

DataMosaik1

*A mapping of danish cultural data sources and the
state of data practices in the cultural sector*

”I could get my grandmother’s old Pentium computer to run all registrations and payouts, if only the data were there, if only they were shared, if they were complete.”

– Morten Hjort, Head of Data and Distribution,
Music Producers’ Collecting Society / MPO

Colophon

Title: DataMosaik1 – a mapping of Danish cultural data sources and the state of data practices in the cultural sector

Authors: Kristian Mortensen and
Åsmund Trolle Boye Lindquist-Kverneland

Graphic design: Maria Daa

Edition: 1st version, Roskilde, January 2026.

License: This publication is available under the [CC-BY-SA 3.0 IGO](#) licence. Reproduction of the report's content is permitted with clear source attribution.

Publisher:
Kulturens Analyseinstitut
Knudsvej 44 C
4000 Roskilde, Danmark

Tel.: (+45) 6916 2669
info@kulturanalyser.dk
kulturensanalyseinstitut.dk

About Kulturens Analyseinstitut: Kulturens Analyseinstitut, or Danish Institute of Cultural Policy Analysis, is a self-governing and independent institution established by law in 2023 on the basis of a political majority in the Danish Parliament. The purpose of Kulturens Analyseinstitut is to create and disseminate knowledge about the importance of culture in society.

ISBN: 97887-85495-99-0

This mapping has been prepared in dialogue with the Danish Agency for Digital Government (Digitaliseringsstyrelsen) and with funds from NextGenerationEU.

Indhold

Introduction	6
Reading guide	8
Purpose and background	8
Case selection: domain specificity vs. generalisability	9
Terminology: key concepts used in this report	10
Executive Summary	12
Key findings	12
Overall summary	12
Main recommendations	14
From recommendations to four pilot projects	15
Methodology	17
Period and purpose	17
Data collection and analysis method	17
Cultural data sources covered by the mapping	19
Ethics and data protection	20
Scope and generalisability	20
How is the term culture delimited and used in this report?	21
Key findings, elaborated	22
Layers of the data infrastructure	22
Challenges in the data layers	23
Gap analysis of key findings	25

Gap 1 - Double entry and redundant data work	25
Gap 2 - Identifier breaks in the value chain	26
Gap 3 - Lack of shared minimum metadata (MVM)	26
Gap 4 - Exchange via files rather than APIs	26
Gap 5 - Governance and weak feedback loops	27
Gap 6 - Missing metadata preservation and cultural heritage preservation	27
Gap 7 - Missing bridges between closed and open identifiers.	27
Interdependencies and consequences of the gaps.	28
Summary of the gap analysis	28
International confirmation of the gaps	29
The gap analysis linked to low data literacy in the cultural sector	29
Recommendations on how to strengthen data literacy in the cultural sector.	30
 Conclusion	 32
 Possible solutions, or paths of action.	 34
The paths of action in the form of four pilot projects	34
 Recommendations for role allocation	 37
 Way forward	 39
 Appendices	 41
Appendix 1: Cultural data sources for datavejviser.dk	42
Appendix 2: Primary data types and structures in the music industry	43
Appendix 3: Data identifiers and data bridges in the music industry	44
Appendix 4: Extended glossary	46
Appendix 5: Interview guide (semi-structured)	49

Introduction

Culture, or art, is ubiquitous in society. As cultural creators and consumers we have personal preferences. Some love metal, haiku poetry or local folk songs; others love comics, choral singing or modern dance; and others again are into ceramics, K-pop or anime. The list is practically endless.

Cultural creation happens for everyone in so many shapes. We have a drive or a desire in our expression or preference. With the creation of art there are several practical aspects - for example buy oil paints, have access to a music studio, improve one's skills, and so on. There is also a fair amount of administration connected to cultural creation, including a substantial data-oriented component.

Cultural creation is, by its nature, often data intensive. Every song, book, performance or exhibition needs data about it, which is what we in this report refer to as *metadata*. Information about authorship, rights, time, place and context must be registered, maintained and shared between actors and audiences. When these metadata processes work, artists' rights are safeguarded, cultural heritage is preserved, and the public's access to culture is ensured. When they do not work, artists lose income¹, institutions lose oversight, and significant parts of our cultural heritage are not preserved correctly.

For example, we have been in dialogue with Fie and her husband. They are both musicians. They earn their living from it and, as part of the growing group of artists, they manage their own business outside the major companies. Spotify's own figures from its Loud & Clear 2024 report², show that approximately half of total payouts go to independent artists and record labels, including a growing share of self-managed artists. It is important to clarify, however, that "independent" in this context does not necessarily mean "self-managed", but more broadly covers artists and companies that are not affiliated with the major record labels. And Spotify does not pay artists directly, but via rights managers and other distribution links.

The couple explain that they administer all their rights themselves. This means that they must handle up to seven different systems, including registering copyright (KODA), recording rights (Gramex), digital distribution, and royalties collection solutions. They often spend 10–15 hours per week on administration alone. A central challenge is that they cannot compare their metadata (song titles, credits, royalty splits and play counts across systems), because each system has its own data structure and cannot be linked.

¹ gaffa.dk/nyheder-fra-gaffa/2025/oktober/paul-mccartney-elton-john-og-thom-yorke-rejser-kritik-mod-musikrettighedskollektiv

² loudandclear.byspotify.com/#takeaway-9

With the democratisation of digital tools - not only to produce artistic works, but also for their distribution - many barriers for artists have been removed. This means that, broadly speaking, cultural releases have exploded in volume.

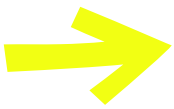
One example is self-publishing of books via print-on-demand services, which has increased markedly over the past 10 years. In 2023, 2.6 million ISBN-registered, self-published books were released in the United States alone, which is 4.7 times as many as the traditional book publishers release³. And there are unknown numbers of publications without ISBN numbers.

If we look broadly at cultural creation, releases within music⁴, books, podcasts⁵, YouTube videos (covering many cultural domains)⁶, theatre⁷, computer games⁸, film and series⁹ are growing year after year.

With the increasing volume of cultural production, the burden on the organisations that are responsible for registration and data management increases. The importance of correct registration regarding ownership of artistic work is more important than ever.

And with AI tools it has become easy both to imitate¹⁰ an artist and to publish it digitally¹¹. There is a need for verifiable, coherent metadata chains from the moment of creation, in order to be able to claim artistic works and artistic identity.

3 publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/96468-self-publishing-s-output-and-influence-continue-to-grow.html
4 musicbusinessworldwide.com/its-happened-100000-tracks-are-now-being-uploaded
5 grandviewresearch.com/industry-analysis/podcast-market
6 blog.youtube/news-and-events/2024-us-youtube-impact-report
7 edfringe.com/about-us/news-and-blog/weve-published-our-review-of-the-year-2024
8 app.sensortower.com/vgi/assets/reports/VGI_Global_Indie_Games_Market_Report_2024.pdf
9 obs.coe.int/en/web/observatoire/-/gbo-in-europe-up-to-eur-6.7-billion-in-2023-cinema-attendance-reached-861-million-tickets-sold
10 politifact.com/article/2025/oct/01/AI-music-charlie-kirk/
11 digitalmusicnews.com/2025/07/22/spotify-lax-controls-exposed-the-blaze-foley-ai-song-debacle



Reading guide

DataMosaik1 was prepared in dialogue with the Danish Agency for Digital Government (Digitaliseringsstyrelsen) and funded with resources from NextGenerationEU and Danish Institute for Cultural Policy Analysis (Kulturens Analyseinstitut). We would like to extend our sincere thanks to the Danish Agency for Digital Government's Office for Architecture and Technology for supporting the initiation of this mapping of the cultural sector's data landscape and practices, and to the participating organisations and respondents for generously sharing experience, data and time.

Purpose and background

The purpose of this mapping is to assess existing Danish cultural data sources and recommend them for publication on www.datavejviser.dk, the Danish Agency for Digital Government's overview of public data sources, as well as to identify opportunities and challenges in Danish data infrastructure and data practices in the cultural sector.

Although www.datavejviser.dk is limited to only containing publicly governed data sources, a meaningful analysis of the cultural sector's data infrastructure requires that private data sources and exchange practices are also included. This is because metadata about artistic works is often created in private parts of the value chain and only later enters public data infrastructures (e.g. libraries, public archives or other registers).

Mapping opportunities and barriers in the data area is in demand among decision makers and industry actors in the cultural sector who want access to higher data quality and data practices that can increase the ability to conduct relevant analyses across the different art domains and ensure data-driven insights that can illuminate cross-sectoral relationships - and thus strengthen the decision basis for both cultural actors and outsiders who engage with the cultural sector.

This mapping therefore primarily covers these three main areas:

1. **Description of a selection of Danish cultural datasets**, focusing on the composition of the selection, how the datasets are collected, updated and made available to different actors. The report provides insight into the methods used to validate datasets and describes the technological tools and standards that support their structure and interoperability. Finally, the datasets poten-

tial to support research, analysis and decision-making in the cultural sector is assessed, including their ability to enable new collaborations and innovation within digital cultural solutions.

