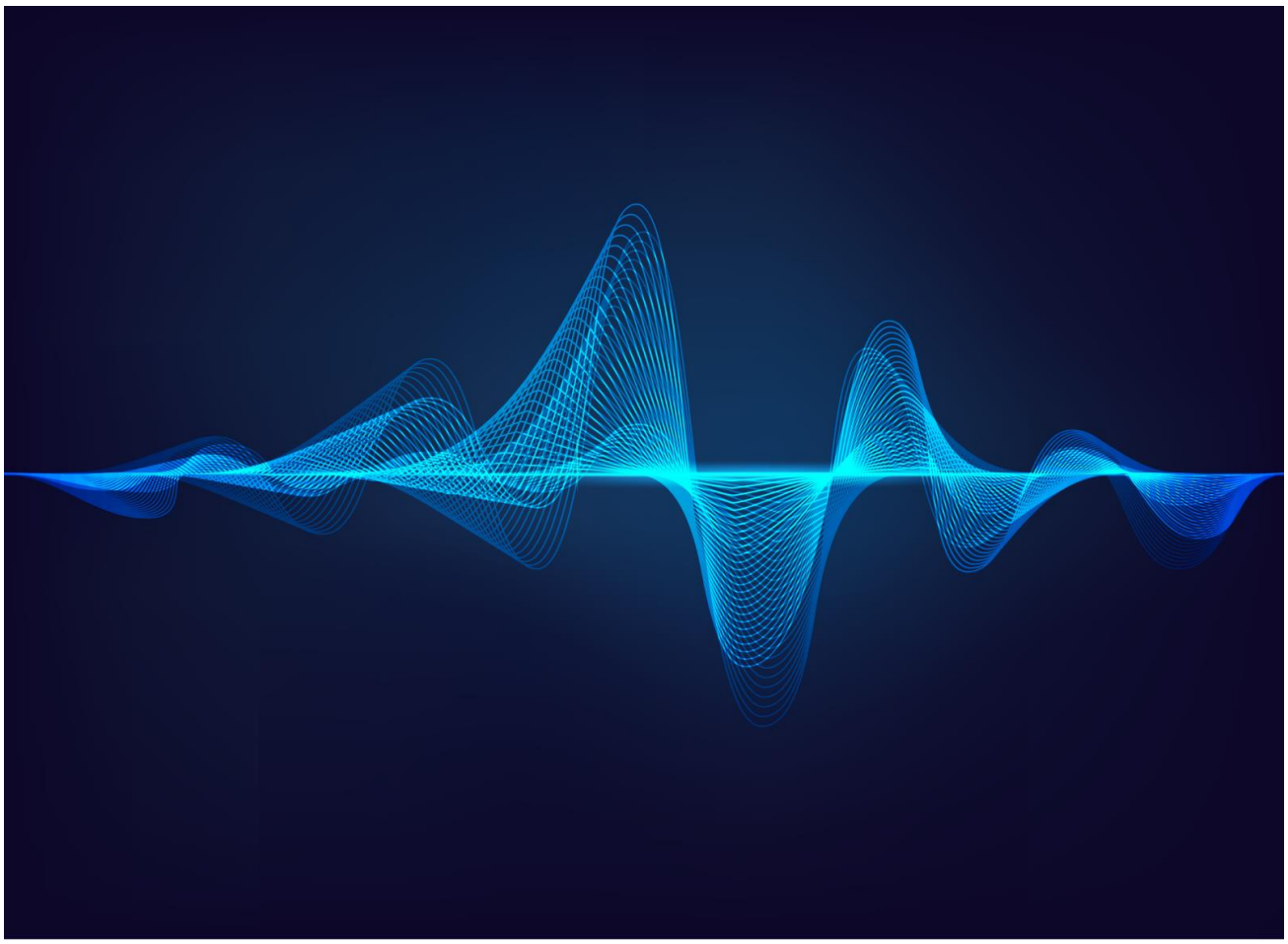
# D5.6 – Policy Brief 1: Music Metadata Mainstreaming and EU Law

Senftleben, M.R.F.; Margoni, T.; Poort, J.P.; Szkalej, K.; Valk, E.G.

**Publication date**
2024
**Document Version**
Final published version



# D5.6 – Policy Brief 1: Music Metadata Mainstreaming and EU Law

## *OpenMusE*

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

Funded by the European Union

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

# Project Factsheet

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

# Deliverable Factsheet

| | |
|---|---|
| Number: | **D5.6** |
| Title: | **Policy Brief 1: Music Metadata Mainstreaming and EU Law** |
| Lead beneficiary: | **UVA** |
| Work package: | **5** |
| Task: | 5.2 |
| Dissemination level: | **Public** |
| Submission date: | **29.03.2024** |
| Main author(s): | **Martin Senftleben, Universiteit van Amsterdam** |
| | **Thomas Margoni, KU Leuven / Universiteit van Amsterdam** |
| | **Joost Poort, Universiteit van Amsterdam** |
| | **Kacper Szkalej, Universiteit van Amsterdam** |
| | **Etienne Valk, Universiteit van Amsterdam** |
| Contributor(s): | |
| Quality check: | **Caterina Sganga, Scuola Superiore Sant'Anna** |
| | **Eszter Kabai, Artisjus** |

Document history:

| Revision | Date | Main modification | Author |
|---|---|---|---|
| 0.9 | 20/03/2024 | Submitted to reviewers | See above |
| 1.0 | 29/03/2024 | Submitted to EC | See above |

## Disclaimer of Warranties

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

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

Any dissemination of results must indicate that it reflects only the author's view and that the Commission Agency is not responsible for any use that may be made of the information it contains.

Neither Project Coordinator, nor any signatory party of OpenMusE Project Consortium Agreement, nor any person acting on behalf of any of them:

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

(b) assumes responsibility for any damages or other liability whatsoever (including any consequential damages, even if Project Coordinator or any representative of a signatory party of the OpenMusE Project Consortium Agreement, has been advised of the possibility of such damages) resulting from your selection or use of this document or any information, apparatus, method, process, or similar item disclosed in this document.

# Glossary

AI Act Artificial Intelligence Act

CDSMD Directive on Copyright in the Digital Single Market

CEDS Common European Data Spaces

CISAC International Confederation of Societies of Authors and Composers

CMO Collective Management Organisation

DA Data Act

DDL Data and Digital Legislation

DGA Data Governance Act

DNN Deep neural networks

DOI Digital Object Identifier

EIDR Entertainment Identifier Registry

GRD Global repertoire database

IPI Interested Party Information Number

ISAN International Standard Audiovisual Number

ISBN International Standard Book Number

ISMN International Standard Music Number

ISNI International Standard Name Identifier

ISRC International Standard Recording Code

ISSN International Standard Serial Number

ISWC International Standard Work Code

ML Machine learning

MLC Mechanical Licensing Collective

MMA Music Modernisation Act

NLP Natural language processing

ODD Open Data Directive

TDM Text and data mining

WIPO World Intellectual Property Organisation

# Table of contents

# Executive Summary

In order to enable composers, performers and the music industry to benefit from licensing opportunities in the field of new technologies, such as AI training, it is important to establish a comprehensive music metadata infrastructure that improves the visibility and accessibility of the European music repertoire in digital and algorithmic environments. Recognizing the need for metadata improvement, various European initiatives aim to increase awareness among artists and rightholders, and to build bridges between existing metadata collections and infrastructures. One central factor in the equation, however, has remained underexplored and underused to this day: despite the prohibition of formalities in the Berne Convention, it is conceivable to employ legal mechanisms, such as the notification of work-related information under Article 17(4)(b) of Directive 2019/790 on Copyright in the Digital Single Market, the opt-out mechanism relating to text and data mining that follows from Article 4(3) of the same Directive, and the EU rules on collective rights management, as well as the broader legal framework applicable to data spaces as vehicles to impose an obligation on rightholders to constantly provide updated music metadata in standardised form. If information stemming from these channels is pooled, the resulting accumulation of EU copyright data could lead to a promising reservoir of music metadata that is capable of enhancing and boosting licensing opportunities.

