a. “Data Spaces Blueprint V2.0 — Cross-Data Space Interoperability Considerations in Data Space Design and Operation.” <https://dssc.eu/space/BVE2/1071252241/Cross-data+space+interoperability+considerations+in+data+space+design+and+operation>.
- . 2025b. “Data Spaces Blueprint V2.0 — Introduction: Key Concepts of Data Spaces.” <https://dssc.eu/space/BVE2/1071251613/Introduction+-+Key+Concepts+of+Data+Spaces>.

- DataCite. 2021. *DataCite Metadata Schema: Mapping to Dublin Core*. DataCite. https://schema.datacite.org/meta/kernel-4.4/doc/DataCite_DublinCore_Mapping.pdf.
- Diefenbach, Dennis, Max de Wilde, and Samantha Alipio. 2021. “Wikibase as an Infrastructure for Knowledge Graphs: The EU Knowledge Graph.” *ISWC 2021* (Online, France), October 2021. <https://hal.science/hal-03353225>.
- Dinler, Mesut. 2025. “Metadata and Linked Open Data in Digital Heritage for Decolonization.” In *The Palgrave Encyclopedia of Cultural Heritage and Conflict*, edited by Ihab Saloul and Britt Baillie, 1–9. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-61493-5_52-1.
- 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.
- European Commission. 2020. *A European Strategy for Data*. European Commission. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020DC0066>.
- . 2021. *Commission Recommendation (EU) 2021/1970 of 10 November 2021 on a Common European Data Space for Cultural Heritage*. <http://data.europa.eu/eli/reco/2021/1970/oj/eng>.
- 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. 2014. “Directive 2014/26/EU of the European Parliament and of the Council of 26 February 2014 on Collective Management of Copyright and Related Rights and Multi-Territorial Licensing of Rights in Musical Works for Online Use in the Internal Market.” In *Official Journal of the European Union*, L 84, 72–98. <https://eur-lex.europa.eu/eli/dir/2014/26/oj>.
- . 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).” In *Official Journal of the European Union*, L 119, 1–88. <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).” In *Official Journal of the European Union*, L 152, 1–44. <https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng>.
- . 2024. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>.
- European Union Agency for Fundamental Rights. 2020. *Getting the Future Right. Artificial Intelligence and Fundamental Rights*. Luxembourg: Publications Office of the European Union. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf.
- Europeana. 2017. *Definition of the Europeana Data Model V5.2.8*. Europeana. https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf.
- EuroSDR. 2021. *Sustainable Open Data Business Models for NMCAs. EuroSDR Official Survey Report*. <https://lirias.kuleuven.be/retrieve/696771>.
- Eurostat. 2023. *Culture Statistics — 2023 Edition*. Publications Office of the European Union. <https://doi.org/10.2785/537327>.
- Fagervig, Alicia. 2023. “Wikidata for Authority Control: Sharing Museum Knowledge with the World.” *Digital Humanities in the Nordic and Baltic Countries Publications* 5 (1): 222–39. <https://doi.org/10.5617/dhnbpub.10665>.
- 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>.
- 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>.
- Hyvönen, Eero, and Jouni Tuominen. 2024. “8-Star Linked Open Data Model: Extending the 5-Star Model for Better Reuse, Quality, and Trust of Data.” In *Posters, Demos, Workshops, and Tutorials of the 20th International Conference on Semantic Systems (SEMANTiCS 2024)*, edited by Daniel Garijo, Alessandra Lo Gentile, Ana Kurteva, An-

drea Mannocci, Francesco Osborne, and Sahar Vahdati, vol. 3759, 3759. CEUR Workshop Proceedings. Aachen: CEUR-WS.org. <http://hdl.handle.net/10138/586661>.

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 2108:2017 (En), Information and Documentation — International Standard Book Number (ISBN)*. 2108. Version 2016. International Organization for Standardization. <https://www.iso.org/standard/65483.html>.

———. 2017b. *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>.

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

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

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

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

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

———. 2022. *International Standard Musical Work Code (ISWC)*. ISO 15707:2022. 15707.