2. Description of the consequences of incomplete/incorrectly registered data, looking more closely at the practical and strategic challenges that arise when data contains errors or is incomplete, using the music industry as a case domain. For this reason, the report examines implications for rights management, economic distribution and cultural heritage preservation, and how data errors can lead to lost opportunities, reduced insight, lost income or incorrect representation of cultural heritage. Recommendations are also presented that can strengthen data literacy in the cultural sector, including concrete measures and competence development that can promote better data registration procedures and data sharing practices.

3. Identified barriers and incentives for data literacy and data sharing, including a review of key challenges and opportunities, an assessment of existing practices for data sharing between actors, and proposals for concrete initiatives that can promote data competence and strengthen collaboration on sharing and using data in the cultural sector. In addition, the mapping identifies the most important barriers to increased data sharing (e.g. technical, legal or organisational obstacles) and assesses which incentives can support a more open and effective data exchange.

Case selection: domain specificity vs. generalisability

In addition to qualifying existing cultural datasets for potential inclusion in www.datavejviser.dk, the report's main focus is the music domain - particularly the handling of data on musical works - where there is high transparency in work processes and relatively easy access to information about metadata for musical works, as well as the consequences when these data are incorrec¹². There is also widespread use of international identifiers such as ISRC and ISWC, as well as the DDEX exchange standard, which enables direct data sharing and metadata validation. At the same time, actors in the industry have shown strong willingness to participate in interviews and help qualify the scope and insights of the study.

12 musicweek.com/publishing/read/noctil-ceo-founder-jacob-varghese-on-the-hidden-cost-of-bad-data/092756

Danish music industry actors also have extensive experience documenting and managing the challenges that arise from incorrect or incomplete registration/exchange of data on musical works, and they are very aware of how much this matters for the payment of musicians' royalties. See transparency reports from Danish organisations such as KODA¹³, Gramex¹⁴ and the Music Producers' Collecting Society (MPO)¹⁵, which manage their members' rights in musical works.

The music domain has therefore been chosen as the central case study to illuminate opportunities and barriers in data practice, but the findings indicate that similar issues apply - albeit in different forms - across all cultural domains.

Terminology: key concepts used in this report

The report uses concepts from data standardisation, data interoperability and data space thinking. The purpose is to reduce long, repetitive explanations in the text and - hopefully - to contribute to standardising data terminology. For an extended glossary, see [Appendix 4: Extended glossary](#).

API (Application Programming Interface): Refers to a software interface or solution that allows systems to exchange data automatically with each other, rather than data exchange being handled by humans.

Data space(s): A data space is a decentralised, shared framework for trust-based, secure and standardised exchange of data between multiple independent organisations. In a decentralised (often referred to as federated) data space, data are not shared in a central database; they remain with the actors that produce or manage them, without requiring everyone to give up ownership or control over their own data.

(Data) identifiers: Unique numbers or codes used to identify a specific work, recording, artist or release. They can be proprietary (only in one system) or interoperable (used across systems).

Interoperable data/systems/formats: The ability for different systems, data types, products or organisations to work together and exchange data smoothly - even if they are built by different companies or organisations.

13 koda.dk/pdf/Koda_Gennemsigthedsrapport_2024

14 gramex.dk/wp-content/uploads/2025/04/GF25_Gennemsigthedsrapport-2024_final_.pdf

15 mpo.dk/download/2184/?tmstv=1763472032

Cultural data sources: Organised collections of data about culture - i.e. data sources, datasets or databases that contain information about cultural works, activities and/or actors.

Metadata: Data about data, i.e. all the information that describes an artwork but is not the work itself. Metadata describe an artwork, a recording, a release or an actor (e.g. authorship, rights, time, place, roles, material, format).

MVM (Minimum Viable Metadata): “Minimum metadata that MUST be present.” The smallest set of information that is absolutely necessary for the work to function in data systems with respect to cultural heritage, archiving and attribution - i.e. data that must be completed to ensure traceability, settlement and preservation.

Royalties: The sum/payment that rights holders receive when their works are used.

Executive Summary

Key findings

1. Our mapping has identified 16 cultural data sources that are recommended for inclusion in www.datavejviser.dk. See Appendix 1: Cultural data sources for datavejviser.dk for details.
2. In our assessment of the openness, quality and usability of cultural data sources, we found that, across actors, there is a lack of shared definitions for data standards and minimum requirements for metadata.
3. We see inconsistent use of data identifiers and missing data bridges that could enable data exchange between actors - particularly between closed/internal identifiers that can only be used in one system and shared/open identifier systems.
4. Data literacy – or the ability to read, understand and use data - is not yet at an appropriate level across the cultural sector in our conclusion. The practical consequence includes unnecessary, repeated data entry and limited or reduced data sharing.
5. We see no shared standards or coordinated development of how digital cultural heritage is registered at this point in time. This means that information is often scattered and cannot be used together across systems, functions and institutions. Opportunities for insight into patterns around use, authorship, roles, etc. are lost.

Overall summary

In this mapping we assessed Danish cultural data sources in terms of whether (at the time of this mapping) they were included on www.datavejviser.dk, their potential use in analysis and commercial innovation, and their degree of openness. Here, openness is understood as access/availability - for example whether the data can be accessed via an API, or whether they are “only” available through a user interface on a website. As an individual it is easy to look things up on websites to gain concrete knowledge about a work of art, but websites are difficult to use to make two systems “talk” to each other.

In mapping the Danish cultural data infrastructure and data practice, we went deepest into the music domain and used cultural heritage preservation, data quality and data literacy as key perspectives. We found that the digital infrastructure is, in many cases, fragmented or split

into isolated silos. High-quality data sources exist, but they do not flow systematically between actors - among other reasons because identifiers are inconsistent or because there are no automated exchange mechanisms. This often results in manual double entry, loss of the data's context and significant difficulties both for cultural heritage preservation and for payments of royalties to rights holders. The latter can be seen, for example, in the amount of collected royalties that are not distributed year after year by organisations such as Gramex, MPO and KODA¹⁶.

These problems and challenges exist across the cultural sector. What happens in the music domain also happens in other cultural domains. Fragmentation, double entry and different understandings of standards weaken data quality and thereby, for example, royalty payments, innovation opportunities and cultural heritage preservation.

The data exchange challenges can be compared to the time when we did not have a requirement for a single standard for mobile phone charging connectors. One group had Nokias, and another had Samsungs. They could not be connected - unless you did it manually with a home-built special solution that works precisely for charging a Nokia 3310 and a Samsung flip phone, and nothing else.

The result is a patchwork of data exchange solutions that often prevents data sharing between actors. Only those who have invested resources in integrating one of the existing solutions get an efficient exchange of data. This creates barriers to collaboration and limits the use of cultural data across the sector.

The report is based on desk research on Danish cultural data sources and semi-structured interviews with representatives from the Danish music industry. In the review of Danish cultural data sources, we focused on identifiers, metadata fields and exchange methods across the data value chain at both Danish and international levels.

The Danish findings are confirmed by two larger European studies from 2025. The CITF report¹⁷ (Copyright Infrastructure Task Force), coordinated by the national libraries of Finland and Latvia, documents identical structural problems across the European copyright ecosystem and recommends federated solutions rather than centralised databases. This conclusion is based on experience with the Global Repertoire Database (GRD), which collapsed in 2014 despite major investments¹⁸. The OpenMuse consortium's report shows, through pilot projects in Slovakia and Latvia, that public-private collaboration around cultural data is both technically and institutionally possible¹⁹.

16 koda.dk/pdf/Koda_Gennemsigthedsrapport_2024
gramex.dk/wp-content/uploads/2025/04/GF25_Gennemsigthedsrapport-2024_final_.pdf
mpo.dk/download/2184/?tmstv=1763472032

17 <https://okm.fi/en/citf-en>

18 prsformusic.com/m-magazine/news/prs-music-issues-grd-statement

19 openmuse.eu/green-paper-on-ai-data-governance-and-metadata-policies-for-europes-music-ecosystem/

Both reports emphasise that fragmentation cannot be solved nationally alone. We assess that, with the right initiatives, Denmark can connect to the current European ambition to strengthen data infrastructure, which is now taking shape. The four pilot projects recommended in this report share the same aim of interoperability.

Main recommendations

This chapter describes a summarized version of the two main recommendations that we propose, and the four pilot projects that would activate them. For an in-depth description, please refer to the chapter “Key findings, elaborated”.

Recommendation 1: Strengthen data literacy in the cultural sector

To strengthen data literacy in the cultural sector we recommend developing targeted training programmes and competence development initiatives for both managers and staff. This could, for example, include courses in data understanding, workshops on best practice in data handling, and networks where actors can share experiences and get sparring on concrete data projects. In addition, guidelines and case examples should be developed to illustrate how stronger data literacy can lead to more effective use of cultural data and reduce errors and duplicated work.