# 1. Introduction

In order to enable composers, performers and the music industry to benefit from licensing opportunities in the field of new technologies, such as AI training, it is important to establish a comprehensive music metadata infrastructure that improves the visibility and accessibility of the European music repertoire in digital and algorithmic environments. Recognizing the need for metadata improvement, various European initiatives aim to increase awareness among artists and rightholders, and to build bridges between existing metadata collections and infrastructures. One central factor in the equation, however, has remained underexplored and underused to this day: despite the prohibition of formalities in the Berne Convention, it is conceivable to employ legal mechanisms, such as the notification of work-related information under Article 17(4)(b) of Directive 2019/790 on Copyright in the Digital Single Market, the opt-out mechanism relating to text and data mining that follows from Article 4(3) of the same Directive, and the EU rules on collective rights management, as well as the broader legal framework applicable to data spaces as vehicles to impose an obligation on rightholders to constantly provide updated music metadata in standardised form. If information stemming from these channels is pooled, the resulting accumulation of EU copyright data could lead to a promising reservoir of music metadata that is capable of enhancing and boosting licensing opportunities.

This first policy brief build on work done by the University of Amsterdam in the context of Task 1.1. It benefits from valuable insights that have been gathered during a hybrid workshop entitled “Metadata mainstreaming in the music industries. Data spaces and beyond”, and in Amsterdam on 14 March 2024, involving academics, policy makers and stakeholders. We would like to express our gratitude to all presenters and participants of this workshop for their active contributions.

# 2. Inadequate data infrastructure

The problem of insufficient metadata quality in the field of music – and literary and artistic works more broadly – is not new. Missing, inaccurate or non-interoperable metadata can lead to a lack of adequate information on protected contributions, current rightholders and work- or production-related characteristics (title, genre, content, year of creation, etc.). In monetary terms, missing or inaccurate metadata can lead to foregone licencing opportunities, as works are not found or rightholders cannot be identified. As the current discussion on disruptive effects of generative AI systems and the plea for equitable remuneration for the use of human musical works in AI training processes shows,[1] new technological developments further enhance the need for adequate data solutions in the area of licensing.[2] In economic terms, this causes welfare losses both for rightholders and potential users of musical works, and for the wider public. Second, it can increase search costs (which entail economic losses as well) and costs associated with data cleaning and correction. Third, it can lead to instances in which musical works are used while rightholders do not receive the remunenation that is due for the reason that they cannot be identified. If an overall music database is capable of serving as a one-stop shop for acquiring rights and paying royalties, it could boost the exploitation of protected material in unexpected ways and bring increased revenue to authors, performers and the creative industry.

Missing and inaccurate metadata can also have more indirect and subtle effects that can be both economic and cultural in nature. It can for instance cause biases is recommender systems for music streaming platforms, and can favour repertoires from largers markets or language areas over those of smaller ones, as a consequence of economies of scale associated with larger repertoires. A lack of interoperability of metadata systems can also favour large repertoires or platforms over smaller ones.[3]

In the music segment of the creative industries, there are several well-known examples of existing data infrastructures, such as the Common Information System (CIS) of the International Confederation of Societies of Authors and Composers (CISAC). With its various nodes in several regions of the world, the CIS-Net system and its associated standards represent a global tool to facilitate music licensing and revenue distribution.[4] In terms of music data standardisation, the music publishing industry's

[1] European Composer and Songwriter Alliance/European Writers' Council et al. (2023), *Joint Statement from Authors' and Performers' Organisations on Artificial Intelligence and the AI Act – True Culture Needs Originals: Transparency and Consent are Key to the Ethical Use of AI*, available at: https://screendirectors.eu/joint-statement-from-authors-and-performers-organisations-on-artificial-intelligence-and-the-ai-act/; Initiative Urheberrecht (2023), *Joint Statement: Authors and Performers Call for Safeguards Around Generative AI in the European AI Act*, 19 April 2023, available at: https://urheber.info/diskurs/call-for-safeguards-around-generative-ai; European Guild for Artificial Intelligence Regulation (2023), *Manifesto for AI Companies Regulation in Europe*, available at: https://www.egair.eu/#manifesto (last visited on 29 March 2024).

[2] Cf. M.R.F. Senftleben, 'Generative AI and Author Remuneration', *International Review of Intellectual Property and Competition Law* 54 (2023), 1535-1560; Geiger C, Iaia V (2023), 'The forgotten creator: towards a statutory remuneration right for machine learning of generative AI', *Computer Law and Security Review* (forthcoming). https://ssrn.com/abstract=4594873 (last visited on 29 March 2024); Frosio G (2024), 'Should we ban generative AI, incentivise it or make it a medium for inclusive creativity?', In: Bonadio E, Sganga C (eds.) *A research agenda for EU copyright law*. Edward Elgar: Cheltenham 2024 (forthcoming). https://ssrn.com/abstract=4527461 (last visited on 29 March 2024).

[3] See for a more detailed discussion of these perspectives Senftleben/Margoni/Antal et al. (2022). See also: European Commission, Directorate-General for Communications Networks, Content and Technology, (2022). *Study on copyright and new technologies : copyright data management and artificial intelligence*, Publications Office of the European Union. https://data.europa.eu/doi/10.2759/570559, Section 2.3.

[4] See https://www.cisac.org/services/information-services/cis-net.

*International Standard Work Code* (ISWC),[5] the recording industry's *International Standard Recording Code* (ISRC), the *Interested Party Information Number* (IPI) and the *International Standard Name Identifier* (ISNI) continue to be prime examples of existing initiatives aimed at enabling the exchange of accurate data to identify repertoire or reduce transaction costs associated with the processing of licensing agreements.

At the same time, these examples highlight data deficiencies and interoperability problems resulting from different metadata sets and different approaches to identifying and verifying data. To date, initiatives to harmonise ISWC and ISRC metadata and integrate them into an overarching, comprehensive database have failed. In the EU, former Commissioner Neelie Kroes set up a working group in 2008 to explore the possibilities of establishing a global repertoire database (GRD). The participants of the working group, which included producers, collecting societies and distribution platforms, did come up with recommendations on how to proceed.[6] Ultimately, however, the project was buried in 2014.[7] Other unsuccessful attempts were the *International Music Joint Venture* from 2000, which was founded by several collecting societies in Europe and North America, and a project initiated by the World Intellectual Property Organisation (WIPO) in 2011, which aimed to establish a joint rights database and also did not produce any concrete results.[8] Later, in 2020, the WIPO PROOF service was launched. The digital business service was meant to provide a date- and time-stamped digital fingerprint of any file, proving its existence at a specific point in time.[9] However, the service was formally discontinued on February 1, 2022. The WIPO Secretariat has explained to member states that "since the initial feasibility studies, the market has evolved quickly, driven by the accelerated digitalization".[10]

These unsuccessful attempts already show that – despite existing metadata infrastructures such as the CIS-Net system and the ISWC/ISRC standards – there is a need in the European music industry to combine work- and rights-based databases to a much greater extent in order to ultimately create overarching licensing platforms.[11] Recent initiatives point in the same direction. The *Technical Online Working Group Europe* (TOWGE), for example, brings together a large group of European collecting societies, music publishers and rights agencies to develop a digital royalty processing system.[12] An initiative with similar goals was taken by the Finnish collecting society Teosto. The collaboration

[5] The ISWC was developed by CISAC in collaboration with the International Organisation for Standardisation (ISO) as "a unique, permanent, and internally recognised reference number for the identification of musical works". Another example of an identification system is the GRiD (Global Release Identifier) developed by IFPI. See Katz, Journal of Competition Law and Economics 1 (2005), 276.

[6] Cf. Isherwood, Global Repertoire Database, presentation at the WIPO Meeting wipo_cr_doc_ge_11, 13-14 October 2011, Geneva: WIPO 2011, available at: https://www.wipo.int/meetings/en/2011/wipo_cr_doc_ge_11/prov_program.html.