- Version 2022. International Organization for Standardization. <https://www.iso.org/standard/83125.html>.
- . 2023a. *ISO/IEC 11179-1:2023(en), Information Technology — Metadata Registries (MDR) — Part 1: Framework*. London:United Kingdom: International Organization for Standardization. <https://www.iso.org/obp/ui/en/#iso:std:iso-iec:11179-1:ed-4:v1:en>.
- . 2023b. *ISO/IEC 2382:2015(en), Information Technology — Vocabulary*. London:United Kingdom: International Standards Organisation. <https://www.iso.org/obp/ui/en/#iso:std:iso-iec:2382:ed-1:v2:en>.
- ISO/IEC. 2022. *Information Technology — Artificial Intelligence — Concepts and Terminology. ISO/IEC 22989:2022*. 22989. Version 2022. International Organization for Standardization. <https://www.iso.org/standard/74296.html>.
- . 2023. *Information Technology — Artificial Intelligence — Management System. ISO/IEC 42001:2023*. 42001. Version 2023. International Organization for Standardization. <https://www.iso.org/standard/81230.html>.
- Jachimczyk, Bartosz, and Katarzyna Nowak. 2024. “Exploring Open Government Data (OGD) Portals: A Usability Study Across the EU.” *arXiv Preprint*, 2024. <https://arxiv.org/abs/2406.08774>.
- Jegan, Robin, Leon Fruth, Tobias Gradl, and Andreas Henrich. 2023. “Integrating Access to Authority Data for Improved Interoperability of Research Data in the Digital Humanities.” *Datenbanksysteme Für Business, Technologie Und Web (BTW 2023)*, 2023, 829–36. <https://doi.org/10.18420/BTW2023-54>.
- Kesäniemi, Joonas, Mikko Koho, and Eero Hyvönen. 2022. “Using Wikibase for Managing Cultural Heritage Linked Open Data Based on CIDOC CRM.” In *New Trends in Database and Information Systems*, edited by Silvia Chiusano, Tania Cerquitelli, Robert Wrembel, Kjetil Nørvg, Barbara Catania, Genoveva Vargas-Solar, and Ester Zumpano, 542–49. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-15743-1_49.
- 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>.
- Klímek, Jakub, Alexandros Chortaras, Jan Mísek, Jie Yang, Sigmund Skagemo, and Vasilios Tzouvaras. 2023. “Semantic, Technical and Legal Interoperability of European Company Open Data in Practice: The STIRData Approach.” *Proceedings of the 12th International Conference on Data Science, Technology and Applications (DATA 2023)*, 2023, 183–94. <https://doi.org/10.5220/0012132100003541>.