Strengthening the cultural sector’s level of data literacy is critical for developing a stronger cultural life and future-proofed cultural heritage. We therefore recommend a stronger effort that both improves the practical use of cultural data from the creators’ perspective and works with exchange and use of standards. Rather than creating one central system, we recommend that the effort is steered through clear rule sets. The focus should be on developing practical solutions that many can use and on establishing shared governance between actors.

We recommend the following guidelines to strengthen Danish cultural data infrastructure and data practice:

Shared purpose: Cultural data must also serve institutional needs, public goals and social benefit (e.g. ensuring correct and time-efficient settlement of royalties relating to copyright), beyond operational optimisation.

Shared reference frameworks: There should be minimum data requirements and shared data identifiers that enable collaboration and data exchange across actors.

Infrastructure, not software: Data exchange should be based on APIs, standards and easy integration - not only one system.

Collaboration-driven governance: Cross-disciplinary steering/working groups should govern classifications and concepts, the inclusion of new standards, and organisational buy-in.

Recommendation 2: Establish a data space for the Danish cultural sector

Based on the mapping findings and the recommendations above, we believe a coordinated effort should be launched to establish a Danish data space for actors in the cultural sector.

A data space is, overall, about organising data exchange²⁰. This model is gaining momentum in the EU and globally across many sector areas. The term “data space” describes a model for decentralised data infrastructure that enables secure data exchange across different actors, organisations and sectors through shared rules, principles and standards. Kulturens Analyseinstitut is also a member of the Danish Agency for Digital Government’s network, Danish Data Space Forum.

The Danish Agency for Digital Government describes the forum as follows:

“Danish Data Space Forum works to promote the development and use of data sharing in Denmark. The forum enables knowledge sharing, networking and supports the spread of data spaces in Denmark, with a particular focus on increased data use among Danish authorities and businesses. The initiative is part of the Government’s Digitalisation Strategy, which includes establishing a Danish Gaia-X hub that will function as a national contact and coordination point for Danish actors within the European data economy.”

We regard this decentralised data space organisation as the best strategy to support increased data sharing and data literacy in the cultural sector.

From recommendations to four pilot projects

There is a need to establish shared guidelines for which information (metadata) should, at a minimum, accompany works of art - a kind of data checklist. At the same time, better solutions are needed so systems can communicate better, even when they use different codes or ways of identifying things. Finally, shared rules and strong reasons to participate should be established (e.g. faster case handling or specific benefits), and this work can be initiated through pilot projects.

These four pilot project initiatives are described in more detail in the chapter “Possible solutions, or paths of action”.

Pilot A - Danish data space for the cultural sector

Pilot A aims to establish the organisational foundation for a Danish data space for arts and culture by setting up working groups and a steering group that defines principles for data exchange, an MVM profile, and incentive structures for the other pilot project initiatives. See subchapter “Pilot A – Governance framework and data space principles” for more details.

²⁰ digital-strategy.ec.europa.eu/en/policies/data-spaces

Pilot B - Danish ISNI registration initiative

Inspired by the Finnish ISNI project²¹, we recommend initiating ISNI registration and matching across data sources in Denmark. See subchapter “Pilot B – ISNI registration and identity matching” for more details.

Pilot C - Verification of cultural datasets from actors

The purpose of Pilot C is to establish methods for automated data exchange, with neutral operators or intermediaries that enable secure data sharing between actors without centralising ownership or disclosing business-sensitive information. Our approach to intermediaries and operators is based on the Data Spaces Support Centre’s definitions²². See subchapter “Pilot C – Verified metadata via intermediaries” for more details.

Pilot D - Cultural heritage data enrichment

Pilot D aims to create a more systematic approach to verification and enrichment of cultural heritage and cultural data. For example, by registering where and when a work was created, cultural data and traceability are improved. Museums can verify information by sharing digital collections on a platform where metadata are automatically checked for quality and standards. Data matching between, for example, record labels and libraries enables the linking of biographical data and work descriptions, giving users better access to relevant information and revealing errors or duplicates. See subchapter “Pilot D – Cultural heritage enrichment with geographic and temporal anchoring” for more details.

²¹ kansalliskirjasto.fi/en/services/isni

²² dssc.eu/space/BVE2/1071253470/Intermediaries+and+Operators

Methodology

Period and purpose

The DataMosaik1 project ran from May to December 2025 as a qualitative mapping process with a focus on practice-oriented analysis, with the following goals:

- Mapping and categorisation of existing and potential Danish cultural data sources
 - Assessment of the openness, quality and usability of cultural data sources
 - Metadata descriptions for selected cultural data sources
 - Identification of barriers and incentives for increased data literacy and data sharing
 - Policy recommendations on how to strengthen data literacy in the cultural sector
-

Data collection and analysis method

The mapping sought to cover the entire data value chain from production through distribution to cultural heritage preservation, by examining how Danish cultural data infrastructure works in practice, which factors prevent data exchange between cultural actors, and which initiatives can best respond to the insights the mapping reveals.

The report's quantitative data basis consists of desk research on cultural data sources that were mapped and categorised in relation to identifiers, metadata fields, exchange forms and governance.

The report's qualitative data basis consists of 10 semi-structured interviews with central actors in the Danish music industry and two actors within cultural heritage professions (see Table 1 below). The interviews covered representatives from four actor types: collecting societies, commercial actors, public actors, and political organisations/trade unions.

Name	Actor category	Data area	Cultural domain	Respondent's role/function
GRAMEX	Collective management organisation	Mechanical rights	Music	Data operations
KODA	Collective management organisation	Copyright (authors' rights)	Music	Development and data operations
MPO	Collective management organisation	Producer rights	Music	Data operations
Sony Music Entertainment Denmark	Commercial actor	Label	Music	Data operations
Warner Music Denmark	Commercial actor	Label	Music	Management
Wise Music Group	Commercial actor	Label/Publishing	Music	Data operations
United Music Society	Commercial actor	Label/Publishing	Music	Management and data operations
DBC Digital	Publicly owned actor	Data infrastructure	Cultural heritage	Development and data operations
The Royal Danish Library (Det Kongelige Bibliotek)	Public actor	Data infrastructure	Cultural heritage	Data operations
DPA	Interest organisation	Advisory services	Music	Management
DMF	Trade union	Advisory services	Music	Management
Anna Lidell	Former chair of Autor and KODA	Copyright (authors' rights)	Music	Management

Table 1: interview respondents by organisation/name, actor category, data area, cultural domain and respondent's role/function.

The interviews were structured around four core themes:

1. **Data identifiers:** Which identification systems are used, and how are they used across systems?
2. **Metadata fields/registration practice:** Which data fields are registered, and which are considered critical by the respondent?
3. **Forms of exchange:** How are data shared with others?
4. **Governance:** Who owns the data, who updates them, and how are data gaps and registration errors handled?

Seven of the interviewees were part of, or close to, operations relating to data management, and five were in management layers and could therefore only speak more generally about the core themes. Interviews were recorded and transcribed using AI tools and then spot-checked for errors. To ensure validity, selected informants were given the opportunity to read and approve quotes and factual descriptions of their organisations' practices. This reduced the risk of misunderstandings and of technical details being represented incorrectly.

Cultural data sources covered by the mapping

Below is an excerpt of the cultural data sources mapped in this report. See Appendix 1: Proposed cultural data sources for www.datavejviser.dk for further details regarding source links, data identifiers, data owners, accessibility, etc.

Cultural data source title	Cultural data source owner	Reuse assessment
MediaStream	The Royal Danish Library (Det Kgl. Bibliotek)	Low
Arkiv.dk / Arkibas	Municipal / association-run	Low
Image and poster archive (Billed og plakatarbivet)	Danish Film Institute (Det Danske Filminstitut)	Low
Denmark on Film (Danmark på Film)	Danish Film Institute (Det Danske Filminstitut)	Low
Danish cultural data (Dansk kulturdata)	Europeana	High
DR archive (DR-arkivet)	The Royal Danish Library (Det Kgl. Bibliotek)	Low
Festival history (Festivalhistorik)	Festivalhistorik.dk	Low
The Film Database (Filmdatabasen)	Danish Film Institute (Det Danske Filminstitut)	Medium
Performance data (Forestillingsdata)	Teaterbilletter.dk	Low
Listed and preservation-worthy buildings (Fredede og Bevaringsværdige Bygninger)	Danish Agency for Culture and Palaces (Slots- og Kulturstyrelsen)	Low
Finds and ancient monuments (Fund og Fortidsminder)	Danish Agency for Culture and Palaces (Slots- og Kulturstyrelsen)	Medium
Art Index Denmark & Weilbach's Artist Lexicon (Kunstindeks Danmark & Weilbachs Kunstnerleksikon)	Danish Agency for Culture and Palaces (Slots- og Kulturstyrelsen)	Low
LOAR	The Royal Danish Library (Det Kgl. Bibliotek)	Medium
Museum collections (Museernes Samlinger)	Danish Agency for Culture and Palaces (Slots- og Kulturstyrelsen)	Low
Netarkivet	The Royal Danish Library (Det Kgl. Bibliotek)	Low
Publications	Kulturens Analyseinstitut	Low
Audience surveys (Publikumsundersøgelser)	Applaus	Low
Collections Online (Samlinger Online)	National Museum of Denmark (Nationalmuseet)	Low
Cultural event calendar (Kultureventkalender)	Kulturnaut	Medium

Table 2: excerpt of the table of cultural data sources mapped in this report (see [Appendix 1: Proposed cultural data sources for www.datavejviser.dk](#)).