[7] See Resnikoff, Digital Music News, 10 July 2014, available at: https://www.digitalmusicnews.com/2014/07/10/global-repertoire-database-declared-global-failure/; Schwemer, Licensing and Access to Content in the European Union. In Licensing and Access to Content in the European Union: Regulation between Copyright and Competition Law, 2019, 68-73.

[8] Schwemer, id., 69-70.

[9] See https://www.wipo.int/wipoproof/en.

[10] See https://www.wipo.int/wipoproof/en/news/2021/news_0003.html.

[11] See Gronau/Schaefer, EIPR 2021, 488-494; Schaefer, Kluwer Copyright Blog, 27 November 2020, available at: http://copyrightblog.kluweriplaw.com/2020/11/27/why-metadata-matter-for-the-future-of-copyright/ for a further discussion of the importance of data integration and harmonisation initiatives in the EU. See also Lyons/Sun/Collopy et al, Music 2025 - The Music Data Dilemma: issues facing the music industry in improving data management, 2019, available at: https://www.gov.uk/government/publications/music-2025-the-music-data-dilemma, 34.

[12] See https://www.digitalmusicnews.com/2019/07/26/towge-digital-royalty-group/.

between Teosto and the start-up *Mind Your Rights* has resulted in the platform "Concertify", which aims to provide an efficient and transparent system for cross-border copyright licensing in addition to existing industry structures. Concertify enables artists, rightholders, including collecting societies, music publishers and event organisers, to collaborate and transmit information directly using specific modules, such as a module for the transmission of setlists. With the support of the Slovak Arts Council, a collaboration between the collecting society SOZA and various stakeholders from the music industry has led to the creation of a prototype for a comprehensive database and metadatabase of the Slovak music repertoire. On this basis, the consortium created the prototype of the "Listen Local" recommendation system, which meets the requirements of the recommendations on the topic of trustworthy artificial intelligence (AI) of the European *High-Level Expert Group on Artificial Intelligence.*[13] The accompanying feasibility study showed and quantified the problems arising from incomplete copyright data in existing databases and commercial AI solutions. It estimates that at least 15% of Slovak, Estonian, Hungarian and Dutch works are difficult to exploit due to data problems.[14]

Another recent initiative is the DDEX project, ran by standards-setting organisation Digital Data Exchange LLC. The purpose of DDEX is, in short, to develop standards relating to metadata creation and management in the overall digital music supply chain. One of the DDEX standards, DDEX RDR, describes in detail the required metadata for managing international performance rights. RDx, which stands for Repertoire Data Exchange, puts the DDEX RDR standard into practice since its full operationality in 2020. RDx is a data exchange hub where participating entities can send and receive DDEX RDR data.[15] Somewhat similarly, copyright platform Cube aims to deliver a highly automated copyright system. The platform was commissioned by ICE Services, a joint venture between German, Swedish and UK Performing Rights Organisations.[16]

Other sectors of the creative industries face similar data issues and have also taken initiatives to improve, harmonise and merge data. In the field of book publishing, industry initiatives looking at the establishment of various e-book platforms and catalogues play an important role. Another example is the *Entertainment Identifier Registry* (EIDR): a universal labelling system for film and television data based on DOI technology.[17] Examples of data standardisation initiatives include the *International Standard Book Number* (ISBN), the *International Standard Serial Number* (ISSN) for periodicals, the *International Standard Music Number* (ISMN) for notated music and the *International Standard Audiovisual Number* (ISAN) for audiovisual works. Furthermore, in the field of books, e-books and serial volumes, the standardisation work of the international EDItEUR group is noteworthy.[18] This initiative has led to the ONIX family of standards.[19] With regard to the digital environment, the International DOI Foundation offers the *Digital Object Identifier* (DOI) services and related registration facilities already mentioned: a technical and social infrastructure for the registration and use of persistent interoperable

[13] See https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai and https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.
[14] Antal, Feasibility Study On Promoting Slovak Music In Slovakia & Abroad, 2020, available at: https://reprex.nl/publication/listen_local_2020/.
[15] See https://www.ifpi.org/rdx-recording-industrys-new-data-exchange-service-now-fully-operational/ and https://www.rdx-portal.org/.
[16] See: https://www.iceservices.com/innovation/cube/ and https://www.iceservices.com/about/?_gl=1*xyff4r*_up*MQ..*_ga*MTU5ODY1ODAyLjE3MTEzNjY4NDg.*_ga_QSZEMWD5QD*MTcxMTM3MDMzMS4yLjAuMTcxMTM3MDMzMS4wLjAuMA.
[17] See https://www.eidr.org/.
[18] See https://www.editeur.org/2/About/#Intro.
[19] See https://www.editeur.org/8/ONIX/.

identifiers for use on digital networks, including identifiers for literary and artistic works.[20] Furthermore, relevant within the digital environment is the cross-sectoral initiative for the *International Standard Content Code* (ISCC), which is currently still under development at the *International Organization for Standardization* (ISO).[21] It will be an identification system for digital assets (such as encodings of text, images, audio, video or other content across all media-sectors). The ISCC should serve as a similarity-preserving fingerprint designed to identify digital content in decentralized and networked environments spanning the creative industries.[22]

In the field of visual arts, the *Visual Arts Council* of CISAC has expanded its original work on resale rights and established an online licensing platform under the umbrella of the *International Council of Creators of Graphic, Plastic and Photographic Arts* (CIAGP).[23][24] OnLineArt (OLA) is a one-stop shop for acquiring licences for the online use of works of visual art worldwide, which currently includes works by 60,000 artists.[25] While existing initiatives in the field of visual arts – particularly digitisation initiatives by museums and other cultural heritage institutions – have significantly expanded data collection of visual art works, the situation in the field of photography and illustration is far less transparent.[26] Commercial owners of large fine art libraries, such as Getty Images, can consistently point to existing data management tools. However, the cost of documenting a large number of individual works can quickly become prohibitive for smaller providers of photographic works and illustrations, given the low average value of individual works from the overall collection.[27] Specifically in the field of photo metadata documentation, the IPTC Photo Metadata Standard is widely used and recognized.[28] Despite its prominence, however, it still poses challenges for reliably distinguishing authentic media assets.[29] Compared to the status quo already achieved in the music sector, the process of harmonising, linking and bundling work- and rights-related data in the visual arts sector still seems to be in its infancy.

# 3. Current developments

The brief – by no means exhaustive –outline of existing initiatives in the previous chapter clearly shows that the problem of inadequate data quality in the creative industries, including music metadata, remains complex and unresolved. At the same time, there is an urgent need for improvement. Current developments in the increasingly digital and algorithmic information society make it more important than ever to take a fresh look at the problem and develop stronger regulatory support for music metadata initiatives.

[20] See https://www.doi.org/.
[21] See https://www.iso.org/standard/77899.html.
[22] See https://core.iscc.codes/.
[23] See https://www.cisac.org/services/creator-relations/ciagp-technical-operations.
[24] See http://www.ciagp.org/.
[25] See https://onlineart.info/.
[26] For a more detailed analysis of the specific situation and dynamics in the field of visual arts, see the study by Azzi/El Hage, Les métadonnées liées aux images fixes, 2021.
[27] On this investment dilemma, see Posner, John Marshall Review of Intellectual Property Law 4 (2005), 325.
[28] See study by Mo/Kang et al, Towards Trustworthy Digital Media In The Aigc Era: An Introduction To The Upcoming IsoJpegTrust Standard, 2023 and https://iptc.org/standards/photo-metadata/quick-guide-to-iptc-photo-metadata-and-google-images/.
[29] See Mo/Kang et al and Dang-Nguyen/Sjøen et al, Practical Analyses of How Common Social Media Platforms and Photo Storage Services Handle Uploaded Images, 2023.