- Loutas, Nikolaos, Vassilios Peristeras, Konstantinos Tarabanis, Matthias Jarke, et al. 2022. “Designing Data Spaces: The Ecosystem Approach to Competitive Data Sovereignty.” In *Designing Data Spaces*, edited by Matthias Jarke, Maurizio Lenzerini, et al., 3–22. Springer. https://doi.org/10.1007/978-3-030-93975-5_1.
- Magnus, Bart, and Olivier Van D’huynslager. 2021. “Podiumkunstendata Op Wikidata: De Stap Naar Echte Linked Open Data.” *Kunstenpunt / Flanders Arts Institute*, February 11. <https://www.kunsten.be/nu-in-de-kunsten/podiumkunstendata-op-wikidata-de-stap-naar-echte-linked-open-data/>.
- Mechanical Licensing Collective. 2021. *2021 Annual Report*. <https://www.themlc.com/annual-report-2021>.
- Mikš, Tomáš. 2025. *OpenMusE: Towards a Sustainable Licensing Market for AI Use of Protected Works*. April 29–30, 2025. Vilnius, Lithuania: Slovak Performing; Mechanical Rights Society (SOZA); OpenMusE Consortium; Presentation at the CISAC European Committee Meeting. https://www.openmuse.eu/wp-content/uploads/2025/05/20250427_CISAC_EC_2025_OpenMusE_final.pdf.
- Ministerstvo kultúry SR, and Open Music Europe. 2023. *Memorandum o porozumení o využití výsledkov analýz otvorených politík v kontexte slovenského kultúrneho a kreatívneho priemyslu a sektorových verejných politík v spolupráci s konzorciom pre výskum a inovácie s názvom OpenMuse*. [Memorandum of Understanding on utilizing the Open Policy Analysis results of the OpenMuse Research and Innovation Consortium in the context of Slovak cultural and creative industries and sectors’ public policies]. <https://www.crz.gov.sk/zmluva/7645338/>.
- MIT Sloan School of Management. 2025. *Project NANDA: Enterprise Generative AI Value Creation*. Massachusetts Institute of Technology. <https://mitsloan.mit.edu>.
- Music Moves Europe. 2024. *Music Ecosystem 2025: Study on the Music Ecosystem*. Publications Office of the European Union. Luxembourg: European Commission, Directorate-General for Education, Youth, Sport; Culture. <https://doi.org/10.2766/95340>.
- Nagel, Lars, and Douwe Lycklama, eds. 2021. *Design Principles for Data Spaces. Position Paper. Version 1.0*. Open DEI. <https://doi.org/10.5281/zenodo.5244997>.
- Nakos, Basil, and Lazaros Tsoulos. 2022. “Web-Based Nautical Charts Automated Compilation from Open Hydrosatial Data.” *Journal of Navigation* 75 (6): 1–19. <https://doi.org/10.1017/S0373463322000489>.
- Ni, Wei, Xiangyu Miao, Xin Zhao, Yifan Wu, and Jian Yin. 2023. “Automatic Data Repair: Are We Ready to Deploy?” *arXiv Preprint*, 2023. <http://arxiv.org/abs/2310.00711>.
- Noardo, Francesca, Rob Atkinson, Lucy Bastin, Joan Maso, Ingo Simonis, Alejandro Villar,

- Marie-Françoise Voidrot, and Piotr Zaborowski. 2024. “Standards for Data Space Building Blocks.” *Remote Sensing* 16 (20): 3824. <https://doi.org/10.3390/rs16203824>.
- OpenRefine Community. 2021. *OpenRefine Reconciliation API Standard*. <https://reconciliation-api.github.io/specs/latest/>.
- 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/hitchhikers-guide-high-value-datasets>.
- PwC. 2023. *Digital IQ 2023: Driving ROI on Digital Investments*. PricewaterhouseCoopers International Limited. <https://www.pwc.com/gx/en/industries/technology/digital-iq-survey.html>.
- . 2024. *27th Annual Global CEO Survey*. PricewaterhouseCoopers International Limited. <https://www.pwc.com/gx/en/ceo-agenda/ceosurvey/2024.html>.
- Quine, Willard Van Orman. 1968. “Ontological Relativity.” *The Journal of Philosophy* 65 (7): 185–212. <https://doi.org/10.2307/2024305>.
- Rossenova, Lozana, Paul Duchesne, and Ina Blümel. 2022. “Wikidata and Wikibase as Complementary Research Data Management Services for Cultural Heritage Data.” Paper presented The 3rd wikidata workshop, workshop for the scientific wikidata community, @ ISWC 2022,. *CEUR Workshop Proceedings*, October 24, 2022. https://serwiss.bib.hs-hannover.de/frontdoor/deliver/index/docId/2573/file/rossenova_etal2022-wikidata_research_data_mgmt.pdf.
- 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>.
- 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.
- Terzis, Panagiotis. 2024. “Interoperability and Governance in the European Health Data Space Regulation.” *Proceedings of the 19th International Conference on Internet Technology and Secured Transactions (ICITST-2024)*, 2024, 112–19. <https://doi.org/10.1109/ICITST60470.2024.00025>.
- Varghese, Jacob. 2024. “Beyond the Metadata: How Can We Solve the Black Box Royalty Mystery?” Noctil, August 14. <https://independentmusicinsider.com/editorial-articles/4820/>.
- Veen, Theo van. 2019. “Wikidata: From ‘an’ Identifier to ‘the’ Identifier.” *Information Technology and Libraries* 38 (2): 72–81. <https://doi.org/10.6017/ital.v38i2.10886>.
- 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>.