Ethics and data protection

No sensitive personal data are included in this report. When processing example datasets, we applied privacy-by-design principles, where personally identifiable information was pseudonymised or removed. Informants had the right to withdraw statements, and all quotes are anonymised unless the informant has explicitly granted permission for name/organisation reference.

Scope and generalisability

The report uses music as the case domain for assessing data infrastructure due to the following three conditions:

1. Data transparency: collecting societies' annual reports document data challenges and the financial consequences they entail.
2. Mature international data standards: there is broad implementation of data identifiers.
3. Actor willingness: the music industry has shown strong willingness to participate in interviews and validation of hypotheses and findings.

The report's findings are primarily anchored in an in-depth analysis of the music industry, but through additional interviews and desk research we tested whether these findings also apply to other domains in the cultural sector. What we saw suggests that they face exactly the same challenges: fragmented data systems, entry of the same data in multiple places, and metadata that are not preserved or shared - leading to similar losses of overview, cultural heritage preservation, accessibility and rights management.

During the work we asked ourselves several times: what is culture? In the analysis of what is available on www.datavejviser.dk we found, for example, the continuously updated "culture" data source Denmark's Environmental Portal – Conservation Data (Danmarks Miljøportal – Fredningsdata). This may not have a direct connection to data on the arts, but in our discussion of missing data and where digital cultural data are registered, one can argue that data sources such as this have a direct significance for culture because they visualise distribution and provide overview.

How is the term *culture* delimited and used in this report?

This report applies an operational and analytical definition, or delimitation, of the concept of culture, adapted to the purpose of mapping cultural data sources relevant to data infrastructure, rights management and cultural heritage preservation. The delimitation is functionally motivated and is intended to ensure consistency and comparability in the identification and assessment of arts and culture connected datasets.

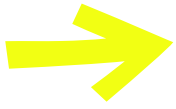
In DataMosaik1, the concept of culture is understood concretely as:

- artistic works and cultural objects (e.g. books, music, film, performing arts, visual arts, museum objects and archival materials),
- associated actors and rights relationships (authors/creators, performers, producers, institutions and collective management organisations),
- cultural activities and forms of distribution (performances, exhibitions, events, loans, ticket sales, streaming, etc.),
- preservation and documentation contexts (bibliographic registrations, archival data, collection and authority registers).

The delimitation is based on an understanding of culture as documentable and institutionally organised artistic production, and as a field of production and distribution consisting of works, actors and institutional relationships. In extension of this, cultural products are understood as works with creative content and symbolic meaning that are typically subject to copyright regulation. This tripartite perspective is particularly relevant in a data-infrastructure context, where metadata, identifiers and rights attribution relationships constitute central structural elements.

The delimitation is consistent with the European Commission's definition of cultural and creative sector(s), which includes the development, production, distribution and preservation of cultural expressions and works. The report thus adopts a domain-specific understanding that links culture to production of works, authorship, rights and institutional practice. Data sources may also be included when they contribute to documentation, preservation or institutional stewardship of cultural heritage.

Broader anthropological conceptions of culture - such as UNESCO's definition of culture as ways of life, values and traditions - lie outside the report's analytical focus.



“I could get my grandmother’s old Pentium computer to run all registrations and payouts [in the music domain, ed.], if only the data were there, if only they were shared, if they were complete.”

Morten Hjort, Head of Data and Distribution, Music Producers’ Collecting Society / MPO

Key findings, elaborated

This chapter reviews the findings that the mapping of the data infrastructure has uncovered, first in a descriptive section with the main insights, and then a gap analysis of these insights.

Layers of the data infrastructure

Based on our desk research and interviews, we identified four layers in the Danish music industry’s data infrastructure. Ideally these layers should be connected, but in practice they are very often isolated from each other. This section first describes the four data layers in the music industry, followed by an account of the challenges that arise in the interaction between them. We have divided the four data layers in the music industry as follows:

Data layer 1: Agent, work and recording registers constitute the first and fundamental layer where identities and musical works are registered or created. This is where the basic identifiers are created, which the rest of the data value chain depends on - both for cultural heritage preservation and commercial interests.

Data layer 2: Distribution and publishing systems include production systems, digital distributors and streaming platforms. Here, data and musical works from layer 1 are received.

Data layer 3: Library registers represent the public preservation infrastructure, primarily DBC Digital’s and the Royal Danish Library’s systems and public service archives. These systems partly use other standards than the production systems in

layer 2, which creates frictions between systems. We identified a number of parameters that can be problematic in relation to cultural heritage preservation and local identity anchoring - for example a lack of systematic registration of the geographical origin of works and cultural productions.

Data layer 4: Tools for aggregation and reporting include settlement and member portals (e.g. Gramex, MPO and KODA), artist portals (e.g. Spotify for Artists, Bandcamp for Artists and YouTube for Artists), analytics tools (e.g. WARM and Playback Studios), and various distribution systems (e.g. Sleeve, SongTrust, Believe and The Orchard). These systems attempt to assemble data from the other layers but consistently encounter matching challenges due to inconsistent or varying identifiers across sources, and reluctance to share data that are close to the actors' confidential business models.

Challenges in the data layers

Our desk research and interviews point to three central challenges that arise in the interaction between these data layers:

Challenge #1: Inconsistent use of data identifiers

Even though standards exist and are, in principle, adopted by the actors, they are used inconsistently. For example, several interviews and desk research on music data indicated that at least one of the identifier fields is missing or erroneous in between 10% and 15% of music releases²³, and that linkage to rights holders is often only established at royalty settlement (which can sometimes be months/years after release, rather than at production time). This creates “identifier breaks” where traceability is lost between layers. It is important to note that both in Denmark and internationally there is ongoing work on solutions to strengthen the linkage between work data and sound recordings before release.

Challenge #2: Varying metadata field names and metadata models

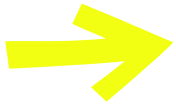
Even when the same information is registered, different field names, structures and controlled vocabularies are used. “Genre” can be free text for one actor, a hierarchical tree for another, and a flat set of 20 tags for a third. “Artist name” can be one field or split into first name/last name. “Role” (e.g. producer, mixer, featured artist) is described without a shared taxonomy. This makes automated mapping between systems difficult and necessitates manual work/corrections - something that is often hard or nearly impossible given today's high frequency of music releases.

23 [billboard.com/pro/unclaimed-black-box-royalties-how-much-money](https://www.billboard.com/pro/unclaimed-black-box-royalties-how-much-money)

Challenge #3: Lack of requirements for minimum metadata quality (MVM)

There is no shared understanding of which metadata are critical for ensuring correct attribution, traceability and cultural heritage preservation. Interviews show that many actors have very different definitions of “minimum requirements/needs for registration fields”. Major labels typically operate with minimum profiles, while some distributors require only title and artist; unfortunately, metadata quality is ultimately reduced to the weakest link in the value chain.

These three challenges are highly transferable to other cultural domains. The coherence and interoperability between the different layers are therefore decisive for how effectively cultural data can be used and reused across the sector. A robust data infrastructure where data flow smoothly between the layers will not only strengthen the cultural value chain but also increase the potential for innovation and cross-actor collaboration, and better future-proof Danish cultural heritage. It is therefore central to address the identified challenges in multiple fora in order to both elevate and future-proof the Danish cultural data ecosystem.



“...it is an advantage if processing of data can take place - untouched by human hands.

Because when a song is played on, for example, Spotify, there should ideally be a settlement to rights holders as shortly thereafter as possible.

However, there are many links in this data chain - registration with digital music distributors, onward transmission, matching, claiming, calculation, quality assurance, and so on.

And the more correctly the different kinds of music data in the chain are registered from the start, the more automated and efficient one can run one's business.”

Kasper Frost Iversen, Head of KODA's Music Data Management team

Gap analysis of key findings

Through the mapping interviews we identified seven central “gaps” that together constitute the most important challenges in the cultural data area. These gaps are not isolated issues; they are connected in a pattern where each challenge reinforces the others.

Gap 1 – Double entry and redundant data work

Artists and administrators register the same metadata repeatedly across systems. A typical music release requires entry in collecting societies' systems, distributors' platforms, labels' production systems and library systems. Interviews documented that, in one example, a couple had to enter identical information - artist name, song title, composers, recording date - into up to seven different systems. This repeated work is not only time-consuming; it also introduces a higher risk of errors, as manually entered data often differ between systems (different interpretations of minimum data, spelling errors, format variations, missing fields, etc.).

The problem arises because systems do not “talk” to each other. Even when an artist has registered complete metadata in one place, these data cannot automatically be shared onwards. Each actor in the value chain operates its own closed system, and integration between systems is either absent or based on periodic file transfers rather than live API exchange.