First, there is the issue of worldwide competition for data hegemony. While European initiatives, as described, have not yet led to an overarching, comprehensive data infrastructure for the individual branches of the creative sector, the creation of a comprehensive database has succeeded in the US, at least in the area of the music industry. The US initiative goes back to the *Music Modernisation Act* (MMA) passed in 2018.[30] Title I of the MMA establishes the *Mechanical Licensing Collective* (MLC) as a one-stop shop for music licensing. For the proper functioning of this new licensing body, a comprehensive database of music rights is indispensable. The MLC ultimately achieved this goal on the basis of close cooperation with major providers of music streaming services, in particular Apple and Spotify.[31] The new licensing hub has managed to provide a US-wide platform for royalty administration, enforcement and processing as of 1 January 2021.[32] With regard to the present inquiry - focusing on the strategic use of legal instruments to improve the copyright data infrastructure – it is of particular interest that these developments in the US can be traced back to a legislative intervention, namely the adoption of the MMA in 2018.

From a European perspective, the creation of this infrastructure once again raises the question of whether large-scale bundling of music metadata can succeed. Otherwise, there may be a threat of other large-scale structures, such as the MLC system based on Apple and Spotify data, expanding to the European continent. Such a loss of data autonomy could in turn disadvantage European countries' repertoires that only play a subordinate role in the concert of major streaming service providers such as Apple and Spotify. In contrast, an independent European initiative could give small repertoires sufficient space.

Second, technical developments and related licensing opportunities increase the need to create an overarching data infrastructure. The aforementioned training of generative AI systems, for example, is still largely reliant on extensive use of human source material that allows the systems to analyse the parameters of musical works. To create licensing opportunities not only for big repertoire holders[33] but also for smaller (country) repertoires, it is imperative to make music metadata – providing content-related and rights clearance information – available in a harmonised and interoperable format.[34] Without machine-readable literary and artistic input from flesh-and-blood authors, an AI system has no template for algorithmic processes to emulate human creativity. Modern data-driven AI often uses text and data mining (TDM) techniques to obtain the data needed for machine learning.

TDM has emerged as one of the most powerful digital tools in the AI environment for extracting patterns, correlations and hidden knowledge from existing content and data.[35] Techniques currently

[30] See House Reports 1551, Public Law 115-264, dated 11 October 2018.
[31] See https://appleworld.today/apple-spotify-to-fund-new-music-royalties-collective/.
[32] See https://blog.themlc.com/press/mechanical-licensing-collective-begins-full-operations-envisioned-music-modernization-act. With respect to the underlying preparatory work, see further U.S. Copyright Office Library of Congress, MLC Comments in Reply to the Designation Proposal of the American Music Licensing Collective, Inc, Docket No. 2018-11, 21, available at: https://bw-98d8a23fd60826a2a474c5b4f5811707-bwcore.s3.amazonaws.com/photos/Proposed_MLC_-_Reply_Comments.pdf.
[33] For an example of an existing licensing success at big repertoire level, see the agreement concluded between Universal Music and Google/YouTube, as described by A. Nicolaou/M. Murgia, 'Google and Universal Music negotiate deal over AI "deepfakes"', Financial Times, 8 August 2023, available at: https://www.ft.com/content/6f022306-2f83-4da7-8066-51386e8fe63b (last visited on 29 March 2024).
[34] Cf. Hugenholtz/Quintais, IIC 2021, 1190 (1212-1213); Senftleben/Buijtelaar, EIPR 2020, 797 (804-808); Burk, Houston Law Review 58 (2020), 263 (270-321); Ginsburg/Budiardjo, Berkeley Technology Law Journal 34 (2019), 343 (395-396); Janssens/Gotzen, Auteurs en Media 2018-2019, 323 (325-327); Pearlman, Richmond Journal of Law and Technology 24 (2018), 1 (4).
[35] See Margoni, in: Calboli/Montagnani, Handbook on Intellectual Property Research, 2021, 487-505.

discussed under the terms machine learning (ML), natural language processing (NLP) and deep neural networks (DNN) require training AI systems on vast amounts of content and data. Required training information is often extracted by automated machine reading techniques from books, journal articles, musical works or films that enjoy copyright protection. It is therefore not surprising that the insatiable appetite of "creative" AI systems for literary and artistic data is often seen as a threat to, but also as a promising new source of revenue for, the creative industries.[36] However, if the creative industries in Europe fail to provide licences for large-scale AI applications at low transaction costs and at the required large scale, the risk arises that AI training projects will take place in other regions, such as the US. If the necessary content and data can be purchased centrally there on a larger scale and with less administrative effort, it is foreseeable that attractive licensing revenues from TDM could be lost to the European creative industries. The creation of an overarching data infrastructure is therefore also desirable from this point of view.

# 4. Use of existing legal mechanisms for metadata improvement

Whilst concerns over improving metadata raise challenges that are inevitably technical in nature, legislation has the potential of streamlining the improvement of music metadata ecosystems. Several existing legal mechanisms provide important reference points that can be used as a basis for this.

## 4.1. Legal protection of copyright metadata

Protecting information that describes attributes of content in the digital environment can act as an incentive to operationalise electronic management of rights and promote interoperability in different rights management ecosystems in the same way as copyright protection incentivises creative activity and investments in the sector. Since 2001 European copyright law provides such protection in the Information Society Directive. Article 7 of the Directive introduces protection against the manipulation of what it calls *rights management information* (RMI), defined as any information provided by rightholders that identifies protected subject matter, the author or any other rightholder, information about the terms and conditions of use of protected subject-matter, or any numbers or codes that represent such information.[37] More specifically, the provision prohibits unauthorised *removal* and *alteration* of RMI and distribution or other making available of copies in relation to which such information has been (knowingly) removed or altered without authority, when the person knows or has reasonable grounds to know that these acts induce, enable, facilitate or conceal copyright infringement.[38] As the provision is based on Article 12 WIPO Copyright Treaty (WCT) and Article 19 WIPO Performances and Phonograms Treaty (WPPT) of 1996, collectively known as the WIPO Internet Treaties, such or similar protection exists also in many countries outside of the EU.

Under the European provision, for RMI to be protected no particular form is imposed except the requirement in the definition of RMI that the information should originate from the rightholder ("information provided by rightholders"). Moreover, information becomes RMI already when any of

[36] See Covington/Adams/Sargin, in: Proceedings of the 10th Acm Conference on Recommender Systems, RecSys '16, 2016, 191-198, available at: https://doi.org/10.1145/2959100.2959190; Jacobson/Murali/Newett et al, in: Proceedings of the 10th Acm Conference on Recommender Systems, RecSys '16, 2016, 373, available at: https://doi.org/10.1145/2959100.2959120.
[37] InfoSoc Directive, Art. 7(2).
[38] InfoSoc Directive, Art. 7(1).

the items of information is *associated with* a copy rather than only when it appears in connection with making it available, for example on a website or service.[39] Thus it does not matter if RMI is embedded in the actual electronic file containing the work, is supplied together with a file or is retrieved from a central database during transmission.[40] According to some commentators, the information need not even be accessible to the user.[41] The provision is designed to protect, however, only copyright-relevant information that identifies the work, author/rightholder, or the terms of use. Therefore, information that identifies the user or users' consumption patterns does not as such constitute RMI in the sense of Article 7 InfoSoc Directive, even though the Directive does recognise that such "user metadata" can be processed by rights-information systems.[42] This circumscription does not mean, however, that RMI, especially terms of use, cannot be user specific.