Gap 2 – Identifier breaks in the value chain

Although international standards exist, they are used unevenly and therefore create identifier breaks in data exchanges. This means that the chain of work/recording/release cannot be traced automatically. In music, for example, when a song is played, systems cannot securely match streams to the correct musical works and thereby the correct rights holders. The result is that manual matching processes are necessary, which often significantly delay royalty settlements and, in the worst case, can cause lost income for artists if works cannot be identified unambiguously.

Gap 3 – Lack of shared minimum metadata (MVM)

There is no shared understanding of which metadata constitute a critical minimum across actors. The analysis revealed widely different definitions: some operate with minimum profiles; others require only title and artist. Collecting societies have their own minimum requirements, distributors have others, and streaming platforms require something else again.

In one system for registering recorded music, one can currently register the musician role “bass” without further specification - which can be interpreted as many things, e.g. electric bass, synth bass, bassoon, double bass or vocal bass. There is also no possibility to register where or when the work was recorded. Such information is registered arbitrarily in other dissemination systems, but is not always preserved, for example, in the National Bibliography’s registration of the music release.

In other words, the data chain is only as strong as the metadata that follow the work. Important metadata such as detailed credits, geographic origin, language codes and cultural-historical context drop out because they are not “mandatory” metadata fields at all links in the chain.

Gap 4 – Exchange via files rather than APIs

CSV files and manual file exchanges still dominate data sharing between many actors. Even though many systems have APIs, these are often insufficiently documented, not systematically versioned, or limited to read-only access.

File-based exchange between people introduces timeliness problems (e.g. data are always out of date), requires manual processes (someone must remember - or have time - to export and import), and makes error handling difficult (if a file is corrupt, it is only discovered upon import, often days after export). The absence of well-documented, standardised APIs with clear data models and error handling is a critical barrier to automation.

Gap 5 – Governance and weak feedback loops

It is unclear who owns the data and the exchange rules between systems. When genre lists must be updated, who decides the new genres? When new role types (e.g. “co-producer” or “recording engineer”) must be added, which process is followed? Interviews pointed to ad hoc decisions, lack of documentation of changes, and no systematic feedback loop from data consumers to data producers.

This governance vacuum means that data quality is not improved systematically. Errors are not reported back to the source, so they recur. New needs (e.g. AI training data, ownership of such data, or opting out of being included in AI training sets) cannot be integrated quickly because there are no processes to coordinate changes across actors. Responsibility for data quality is diffuse, and no party has a mandate to drive improvements that require changes by others.

Gap 6 – Missing metadata preservation and cultural heritage preservation

“Where was this song recorded?” “Which local cultural scene does it represent?” “Who had which instrumental roles in the recording?” - these questions can often not be answered from the metadata that follow individual digital works as they exist today.

This gap has direct consequences for cultural heritage preservation. When today’s digital cultural production does not include data that can be anchored geographically and in time periods or in relation to a phenomenon/genre etc. (phonk, dansktop, grunge), future generations lose the possibility of understanding culture’s local roots and development over time. Libraries and archives experience that newer material is harder to contextualise and preserve meaningfully than older physical media, which often had such information printed on the cover or in liner notes.

Gap 7 – Missing bridges between closed and open identifiers

Actors use closed (proprietary) identifiers internally (internal work IDs, label codes, plat-

form-specific barcodes), but these are rarely systematically linked to global open standards. A release can have an internal ID at a record label, a work identifier code at a streaming platform, and a recording identifier at a collecting society, but these identifiers are not connected to the open global identifiers (ISRC, ISWC or ISNI) in a meaningful way.

In Finland, several cultural organisations have adopted ISNI as a global, permanent identification system for rights holders. It strengthens metadata quality, facilitates international information exchange, and can be used together with other identifiers. Their initial experience shows how a shared standard can create interoperability and reduce manual linking in the value chain.

The absence of these “identifier bridges” prevents traceability across systems. One cannot automatically follow a recording from studio to streaming to settlement to archive, because identifiers change at each step and there are no machine-readable mappings. This limits the possibility of automated reconciliation and data-driven insight into cultural production and consumption.

Interdependencies and consequences of the gaps

These seven gaps reinforce each other. Double entry (G1) happens because identifiers do not follow along (G2) and systems cannot talk to each other (G4). Identifier breaks (G2) are worsened by the lack of shared minimum metadata requirements (G3) and the lack of bridges between identifier systems (G7). The absence of governance (G5) means these issues are not solved systematically. And the lack of focus on foundational data (G6) reduces the long-term value of cultural data, even if the technical issues are resolved.

Summary of the gap analysis

G1–G4 directly increase transaction costs and reduce data quality (time spent on entry, error handling, manual matching).

G5 prevents systematic improvement (without governance, databases and infrastructure remain fragmented).

G6 weakens cultural heritage preservation and local cultural identity in the long term.

G7 limits automation, transparency and correct settlement of royalties.

Overall, the gap analysis documents - via the music domain - that the Danish cultural data infrastructure suffers from structural rather than isolated challenges.

International confirmation of the gaps

The seven gaps are not unique to Denmark. The Copyright Infrastructure Task Force (CITF), led by the national libraries of Finland and Latvia, documents similar structural problems across the international copyright ecosystem in its report “Interoperable, trustworthy and machine-readable copyright data in the AI era” (October 2025)²⁴. CITF presents a three-layer model for the future copyright infrastructure, which can be directly linked to the findings in DataMosaik1’s gap analysis:

The fundamental legal layer includes authoritative identifiers for agents and assets. CITF emphasises that these must be open, trustworthy and interoperable - directly addressing Gap 2 (identifier breaks) and Gap 7 (missing bridges between systems).

The semantic layer enables consistent interpretation of meaning across systems without requiring one common ontology. This supports the need for shared reference frameworks and a Minimum Viable Metadata (MVM) requirement (Gap 3) rather than a centralised system.

The technical layer covers APIs, mappings and distribution services. CITF concludes that federated data spaces should be prioritised over centralised databases, and that data exchange should be based on APIs rather than files (Gap 4).

Similarly, the OpenMuse consortium documents in its “Green Paper on AI, Data Governance, and Metadata Policies for Europe’s Music Ecosystem”²⁵ that attempts at centralised solutions have repeatedly failed. The Global Repertoire Database (GRD) project collapsed in 2014 despite massive investments, primarily due to governance conflicts, high costs and lack of trust between actors. OpenMuse concludes that the desired scenario is when “each actor stays in control of its own data, but agrees to shared profiles, identifiers, and rules”. This is the federated approach on which this report’s recommendations are built.

The gap analysis linked to low data literacy in the cultural sector

The mapping identified several systematic barriers to improving data literacy in the cultural sector, which in turn hinder realising the use potential embedded in Danish cultural data.

These barriers operate across technical, organisational and competence dimensions and reinforce each other in a pattern where low data literacy legitimises the continuation of fragmented and problematic data practices.

²⁴ okm.fi/en/citf-en

²⁵ openmuse.eu/green-paper-on-ai-data-governance-and-metadata-policies-for-europes-music-ecosystem/

At the individual level, artists and cultural producers often lack a basic understanding of why correct metadata are critical for their income and visibility. Interviews documented that artists repeat administrative processes without understanding these connections. This lack of data literacy creates incentives for minimum effort, with a “just get it registered” approach, rather than quality assurance.

When the consequence of incorrect registration only becomes visible months later - for example in a settlement or as missing information on commercial services - the feedback loop is broken (Gap 5 – weak governance), and learning does not take place. Strengthening data literacy at artist level requires practice-oriented guidance that demonstrates the direct relationship between data quality and economic return, as well as concrete tools for self-validation of metadata before sharing productions.

At the institutional level, cultural organisations often operate with limited technical capacity and unclear data governance. Smaller organisations lack resources for systematic data work, while larger institutions struggle to move away from legacy systems and silo-oriented data ownership with proprietary identifiers.

Gap 5, concerning weak governance, manifests as uncertainty about who has the mandate to decide on data standards, how changes are coordinated across systems, and how data quality is measured and improved systematically. The absence of shared structures means that an investment in data quality by one actor does not necessarily benefit that actor directly, which weakens motivation for improvement. Strengthening institutional data literacy requires cross-sector governance structures with clear decision procedures, documented standards and measurable quality requirements - as well as visible benefits for contributing parties. There are few examples internationally and locally where groups have worked on such structures, but the Finnish ISNI example benefits both individuals, institutions and the shared framework.

At the system level, barriers are reproduced through the lack of shared reference frameworks and weak incentives for open data sharing. Without shared minimum data (Gap 3) and openly shareable identifiers (Gap 2), cultural data infrastructure remains fragmented even if individual actors want improvements. Open data sharing is hindered by real and perceived conflicts about competition, ownership and commercial interests. At the same time, positive incentives are missing. There are no recognition mechanisms for organisations that contribute high data quality, and no sanctions for actors that maintain closed systems or deliver low-quality data.

Recommendations on how to strengthen data literacy in the cultural sector

Succeeding requires both political support for shared standards and public resources for competence development, as well as pilot projects that demonstrate concrete value of improved data literacy for all stakeholders.

We therefore recommend the following three focus areas:

1. **Practice-oriented guidance** for artists that demonstrates the link between data quality and income, together with tools for self-validation.
2. **Cross-sector governance structures** with clear decision procedures, documented minimum data and measurable quality requirements.
3. **An incentive-based approach** that rewards open data sharing and high quality through visible recognition, reduced administrative burden for contributing parties, and coordinated implementation where improvements benefit all actors simultaneously.

Conclusion

Danish cultural data infrastructure contains significant untapped potential, spanning from metadata about works in relation to, for example, cultural heritage, ticket sales and rights, to user access and digital registration chains. Through the mapping of cultural data sources, it is clear that freeing up the useability requires a bridge between the cultural sector's domain knowledge and technical data use/data competences.

The mapping documents that Danish cultural data infrastructure is functionally fragmented. Data exist, but generally they do not flow particularly freely - or in an automated form - between actors. Standards exist, but are not applied consistently. Systems exist, but do not communicate well. This fragmentation manifests as double entry for artists and administrative staff, lost royalties due to incorrect settlement or delays that exceed the time limit for collection, reduced traceability for users and the public, and thereby weakened cultural heritage preservation.

At the infrastructure level, the organisations DBC Digital and the Royal Danish Library's collections constitute cultural datasets with high readability and accessibility about published material, with metadata that are standardised and can be used across cultural domains. For musicians and cultural producers, this is decisive infrastructure for correct attribution/registration and rights management. Data from, for example, Statistics Denmark's cultural habits surveys, Applaus audience surveys and Kulturens Analyseinstitut's municipal cultural profiles enable market analyses, audience segmentation and location-based cultural analysis for strategic planning and development of, among other things, evidence-based cultural policy and urban planning.

The place-based dimension constitutes a particular strength through combining conservation data, the Danish National Archives' registers and the Royal Danish Library's collections to map local cultural history. When these are linked with cultural event data from, for example, Kulturnaut, Festivalhistorik and Teaterbilletter.dk, temporal depth is created in the understanding of cultural life - while it is taking place. Art Index Denmark and Weilbach's Artist Lexicon show the potential in standardised person data, which can be expanded to, for example, composers and authors through international standards such as the Finnish ISNI solution.

There is great potential in cross-sector collaboration and combining datasets. When bibliographic systems are linked with, for example, event data, audience data and geodata, opportunities arise for AI training, many types of commercial innovation, cultural-historical research and evidence-based cultural policy. This requires strengthened data literacy in the cultural sector, as well as shared standards for openness and data quality, which can thereby realise the full use potential of cultural datasets across institutions and domains.

Through mapping cultural data sources and interviews we documented that Danish cultural data infrastructure exists across four layers as described in the section “Layers of the data infrastructure”. The four layers are: agent and work registers, distribution and release systems, library registers, and aggregation and reporting tools. Although these four layers should connect in a cultural data ecosystem, the cultural data sources are functionally separated or only partially exchanging data. Data exist and are produced continuously, but do not flow stably between actors due to widespread practice - and often necessity - of manual (double) registrations, uneven use of standards, and the absence of shared minimum requirements and governance.

The analysis identified seven mutually reinforcing gaps: Double entry (G1) arises because there are identifier breaks (G2) and missing API-based data exchange (G4) that force artists to enter the same information in many different systems. This is worsened by the absence of shared minimum metadata requirements (G3) and missing bridges between closed and open identification systems (G7). The problems are reproduced systematically because weak governance (G5) prevents coordinated solutions, while lack of focus on foundational data (G6) undermines the cultural heritage value of the data. The consequences are measurable: several respondents confirm that at least 10% of music rights registrations contain errors due to metadata problems; artists spend multiple hours per week on redundant administration; and cultural heritage preservation is weakened through loss of context, etc. The Finnish ISNI project demonstrates how coordinated implementation of shared standards can reduce these frictions and errors significantly.

Possible solutions, or paths of action

We recommend the following three parallel paths of action to address the challenges:

Minimum Viable Metadata (MVM) with validation

A common MVM profile defines critical metadata across the value chain. Validation should take place at the source through quality checks (e.g. that all variables are completed and that reference look-ups are resolved via operators/intermediaries) before the cultural work is (re) distributed.

ID bridges and technical exchange methods

Systematic linking between identifiers through documented APIs rather than file handovers. This requires both technical implementation (versioned APIs with error handling) and organisational coordination.

Governance and incentive framework

A cross-sector governance structure with clear decision processes, measurable quality requirements and “trusted partner” status (reduced case-handling time) for actors with documented high data quality.

The paths of action in the form of four pilot projects

Operationalising the three paths of action above points toward a Danish cultural data space. The four pilot projects are intended to be initiated simultaneously, as they are interdependent and address the challenges identified in the Gap Analysis section and build on other organisations’ experiences and results.

Pilot A – Governance framework and data space principles

The purpose of Pilot A is to create the organisational foundation for a Danish data space for the cultural sector. This will be done by establishing working groups and a steering group that together develop and anchor the principles for how data are exchanged between actors in the sector. A key task for the steering group is to define and document who makes which decisions in the collaboration, and how competitive considerations among participants are handled. In addition, a central focus is

to define and describe the shared “minimum viable metadata” profile (MVM), which will form the basis for effective, high-quality data exchange between parties. Finally, the pilot will establish the incentive structures that support collaboration and motivate actors to deliver high data quality.

Pilot A deliverables:

- Establishment of a steering group with decision mandate
- Documentation of current practice among participating actors
- Development of MVM v1.0
- Definition of “trusted partner” criteria
- Publication of shared terminology

Pilot A goals:

- MVM v1.0 approved by the steering group and implementation initiated among actors
- Letters of intent from participating actors
- Governance document with clear decision procedures developed by the steering group
- Established meeting structure (steering group quarterly, working groups monthly)
- Documented “trusted partner” programme with benefits
- Agreement among participants on prioritised standards

Pilot B – ISNI registration and identity matching

The purpose of Pilot B is to implement the ISNI identifier system as a shared identifier for artists and rights holders across Danish data systems in the cultural sector.

The pilot implements the Finnish ISNI project, where the Finnish National Library functions as a neutral coordinator of a rights holder database. The Finnish collecting society Teosto now assigns ISNIs to its members, giving authors and composers persistent identifiers that connect to, for example, ISRC, VIAF, ORCID and Wikidata. Experience shows how a shared standard can create interoperability and reduce manual linking in the data exchange chain.

The CITF report points out that national libraries already curate large amounts of copyright-protected material and maintain authoritative identifiers, but remain largely disconnected from rights metadata workflows. The report recommends treating national libraries as copyright infrastructure actors, not only cultural heritage stewards, an insight that is directly relevant for the Royal Danish Library’s and DBC’s potential roles in the pilot.

Pilot B deliverables:

- Mapping of existing artist identifiers in participating organisations
- Establishment of matching processes between local IDs and ISNI
- Publication of guidelines for ISNI use in the Danish context

Pilot B goals:

- A defined number of Danish artists registered with ISNI within the pilot
- Published API documentation for an ISNI lookup service

Pilot C – Verified metadata via intermediaries

The purpose of Pilot C is to establish trusted data exchange through neutral operators (intermediaries) that enable secure data sharing between actors without centralising ownership. The pilot demonstrates the “publish once, distribute many” principle, where metadata are validated at the source and distributed automatically to all relevant systems.

Pilot C deliverables:

- Design and implementation of a data space API
- Automated MVM validation when processing data from actors
- Push notifications to relevant systems for new releases
- A transparency report structure developed and published on a regular basis

Pilot C goals:

- A defined number of releases processed through a verified flow
- Reduction of double entry measured among actors (fewer cases requiring manual re-entry)

Pilot D – Cultural heritage enrichment with geographic and temporal anchoring

The purpose of Pilot D is to strengthen cultural heritage preservation through systematic registration of production location and production time for works. The pilot demonstrates how enriched metadata create value for research, local identity anchoring and future cultural-historical understanding, and it addresses Gap 6 directly.

The OpenMuse report uses the concept “frozen assets” about cultural heritage that remains “frozen” because documentation costs exceed the current commercial value. This applies particularly to non-commercial works, such as indie label releases, which can be culturally valuable but not mainstream enough to break through contemporary marketing walls. Without affordable workflows, these works cannot enter modern distribution systems regardless of cultural or artistic significance. Pilot D will contribute to cultural heritage safeguarding so that works can be rediscovered, contextualised and thereby “thawed” for future use.

Pilot D deliverables:

- Expansion of the MVM profile with geographic location and production time
- Enrichment of a defined number of releases with location and time data in collaboration with artists and actors
- Integration with existing geodata
- Publication of best-practice guidance for artists on cultural heritage registration

Pilot D goals:

- A defined number of works registered with geographic origin and production time
- Implemented search functionality with geo/time filters in selected systems (e.g. DBC, the Royal Danish Library)

Recommendations for role allocation

To make the recommendations actionable, a possible role and responsibility model as well as a possible phase breakdown are outlined here.