RMI protection as a supplementary layer of protection in the copyright system has received relatively little attention as a separate object of inquiry and the exact parameters of the provision remain largely unexplored, having remained mostly in the shadow of the sister provision targeting circumvention of DRM systems (technological protection measures) in Article 6 InfoSoc Directive, and the international equivalents in Articles 11 WCT and 18 WPPT.[43] As the direct impulse for both provisions is the fear that unlawful activities may be carried out by users of copies,[44] they are both often conceived of as being relevant only after a copy of the work has left the rightholder and is accessible to the user, protecting against tampering with the technology and/or information to either obtain unlawful access or share unprotected copies. Indeed, liability is conditioned on knowledge (or reasonable grounds to know) that the activities induce, enable, facilitate or conceal copyright infringement. However, these are very broad concepts that may additionally be governed by national conceptions of secondary liability, i.e. contribution to copyright infringement, which may extend over a very broad set of activities that are not necessarily in direct proximity to the infringement. With indications in case law that breach of a copyright licence can constitute copyright infringement,[45] removal or alteration of RMI from a copy or a database concerning terms of use that can cause a bona fide user to technically carry out such contractual breach could even be perceived as facilitating copyright infringement.[46] Importantly, actual infringement of copyright somewhere on the content distribution chain is not a prerequisite for liability to ensue under the RMI protection regime.

RMI protection operates as a form of data protection, designed to preserve the integrity of information associated with a copy. At its core it supplements efficient management of rights and contracting in the

[39] InfoSoc Directive, Art. 7(2).
[40] Geiger, Schönherr, Stamatoudi, Torremans, Karapapa (2021), The Information Society Directive in Stamatoudi/Torremans (eds), *EU Copyright Law: A Commentary* (2nd ed, Edward Elagar), p. 361
[41] Bechtold (2016), Information Society Directive in Dreier/Hugenholtz (eds), *Concise European Copyright Law* (Kluwer Law International), p. 483.
[42] InfoSoc Directive, recital 57.
[43] See however Perry, 'The Protection of Rights Management Information: modernization or cup half full?' in Geist (ed.), *From "Radical Extremism" to "Balanced Copyright": Canadian Copyright and the Digital Agenda* (Irwin Law, 2010), 304-326, and more recently Wilkinson, 'Is protection of data through data exclusivity, technological protection measures or rights management information actually intellectual property?' in Gervais (ed.), *The Future of Intellectual Property* (Edward Elgar 2021), 169-192; Cadavid, *The Origin and Purpose of Legal Protection for the Integrity of Copyright Metadata* (2023) IIC 54 1179-1202.
[44] InfoSoc Directive, recital 56 and 47 respectively.
[45] C-666/18 *IT Development SAS* v *Free Mobile SAS*.
[46] Similarly, Geiger, Schönherr, Stamatoudi, Torremans, Karapapa (2021), p. 362 assessing that "One induces, enables, facilitates or conceals an infringement when he performs any act, which leads a third party in good faith to further use/license/exploit a work on the basis of the information carried by it and which he trusts as being the authentic".

online environment as confirmed by recital 55 and is a tool for safeguarding moral rights of creators.[47] Whilst the protection regime is designed as a general liability rule, and should therefore not be conceived of as a property right over data as such, it follows from Article 7(1) that removal, alteration and distribution of altered copies is prohibited *if it is done without authority*, which suggests that when such authority exists, liability cannot not ensue even if other requirements are satisfied. While such a "fail-safe" mechanism supersedes any differences to and national conceptions of liability, and prevents absurd outcomes from taking place, such as preventing rightholders themselves from becoming liable for removal or alternation of RMI, it also acts as an incentive to obtain authorisation by anyone who wants or needs to remove or alter RMI. For this very reason, alteration of RMI, the specific form it takes or what it is intended to include for each relevant copy, can become an object of contractual bargaining and potentially facilitate endeavours to realise an interoperable infrastructure.

In view of these qualities, dispelling Article 7 InfoSoc Directive because it supposedly is relevant only at the end of the content distribution chain seems impertinent. As rights management becomes an increasingly complex business, the need for preserving and maintaining the integrity of information about copies becomes increasingly more important, and so does legal protection of RMI. The definition of RMI confirms that it is intended to be a flexible and mutable concept designed to incorporate changes in ownership throughout the life cycle of content and metadata that is project-specific, very much supporting activities at the beginning of the content distribution chain where rights management commences and exploitation strategies are designed. Having regard to the variety of arrangements, project-specific RMI could include information and terms that are relevant for specific purposes or territories,[48] categories of users, or categories of licences. This breadth also means that the notion of rightholder, i.e. the source of the information that becomes RMI, potentially is broad and includes not only the actual creator but likely also anyone else that legitimately can supply the information, whether on the basis of default copyright rules (statutory ownership of rights) or on the basis of a contract or rights management mandate (or alternatively render information supplied by the latter as having been done with authority and therefore not contrary to RMI protection). This is particularly relevant for downstream uses where a commercial licensee, such as an online music service, often determines the specific terms and conditions of use of a licensed copy and applies those terms on the service, but may be as relevant in case of collective rights management and in particular in case of registering and effecting extended collective licence (ECL) opt outs.

Although use of RMI is voluntary, an evident shortcoming is that the framework lacks a right of access to the information or a duty to maintain correct RMI in the event that it is used. Arguably, this may originate from early conceptions of the digital market as it was imagined in the 1990s and assumptions about market dynamics, such as alignment of interests on the supply end of the content distribution chain (denoted sometimes as a harmony of interests), rather than being a careful policy choice to exclude creators or their representatives from being able to view and rectify information. But as we shall see below, such a right does exist in the narrow context of collective rights management.

## 4.2. Technical and administrative capacity to create and maintain the accuracy of the repertoire

[47] Most obviously the right to be named as author, protected by Article 6$^{bis}$ Berne Convention and incorporated into the WCT framework by Article 1(4) WCT.
[48] Arguably regional coding of DVDs constitutes such territory-specific RMI.

Adopted in 2014, the CRM Directive[49] imposes legal obligations that create a harmonised approach to copyright metadata in a narrow but essential copyright sector, namely collective rights management. The Directive contains general rules on governance, transparency, and accountability for the collective management of copyright and creates a legal framework for the development of multi-territory licensing by collective rights management organisations (CMOs).[50] A great portion of this framework focuses therefore on general aspects of operating a CMO, however several provisions address directly concerns over the ability to aggregate information about a global repertoire and metadata in a cross-border environment. In a nutshell, these provisions are designed to address the capacity of CMOs to electronically identify works, rightholders, uses, and accuracy of the repertoire in the context of multi-territorial licensing of online rights.[51] Additionally, the Directive establishes an interface between CMOs and online service providers.

Overarchingly, pursuant to Article 24(1) CMOs which grant multi-territorial licenses must have capacity to efficiently and transparently process electronically data necessary for administration of such licences, such as being able to identify the repertoire and monitor its use. In other words, CMOs are required to have a technical and administrative capacity to create and maintain the accuracy of their repertoire. Article 24(2) itemises, however, concrete features that cumulatively amount to such capacity as a minimum.

As a first set of obligations (Article(24)(a)-(c)), CMOs are required to have the ability to *accurately identify* musical works, *wholly* or *in part*, and to *accurately identify*, *wholly or in part*, with respect to *each relevant territory*, the *rights* and their corresponding *rightholders* for *each musical work* or *share* therein, which the CMO is authorised to represent. This is not limited to the CMOs own repertoire as such representation may originate from representation agreements between CMOs.[52] Moreover, CMOs must make use of *unique identifiers* in order to *identify* rightholders and musical works, taking into account, as far as possible, *voluntary industry standards and practices* developed at international or Union level. These obligations are complemented by corresponding prerogatives in Article 26(2), pursuant to which CMOs are required to provide *rightholders* whose musical works are included in its own music repertoire, as well as those represented on the basis of a representation agreement between CMOs (Article 26(3)), the means of submitting to it electronically information concerning their *musical works*, their *rights* in those works and the *territories* in respect of which the rightholders *authorise* the organisation. Also in this respect, CMO and rightholders are expected to take into account to the extent possible *voluntary industry standards or practices* regarding the exchange of data developed at international or Union level, so that rightholders can specify the musical work, wholly or in part, the online rights, wholly or in part, and the territories in respect of which they authorise the organisation. Needless to say, the Directive does not oblige CMOs to develop any industry standards or practices as such. However, CMOs are certainly envisaged to be that actor that relies on such standards and practices, which implies that they ought to follow developments and, to the extent possible, adapt their technical and administrative capacity to any applicable industry standard or practice. But even if the legal framework currently assumes that the market will figure it out, such an expectation may, nevertheless, incentivise CMOs to undertake endeavours to establish such standards or practices, if

[49] 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 uses in the internal market.
[50] CRM Directive, recitals 8-9, 40.
[51] Defined under Article 3(m) CRM Directive as licences which cover the territory of more than one Member State.
[52] See CRM Directive Articles 29-30.

only to make it easier for themselves. After all, as inevitable intermediaries between rightholders and services, CMOs are best placed to observe issues at both ends of the music ecosystem.