Kulturens Analyseinstitut can take on the task as a neutral coordinator and function as secretariat for the programme. The institute can, for example, facilitate processes, document activities and results, continuously evaluate the effort, and disseminate knowledge to relevant parties. In addition, Kulturens Analyseinstitut could handle communication to external stakeholders so that the programme's activities align with the broader data initiatives in the cultural sector. The Danish Agency for Digital Government (Digitaliseringsstyrelsen) should participate as a partner that ensures anchoring of the programme in Danish Data Space Forum. The agency can provide legitimacy and credibility regarding the programme's importance. At the same time, the agency can function as a bridge to other relevant authorities, for example the Danish Business Authority (Erhvervsstyrelsen) and their focus on data quality, and thereby ensure broader anchoring and cross-sector collaboration.

Creative industry actors such as KODA, GRAMEX, MPO, NCB, publishers and distributors contribute as domain experts and implementation partners. They articulate specific needs, commit to MVM requirements, test ID bridges and participate actively in changes to registration practices so that solutions are embedded in the sector's daily work.

Public infrastructure operators such as DBC Digital and the Royal Danish Library have central roles regarding cultural heritage and registration standards. They provide authoritative identifiers and work for semantic harmonisation across systems. At the same time, they can act as potential intermediaries between different actors in the ecosystem.

Private-sector technical operators and suppliers are responsible for developing the technical solutions, including connectors, APIs, validation components and access mechanisms. All development work should be carried out in accordance with data space principles to ensure interoperability and secure data sharing between parties.

Recommended action matrix for pilot projects

Pilot	Primary leads	Co-responsible partners	Core deliverables	Decision needs
A: Governance framework and data space principles	Kulturens Analyseinstitut (secretariat) + steering group	Digitaliseringsstyrelsen, the Royal Danish Library, DBC Digital, selected industry actors	Steering group with mandate, MVM v1.0, terminology, trusted partner criteria	Terms of reference, standard prioritisation, incentive model
B: ISNI registration and identity matching	The Royal Danish Library, DBC Digital (authoritative IDs)	Kulturens Analyseinstitut, industry actors (e.g. KODA, Gramex, MPO), relevant registers	Matching between local IDs and ISNI, guidelines, lookup service/API	Data responsibility, process for error correction, financing
C: Verified metadata via intermediaries	Technical operator (to be appointed) + Kulturens Analyseinstitut (requirements)	The Royal Danish Library, DBC Digital, technical suppliers, Digital Lead	Data space API solutions, MVM validation, transparency report	Choice of operator, access model, contracts
D: Cultural heritage enrichment (geo/time) and context	The Royal Danish Library, DBC Digital	Kulturens Analyseinstitut, geodata partners, cultural institutions, industry actors	Extended MVM (geo/time), enrichment flow, best practice, pilot search functionality	Field standards, legal basis/consent where relevant

Table 2: Pilot projects connected to suggested stakeholders' roles and deliverables.

Implementation phases

Proof of Concept: limited MVM v1.0 and exchange between a few actors, baseline measurement of errors and double entry, iterations, etc.

Pilot: rollout to more actors, ISNI matching, verified flow via intermediaries, first governance version in operation.

Scaling: extension to additional cultural domains, integration with cultural heritage registers, European interoperability as a design requirement.

Way forward

This report concludes that realising these improvements requires three prerequisites:

1. Political support for shared data standards through clear requirements, policy signals and resources.
2. Cross-sector coordination through governance structures with a mandate for binding decisions.
3. A clear value proposition through the focused pilot projects that document concrete value for artists, organisations, cultural heritage and the public.

The proposed pilot projects should not be developed or implemented separately, nor in isolation from the outside world. The European situation shows that cultural data challenges are structural across countries, and that solutions require coordination.

The CITF report explicitly warns about an unresolved conflict between GDPR's data protection requirements and the Berne Convention's requirement for author attribution. Names and attribution data constitute rights management information protected under EU legislation, but are simultaneously personal data under GDPR. This conflict creates legal uncertainty that undermines both AI governance and fair remuneration, and can only be resolved at the European level.

The OpenMuse consortium also warns about the paradox that strengthened metadata quality increases the visibility of works, but simultaneously increases the risk that AI models can more easily exploit works for imitation and substitution. Therefore, both data spaces and metadata governance are political matters - not only technical organisational forms.

Without coordinated action, increasing release volumes and AI-driven production tools will raise both complexity and the need for verifiable metadata chains. International initiatives such as the EU Cultural Data Space, CITF and the UK's streaming metadata working groups are moving forward.

We therefore recommend that Danish actors actively participate in the following contexts:

1. **Open Music Observatory** – the federated European initiative for music data that builds on the same principles as this report's recommendations.
2. **Copyright Infrastructure Task Force (CITF)** – continued work on interoperable copyright data standards, including in relation to AI use.
3. **European Collaborative Cloud for Cultural Heritage (ECCCH)** – for standardisation of cultural heritage data formats and cultural heritage data sharing.
4. **Danish Data Space Forum** – as the national coordination point for data space initiatives.

The pilot projects should be designed with explicit interoperability with these European initiatives. This is to ensure that Danish cultural heritage and rights are visible and protected in the emerging European cultural data ecosystem, and to counteract Denmark creating yet another isolated national ecosystem.

Danish cultural data hold significant potential. The missing element is not data quality, technical capacity or data volume, but stronger willingness for coordination, governance and consensus around shared reference frameworks and standards.

Appendices

Appendix 1: Cultural data sources for datavejviser.dk

[Link to xlsx-file in table format hosted on Kulturens Analyseinstituts files server.](#) If the file is unavailable, please contact aasmund@kulturanalyser.dk.

Below is an overview of the mapped cultural data sources, including their owner and an assessment of the potential for reuse. (The full annex file contains further details such as identifiers, access methods and notes.)

Cultural data source title	Cultural data source owner	Reuse assessment
MediaStream	The Royal Danish Library (Det Kgl. Bibliotek)	Low
Arkiv.dk / Arkibas	Municipal / association-run	Low
Image and poster archive (Billed og plakatarbivet)	Danish Film Institute (Det Danske Filminstitut)	Low
Denmark on Film (Danmark på Film)	Danish Film Institute (Det Danske Filminstitut)	Low
Danish cultural data (Dansk kulturdata)	Europeana	High
DR archive (DR-arkivet)	The Royal Danish Library (Det Kgl. Bibliotek)	Low
Festival history (Festivalhistorik)	Festivalhistorik.dk	Low
The Film Database (Filmdatabasen)	Danish Film Institute (Det Danske Filminstitut)	Medium
Performance data (Forestillingsdata)	Teaterbilletter.dk	Low
Listed and preservation-worthy buildings (Fredede og Bevaringsværdige Bygninger)	Danish Agency for Culture and Palaces (Slots- og Kulturstyrelsen)	Low
Finds and ancient monuments (Fund og Fortidsminder)	Danish Agency for Culture and Palaces (Slots- og Kulturstyrelsen)	Medium
Art Index Denmark & Weilbach's Artist Lexicon (Kunstindeks Danmark & Weilbachs Kunstnerleksikon)	Danish Agency for Culture and Palaces (Slots- og Kulturstyrelsen)	Low
LOAR	The Royal Danish Library (Det Kgl. Bibliotek)	Medium
Museum collections (Museernes Samlinger)	Danish Agency for Culture and Palaces (Slots- og Kulturstyrelsen)	Low
Netarkivet	The Royal Danish Library (Det Kgl. Bibliotek)	Low
Publications	Kulturens Analyseinstitut	Low
Audience surveys (Publikumsundersøgelser)	Applaus	Low
Collections Online (Samlinger Online)	National Museum of Denmark (Nationalmuseet)	Low
Cultural event calendar (Kultureventkalender)	Kulturnaut	Medium

Table: Cultural data sources mapped in the project and proposed for publication on www.datavejviser.dk (abbreviated overview).

Appendix 2: Primary data types and structures in the music industry

Data types

Data type	Format	Standards	Interoperability
Music rights	XML/API	ISWC, IPI	High (DDEX compatible)
Song	Proprietary	ISRC	Medium
Streaming data	JSON	ISRC, UPC	Varies
Cultural heritage data	MARC21	ISBN, ISSN	High

Table: Data types and interoperability in the music industry.

Metadata fields

Metadata field	Data standard	Use	Harmonisation
Song title	Text string	Universal	Low
Composer	IPI/ISNI	Credit identification	Medium
ISWC	T-XXXXXXXX-X	Work identification	High
ISRC	CC-XXX-YY-NNNNN	Track identification	Very high
Label	IFPI LC	Record label ID	High
Genre	Text string	Categorisation	Very low lav

Table: Minimum metadata fields, standards and harmonisation needs (music domain).

Appendix 3: Data identifiers and data bridges in the music industry

1. Proprietary (non-open) identifier systems

Below is a selection of widely used proprietary identifiers in the music industry. They are often necessary in individual systems, but they become a barrier to interoperability if they are not linked to open standards.