As a second set of obligations, CMOs are under a duty to monitor the metadata record and ensure its accuracy. As a component of their envisaged technical and administrative capacity under Article 24, CMOs must be able to identify and resolve in a timely and effective manner inconsistencies in data held by other CMOs granting multi-territorial licences for online rights in musical works (24(2)(d)). Admittedly, this requirement merely imposes an obligation on CMOs in relation to data that is held by other CMOs. However, on the one hand, Article 25(1) contains a corresponding transparency obligation that enables CMOs to exercise its monitoring duty. The provision requires of (data storing) CMOs to provide electronically up-to-date information that enables the identification of the online music repertoire it represents to other CMOs, as well as to rightholders it represents and online service providers, following a 'duly justified' request. The information provided under the transparency obligation must include the musical works represented, rights (wholly or in part) and territories covered. CMOs may, however, take reasonable measures to protect the accuracy and integrity of the information, to control its reuse, and to protect commercially sensitive information (Article 25(2). On the other hand, the monitoring requirement under Article 24(2)(d) is complemented by Article 26(1) which imposes a corresponding set of obligations on CMOs in relation to their own repertoire. Under this provision CMOs are required to have in place arrangements that make it possible for *other CMOs, rightholders*, and *online service providers* to request, in relation to their respective online rights, *a correction* of any of the data that is generated by a CMO as a result of the obligations imposed by Article 24(2) and Article 25. Although the provision operates as a form of prerogative for rightholders and service providers, CMOs are under a duty to use the mechanism as their monitoring obligation follows from Article 24(2)(d).[53] Whilst information must be corrected by the CMO without undue delay, for such correction to actually take place, the claim must be *sufficiently substantiated* on the basis of reasonable evidence that the information is inaccurate. This burden of proof, imposed on CMOs, rightholders and service providers alike, eliminates frivolous claims, and therefore the administrative burden, but also ensures authenticity of the claim and reliability of the information in the record once it is corrected.

In relation to the interface between the activities of CMOs and online service providers, Article 27 imposes another set of obligations that could be useful for the metadata record apart from the obligations under Article 25 and 26 (which also encompass online service providers). Notably, in connection with their obligation to monitor the use of online rights in musical works by online service providers (Article 27(1)), CMOs must offer such providers the possibility to report actual use while the service providers are, correspondingly, obliged to *accurately* report such use (Article 27(2)). This interface, though designed with traditional rights management tasks in mind (such as invoicing and distribution of remuneration to rightholders), supports the obligations and prerogatives under Articles 24, 25 and 26 to identify and correct inconsistencies in the metadata record. Whilst the provision requires of CMOs to offer at least one method of reporting which is based on industry standards or practices, it being voluntary for the service provider to use such a standard, the CMO may reject reporting in a proprietary format when a widely used industry standard for electronic exchange of data is offered by the CMO (Article 27(2)).

Good governance presumes that mechanisms to efficiently manage rights keep pace with technological development and demands of the market. The obligations under the CRM Directive, designed to

[53] As mentioned above, in case of the CMOs own repertoire, the duty to maintain the metadata record follows from the first set of obligations, particularly Article 24(2)(a)-(b).

establish certain database prowess that is responsive to such developments, clearly place CMOs in a position of stewardship over the metadata record. They are expected to store, monitor, and correct. Whilst the obligations apply to the narrow context of multi-territorial licensing by CMOs, they can serve as a blueprint for the implementation of rules to harmonise and improve copyright infrastructure more broadly across European creative sectors and territories.

## 4.3. Work notifications for content blocking

As to existing rules that may serve as catalysts to generate, harmonise and improve copyright metadata in the EU, several provisions of the 2019 CDSM Directive seem of particular interest. The work notification mechanism in Article 17(4)(b) CDSMD sheds light on a first avenue that could lead to the establishment of legal obligations to create and harmonise copyright metadata. Arguably, it even offers a promising opportunity for data improvement, especially with regard to categories of literary and artistic works that regularly play an important role on user-generated content platforms. Music, film, photography and other forms of visual art seem particularly relevant in this context.[54]

Article 17(4)(b) requires online content-sharing service providers (OCSSPs)[55] to use their best endeavours to ensure the unavailability of works and other protected subject matter for which rightholders have provided them with "relevant and necessary" information. Thus, this legal provision initiates a flow of data from rightholders to online platforms. The notification of works opens up the possibility of ensuring the application of measures to block and remove infringing content. In this context, it can be assumed that "relevant and necessary" information in the sense of Article 17(4)(b) goes beyond mere work-related data. A copyright owner submitting information cannot avoid also informing the online platform of his identity, address and other contact details, as well as the nature and (territorial) scope of the rights asserted. According to Article 17(8) CDSMD, OCSSPs must provide rightholders, at their request, with appropriate information on the operation of their procedures for cooperation under Article 17(4) CDSMD. Without contact information, this reporting obligation cannot be fulfilled. When it comes to complaint and redress procedures under Article 17(9) CDSMD, rightholders must also "duly justify" the reasons for their request to block content. Obviously, the exchange of information between rightholders and online platform providers is thus not only intended to ensure up-to-date information on works enjoying copyright protection, but also to bring about an accurate and constantly updated collection of data on rightholders and contact information. Otherwise, OCSSPs will hardly be able to report on their practice of blocking content and request rightholders to substantiate blocking requests in the context of complaint procedures.

However, before painting an overly positive picture of Article 17(4)(b) as a cure for all kinds of music metadata problems, it is important to point out that the provision is only one building block in a more complex puzzle. The regulatory framework of Article 17 CDSMD – as can be seen from the first two paragraphs of the provision – focuses on the right of communication to the public and making available to the public. Accordingly, the notification mechanism resulting from Article 17(4)(b) also concerns

[54] See Senftleben, in Aplin, Research Handbook on IP and Digital Technologies, 2020, 136-162; Triaille/Dusollier et al, Study on the Application of Directive 2001/29/EC on Copyright and Related Rights in the Information Society, 2013, 457-510; Helberger/Guibault et al, Legal Aspects of User Created Content, 2009; Wong, Vanderbilt Journal of Entertainment and Technology Law 11 (2009), 1075; Lee, University of Illinois Law Review 2008, 1459; OECD, Participative Web: User-Created Content, Doc. DSTI/ICCP/IE(2006)7/Final, 2007, available at: https://web-archive.oecd.org/2012-06-15/135484-38393115.pdf.

[55] See the definition of relevant service providers in Article 2(6) CDSMD.

these exclusive rights. The right of communication to the public and the right of making available to the public may be central to OCSSPs and various other forms of digital services. However, new technologies that offer promising exploitation opportunities for literary and artistic works may also predominantly affect the reproduction right and require the acquisition of rights of use in this area. The aforementioned use of human musical productions for AI training purposes can serve as an example.[56] As the provisions on text and data mining in Articles 3 and 4 CDSMD show, the reproduction right is central in this context.

The question therefore arises whether metadata stemming from Article 17(4)(b) notifications can provide useful information for initiatives aimed at identifying works and clarifying rights in new technology areas, such as the AI sector. The answer to this question depends on the phrase "relevant and necessary" information in Article 17(4)(b). In order to ensure the unavailability of protected works on online platforms, it seems sufficient to know who is entitled to prohibit the sharing of user-generated content because they are the holders of rights of communication to the public and/or making available. However, this fact does not preclude the further enrichment of data transfers. As already pointed out, it is the overarching objective of metadata improvement to increase the visibility and accessibility of protected works and create new licensing opportunities. Rightholders who support these objectives may therefore be willing to go beyond the information necessary for Article 17(4)(b) content blocking and provide additional information covering a broader range of exclusive rights. This wider provision of metadata may include, for example, the reproduction right. Article 17(4)(b) CDSMD may, in other words, have the effect of jump-starting a broader process of aggregating copyright metadata. This broader process may include additional exclusive rights, such as the reproduction right.

It should also be taken into account that rightholders provide work-related information under Article 17(4)(b) in order to prevent unauthorised user uploads to online platforms. The data provided serve to identify the work and infringing copies. Given this objective, Article 17(4)(b) notifications may not reveal the nature and content of the work itself. A potential user searching for a specific type of work, such as an AI developer searching for a specific category of human expression, may therefore find the information derived from Article 17(4)(b) notifications insufficient.

Again, it is important to bear in mind that reliance on Article 17(4)(b) notifications is an element of a broader strategy to use legal obligations to enhance the visibility and accessibility of the European repertoire for use and licensing in digital and algorithmic contexts. These benefits can be a strong incentive for rightholders to go beyond raw work identification data and provide additional information reflecting the nature and content of the work. The stakeholder dialogue which the Commission will initiate on the basis of Article 17(10) CDSMD may also address the issue of copyright data. It could include a discussion of "best practices for cooperation" in the area of metadata improvement and lead to the establishment of appropriate reporting standards and metadata enrichment strategies that go beyond the information that is strictly necessary for content blocking under Article 17(4)(b).

[56] Ducato R, Strowel A (2019) Limitations to text and data mining and consumer empowerment: making the case for a right to "machine legibility." IIC Int Rev Intellect Prop Competition Law 50:649; Geiger C, Frosio G, Bulayenko O (2018b) Text and data mining in the proposed Copyright Reform: making the EU ready for an age of big data? IIC Int Rev Intellect Prop Competition Law 49:814; Otero BG (2021) Machine learning models under the copyright microscope: is EU copyright fit for purpose? GRUR Int 70:1043; Senftleben M (2023) Generative AI and author remuneration. Int Rev Intellect Prop Competition Law (IIC) 54:1535–1560. https://doi.org/10.1007/s40319-023-01399-4; Kretschmer, M., Margoni, T. & Oruç, P. Copyright Law and the Lifecycle of Machine Learning Models. *IIC* **55**, 110–138 (2024). https://doi.org/10.1007/s40319-023-01419-3; Emanuilov, I., Margoni, T., Forget me not: memorisation in generative sequence models trained on open source licensed code, (2024), prepring available at https://zenodo.org/records/10635479.

## 4.4. Text and data mining rules

The content blocking mechanism in Article 17(4)(b) CDSMD is not the only legal tool that can be employed to support initiatives aiming at copyright data improvement. The TDM rule in Article 4 CDSMD – and in particular the opt-out mechanism enshrined in the third paragraph of this provision – offers a further opportunity to provide regulatory support. Considering the aforementioned focus of Article 17(4)(b) CDSMD on the right of communication to the public and the right of making available to the public, Article 4(3) CDSMD seems a particularly important counterpart. As already indicated, the harmonised TDM provisions concern the right of reproduction.

In addition to the exemption of scientific TDM in Article 3 CDSMD, Article 4(1) CDSMD contains a more general exemption. Under this additional provision, anyone may make copies of works or databases for the purposes of TDM and retain them as long as necessary for the TDM process.[57] With regard to this broader category of TDM outside the scope of the scientific research rule in Article 3 CDSMD, Article 4(3) CDSMD adds an important nuance by stipulating that rightholders can reserve their rights. The provision contains the following opt-out mechanism:

> The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.[58]

Before examining how this opt-out mechanism could contribute to data improvement initiatives, it seems important to clarify the scope of the provision, in particular with regard to the important case of using literary and artistic works for the training of generative AI systems. As the CDSM Directive dates back to 2019, it may be argued that rights reservations under Article 4(3) CDSMD do not cover these AI training activities. Arguably, the EU legislator did not have in mind the use of copyrighted material as mere data input for the training of generative AI systems which did not exist in 2019.[59] In the TDM debate, it has been underlined around the globe that TDM copies have a specific nature. They fall outside the concept of reproduction in the traditional sense of making copies for the purpose of consulting and enjoying a work.[60] From a US perspective, Michael Carroll has pointed out that in the context of TDM:

> copies are made only for computational research and the durable outputs of any text and data mining analysis would be factual data and would not contain enough of the original expression in the analysed articles to be copies that count.[61]

Explaining the outright exemption of TDM activities in Article 30-4(ii) of the Japanese Copyright Act, Tatsuhiro Ueno has pointed out that:

> if an exploitation of a work is aimed at neither enjoying it nor causing another person to enjoy it (e.g. text-and-data mining, reverse engineering), there is no need to guarantee the opportunity of an author or copyright holder to receive compensation and thus copyright does not need to cover such

[57] Article 4(1) and (2) CDSMD. As to the relevance of Article 4 CDSMD to generative AI systems, see Quintais (2023).
[58] Article 4(3) CDSMD.
[59] For a discussion of this argument with regard to the right of reproduction in international copyright law, see Senftleben (2022c), 1493-1502.
[60] Cf. Senftleben (2022c), 1495-1502.
[61] Carroll (2019), 954.

> exploitation. In other words, exploitation of this kind does not prejudice the copyright holder's interests protected by a copyright law.[62]

Criticizing the regulation of TDM in the EU, Rosanna Ducato and Alain Strowel described the following alternative approach:

> when acts of reproduction are carried out for the purpose of search and TDM, the work, although it might be reproduced in part, is not used as a work: the work only serves as a tool or data for deriving other relevant information. The expressive features of the work are not used, and there is no public to enjoy the work, as the work is only an input in a process for searching a corpus and identifying occurrences and possible trends or patterns.[63]

In fact, the distinction between use of "works as works" and use "as data" is not entirely new in the European copyright debate. In 2011, Maurizio Borghi and Stavroula Karapapa already developed the concept of "de-intellectualized use"[64] against the background of mass digitization projects, such as the Google Book Search. As Borghi and Karapapa point out, mass digitization turns protected content into mere data – with the result that "the expression of the idea embodied in the work is not primarily used to communicate the 'speech' of the author to the public but rather to form the basis of machine-workable algorithms."[65]

In the light of these comments, it seems to be an open question whether the use of copyrighted works during AI training falls within the scope of copyright. As use in this specific setting does not constitute use of an author's individual expression for communication purposes, copyright may be inapplicable from the outset and copyright data improvement may be beyond reach. Luckily, the AI Act (AIA) – adopted after the generative AI revolution and containing provisions on generative AI[66] – clarifies the matter. Recital 60i AIA confirms that the use of literary and artistic works for AI training purposes has copyright relevance and involves acts of text and data mining that require the authorisation of rightholders: "[a]ny use of copyright protected content requires the authorisation of the rightholder concerned unless relevant copyright exceptions and limitations apply."[67]

In line with this clarification in Recital 60i AIA, it can be assumed that EU copyright law brings all forms of TDM, including TDM for generative AI training purposes, under the umbrella of the right of reproduction and, accordingly, offers opportunities for copyright data improvement with regard to this exclusive right. More concretely, this configuration of the right of reproduction means that EU copyright law brings commercial AI training falling under Article 4(1) CDSMD within the reach of rightholders seeking to receive a remuneration for the use of their works.[68] Referring to the opt-out mechanism in Article 4(3) CDSMD, the AI Act confirms the intention to give rightholders the opportunity to exercise control over the use of their works for AI training purposes in Article 4 CDSMD scenarios:

> Where the rights to opt out has been expressly reserved in an appropriate manner, providers of general-purpose AI models need to obtain an authorisation from rightholders if they want to carry out text and data mining over such works.[69]

[62] Ueno (2021), 150-151.
[63] Ducato/Strowel (2021), 334.
[64] Borghi/Karapapa (2011), 45.
[65] Borghi/Karapapa (2011), 44-45.
[66] For an overview, see Senftleben AI Act #.
[67] Recital 60i AIA.
[68] Cf. Keller (2023); Communia (2023).
[69] Recital 60i AIA.