Data identifier (ID)	Used by	Purpose	Format
IPI-nummer (Interested Party Information)	KODA, NCB, PROs (e.g. ASCAP, BMI, PRS)	Identifies songwriters, composers and publishers for performance and mechanical rights	Unique 9–11 digit number
KODA work number	KODA	Internal ID for compositions registered with KODA	Unique per work
NCB work number	NCB	Internal ID for mechanical rights in the Nordic region	Unique per recording
Gramex ID	Gramex	Identifies sound recordings for neighbouring rights in Denmark	Unique per recording
IFPI Label Code (LC)	IFPI	Identifies record labels or publishers that own or administer a sound recording	LC-XXXXX
YouTube Asset ID	YouTube	Identifies audio and video assets for monetisation	Unique per asset
Spotify URI	Spotify	Internal identifier for tracks, albums and artists	spotify:track:xxxxxxxxx

Examples of proprietary identifier systems and their use.

2. Interoperable identifier systems

Interoperable identifiers are global standards that enable linking across systems and actors. Consistent use of these is a key prerequisite for automated settlement and cultural heritage preservation.

Data identifier (ID)	Used by	Purpose	Format
ISWC (International Standard Musical Work Code)	KODA, NCB, PROs (globally)	Identifies musical works (melody and lyrics) globally	T-XXXXXXXX-X
ISRC (International Standard Recording Code)	Gramex, IFPI, streaming services, YouTube Content ID	Identifies sound recordings and tracks globally	CC-XX-YY-NNNNN
UPC / EAN code	NCB, digital distributors (DistroKid, TuneCore, CD Baby, etc.), streaming services	Identifies albums and releases for digital and physical distribution	12–13 digits
ISNI (International Standard Name Identifier)	KODA, NCB, IFPI, PROs, rights agencies	Identifies creators (composers, producers, performers, publishers) across global databases	16 digits
EAN-13 / GTIN (Global Trade Item Number)	Retailers, digital distributors, IFPI	Identifies physical and digital products for sales tracking	13-digit barcode

Table: Examples of interoperable identifiers in the music industry.

3. Data bridges: linking proprietary and interoperable IDs

When proprietary IDs are linked to interoperable standards, “data bridges” are created that enable traceability across the value chain.

Use case	Proprietary ID	Interoperable ID
Songwriter/composer rights	IPI number (KODA)	ISWC
Mechanical rights (physical/digital sales)	NCB work number	ISWC, ISRC
Neighbouring rights (performance/recording)	Gramex Recording ID	ISRC
Streaming monetisation	Spotify URI, YouTube Asset ID	ISRC, UPC
Publisher & composer identification	IFPI Label Code (LC)	ISNI

Table: Examples of data bridges between proprietary and interoperable identifier systems.

Appendix 4: Extended glossary

Practical use

This glossary can be used as:

- Standard translation into a working language (e.g. reports, meetings, applications)
- Reference for consistent terminology
- Starting point for aligning concepts between the cultural sector, data space actors and IT partners

Recommendations

- Use the primary Danish terms consistently.
- Optionally include the English term in parentheses the first time it appears if the target audience works internationally.
- Use alternatives only if the context requires a more precise meaning.

Last updated: February 2026

Data Taxonomy = Danish: Datataksonomi

Alternative translations: Dataklassifikation, Datakategorisering

Definition: A systematic division and classification of data into categories, often hierarchical or thematic.

Use in the cultural domain: Used to structure cultural data across domains (e.g. types of works, types of actors, institutions).

Data Standards = Danish: Datastandarder

Alternative translations: Datakrav, Standardformater

Definition: Common rules and formats for how data must be structured, described and exchanged.

Use in the cultural domain: Crucial for interoperability between cultural institutions and external data partners.

BI Process = Danish: BI-proces (Business Intelligence)

Alternative translations: Analyseproces, Databehandlingsproces

Definition: A process in which data are collected, structured, analysed and turned into decision support.

Use in the cultural domain: Can be used for audience analyses, cultural policy, economic models, etc.

Data Exchange Standards = Danish: Dataudvekslingsstandarder

Alternative translations: Udvekslingsformater, Dataoverførselsstandarder

Definition: Standards that enable automated and consistent data exchange between systems (APIs, XML, JSON, etc.).

Use in the cultural domain: For example DDEX in the music industry or IIIF for image archives.

Data Aggregation = Danish: Dataaggregering

Alternative translations: Datakonsolidering, Sammenstilling af data

Definition: Combining data from multiple sources into one consolidated dataset, often for analysis purposes.

Use in the cultural domain: Used, for example, to combine data about cultural consumption across multiple platforms.

Data Exchange = Danish: Dataudveksling

Alternative translations: Deling af data, Datatransmission

Definition: Transfer or sharing of data between two or more actors/systems.

Use in the cultural domain: The basis for collaboration between cultural institutions, collective management organisations and technology partners.

BI Models = Danish: BI-modeller

Alternative translations: Analysemodeller, Rapportmodeller

Definition: Structured models for turning data into analyses and reports (dashboards, KPI models, etc.).

Use in the cultural domain: For example, cultural economic analyses, impact measurement, and audience segmentation.

Data Literacy = Danish: Datakompetence / Data literacy

Alternative translations: Dataforståelse, Datadannelse

Definition: The ability to read, understand, analyse and use data correctly.

Use in the cultural domain: Central for cultural actors to use data for development, innovation and documentation.

Benchmarking and Surveys = Danish: Benchmarking og surveys / Spørgeskemaundersøgelser

Alternative translations: Sammenligningsanalyser, Spørgeskemaer

Definition: Methods for comparing data or collecting knowledge through questionnaires/surveys.

Use in the cultural domain: For example, to compare audience experiences, institutional performance or cultural habits.

Appendix 5: Interview guide (semi-structured)

Part 1: Introduction and context (5–10 minutes)

Introduction

“Thank you for participating in the DataMosaik1 project. We investigate how cultural data, here specifically how music data, are registered, shared and used in Denmark with a hope to promote better standards and collaboration. The project is carried out in cooperation with the Danish Agency for Digital Government (Digitaliseringsstyrelsen) and aims to expand the number of, or improve the quality of, cultural datasets that are made available on datavejviser.dk.

The interview lasts between 45 and 60 minutes and is conducted as a semi-structured interview with open questions and room for elaboration. We interview key people from the music industry and tech companies connected to the area as suppliers.

Is it okay that we record the interview? It is for internal use only, so that we do not have to write everything down along the way, and the recording will be deleted after transcription. Before we start, I would like to emphasise that it is important to us that you feel safe and comfortable during the conversation. All information you share with us will be treated confidentially and anonymised when we use it in our analysis and report. If at any point during the interview you want us to pause or stop the recording temporarily or completely, please let us know.” (Start recording if relevant.)

Background information

Can you briefly introduce yourself and your role?

- Name, title, organisation
- Responsibilities related to data
- How long have you worked with music data?
-

How would you describe your organisation’s role in the music data ecosystem?

- Primary functions
- Which types of data you work with
- Your position in the value chain
- Age of the organisation

Part 2: Current data practices (10–15 minutes)

Data collection and management

Which types of music data do you collect and maintain?

- Metadata (works, recordings, artists)
- Usage data (plays, sales, streams)
- Rights data
- Other relevant data types

Which systems and tools do you use for data registration?

- Software/platforms
- Manual vs. automated processes
- Integration with other systems

Which identifiers and standards do you use?

- ISRC, ISWC, ISNI, IPI
- Proprietary IDs
- Other standards
- Follow-up: Why these? Do you experience challenges with them?

Part 6: Closing (5 minutes)

Additional input

Are there important aspects we have not touched upon?

- Other challenges
- Opportunities we have overlooked
- Relevant contacts

Do you have concrete examples of situations where better data would have made a difference?

- Successes
- Errors or losses
- Lessons learned

Follow-up

May we contact you again if we have follow-up questions?

Would you like to receive the project results when they are available?

Do you have others in your network we should talk to?

Addendum: niche-specific additional questions

For collective management organisations:

How do you handle cross-border rights?

Which data are you typically missing from rights holders?

How do you view blockchain and other new technologies?

For streaming services:

How do you match content with rights data?

Which metadata are critical for your business?

How do you handle local vs. global standards?

For cultural users:

How do you report your consumption?

With which tools/processes?

Estimated resource use in staff hours?

Estimated budget spend on rights payments?

For cultural institutions:

How do you document your collections digitally?

Which preservation standards do you follow?

How do you make data available to researchers?

For technology suppliers:

Which integrations do your customers request?

How do you see standards developing?

Which technical trends affect music data?

For trade unions:

How do you work with rights management?

What is your assessment of your members' understanding of the area?

How do you work to strengthen your members' understanding of rights management?

For record labels:

Do you deliver legal deposit to the Royal Danish Library (both physical and digital)?

KULTURENS ANALYSEINSTITUT

Kulturens Analyseinstitut
(+45) 69 16 26 69
info@kulturanalyser.dk
Knudsvej 44 C,
4000 Roskilde, Denmark