As in other cases where copyright holders can refuse the permission for a given form of use, this veto right can pave the way for remuneration payments.[70] It is conceivable that the rights reservation option in Article 4(3) CDSMD leads to the evolution of machine-readable rights reservation protocols that express different rightholder standpoints. One standpoint could be a machine-readable rights reservation that signals an outright exclusion of any use of the literary and artistic work at issue for AI training purposes. Using this rights reservation option, rightholders can express their preference for an outright prohibition and prevent TDM of their literary and artistic repertoire altogether. An alternative standpoint, however, could be a machine-readable rights reservation that prohibits use for AI training purposes only if the AI trainer behind the crawler is reluctant to pay remuneration. Using this alternative version, rightholders can express their willingness to permit the use against the payment of remuneration. In other words: the rights reservation option in Article 4(3) CDSMD can lead to generally agreed, machine-readable remuneration protocols that trigger an automated process for the payment of remuneration.

For this overarching rights clearance infrastructure to take shape, however, it is indispensable to enrich the opt-out declaration with relevant metadata. At first glance, Article 4(3) CDSMD does not seem to have much potential in this regard. A mere opt-out statement, for instance in the form of robots.txt embedded in a website, need not offer much detail in addition to the indication that the website content is not available for TDM and the AI crawler should look for training material elsewhere. The moment a rightholder seeks to use the opt-out mechanism as a vehicle to conclude licensing deals, however, the situation becomes markedly different. As a minimum, the rightholder will have to enrich the machine-readable opt-out statement with basic ownership and contact information. Who is the rightholder? What is the (territorial) scope of the rights? How can the rightholder be contacted? To attract attention, it also seems advisable to provide descriptive metadata. Which music repertoire is available? Which genre, style, period of music creation? Finally, and perhaps most importantly, it must be considered that the website with the opt-out statement will not always contain the musical works which the rights reservation concerns. If, for instance, a CMO, such as the French SACEM,[71] exercises the opt-out right following from Article 4(3) CDSMD, the opt-out statement and corresponding robots.txt become elements of the website of the collecting society itself. However, this website is unlikely to contain all the musical works belonging to the repertoire administered by the CMO. If a crawler searching for AI training resources refrains from including the CMO website in the training dataset, the opt out thus remains ineffective. CMO repertoire on other websites may still find its way into the TDM dataset unless these other websites contain the same machine-readable opt-out information. Hence, it can be important to enrich the opt out statement with sufficiently detailed information on the works falling under the rights reservation. In other words: opt-outs under Article 4(3) CDSMD may go far beyond the mere rights reservation statement and offer richer copyright data. For instance, it is conceivable that the opt-out statement contains a link to a database that contains detailed information on composers, performers, titles, rightholders etc. of musical works covered by the opt out. In particular, such a database link makes sense when rightholders, including CMOs, seek to use the opt-out mechanism as an invitation to enter into licensing agreements. Hence, it cannot be

[70] Cf. the positive assessment of the situation by Keller (2023); Communia (2023).

[71] See https://societe.sacem.fr/en/news/our-society/sacem-favour-virtuous-transparent-and-fair-ai-exercises-its-right-opt-out (last visited on 29 March 2024): "Against a backdrop of increasing development of artificial intelligence (AI) tools, Sacem is exercising its right to opt-out on behalf of its members. From now on, data mining of works in Sacem's repertoire by entities developing artificial intelligence tools will require prior authorisation from Sacem, in order to ensure fair remuneration for the authors, composers and music publishers it represents."

ruled out that Article 4(3) CDSMD becomes a propelling force for metadata creation and improvement in the EU.

## 4.5. Common European Data space

In addition to the copyright rules and mechanisms described in the preceding sections, the concept of Common European Data Spaces (CEDS) crystallizes the most recent policy and legislative efforts deployed by European Commission (EC) to regulate data transactions and create value from data. In order to understand the multiple goals and opportunities made possible by CEDS this section offers an overview of the complex and multifaceted regulatory framework. It should be noted at the outset that CEDS and the relevant regulatory framework are underpinned by quite distant goals and legislative approaches in comparison to the copyright *acquis*. In the literature this observation has been labelled as a transition from a property-based to a governance-based model of data regulation.[72] Such a structural difference in the regulatory approach requires to introduce and map CEDS's policy and regulatory options in a way that allows the identification, albeit at a general level, of this idiosyncratic domain.

The extent to which CEDS and associated legislation may play a role in favouring access and portability of music metadata is an interesting and potentially valuable insight. A good example of that could be found in Rec. 2 of the DA where, setting out the reasons underpinning the Regulation, it is stated that "barriers to data sharing prevent an optimal allocation of data for the benefit of society. Those barriers include ... the high level of fragmentation of information in data silos, poor metadata management, the absence of standards for semantic and technical interoperability, ... a lack of common data sharing practices and the abuse of contractual imbalances with regard to data access and use". Whereas music metadata does not find an explicit recognition in the DA, it is difficult not to see a clear link between the (meta)data access and interoperability problems identified by the DA with regards to the data economy and the shortcoming in the music metadata sector identified in the first part of this policy brief.

Nevertheless, due to the recency of the CEDS policy and regulatory environment, which is still largely in the making, it would be difficult to offer a conclusive roadmap of the interaction between music metadata and CEDS relevant legislation at this point in time. Accordingly, this section intends to map and lay the groundwork for a further exploration of the opportunities and potential challenges that CEDS may represent in the field of music metadata.

### 4.5.1. The regulatory framework

CEDS's regulatory framework is vast and could very well encompass any element of EU or domestic law actionable upon a transaction having as an object data, the governance of data, or data infrastructures. Despite this general premise, it is nonetheless possible to identify a few statutory interventions that are particularly significant for CEDS, alongside newly introduced legislation aimed at supporting their

[72] Margoni, T., et al, Data Property, Data Governance and Common European Data Spaces, in Computerrecht: Tijdschrift voor Informatica, Telecommunicatie en Recht, 2023.

advancement:[73] the Open, alongside newly introduced legislation aimed at supporting their advancement [74, 75, 76, 77, 78], even if the precise boundaries of what is included in this definition may vary.

It should further be observed that the relationship between DDL and the *copyright acquis* is very often addressed by way of the "without prejudice" approach. In other words, where DDL and the *copyright acquis* may be in tension, the "rules of engagement" are often summarised by a clause in DDL clarifying that "this regulation is without prejudice to" followed by a list of EU law sources that should not be affected by the regulation containing the "without prejudice" provision. This list usually contains the main EU copyright directives.[79] Whereas there are many good reasons for such a provision, it is also clear, particularly from the perspective of the copyright scholar, that there will be in the future, plausibly within CEDS dynamics, various situations where rules on data sharing usually contained in DDL may contrast with rules protecting works or other subject matter (which could qualify as "data" under DDL) belonging to the copyright field. In these cases, very little guidance is offered. One of the few exception to this situation, also identified as the absence of legal "interlinkers" in the literature,[80] is the DA. Art. 43 DA establishes an "interlinker" with the *copyright acquis* and in particular with the *sui generis* database right (SGDR) established in Art. 7 of the Database Directive, clarifying that the SGDR does not apply when data is obtained from or generated by an IoT product or related service. Similar provisions relating to the SGDR can be identified in other DDL sources.[81]

### 4.5.2. Data as a regulatory narrative

There are various possible starting points to offer a reconstruction of the DDL and of its relevance for CEDS. This is mainly due, on the one hand, to the broad scope of the DDL and, on the other hand, to the dynamic nature of CEDS. The perspective adopted here focuses on the concept of data as a direct object of regulation, i.e., as a concept that not only receives a direct definition, but which represents a whole taxonomy of data types, which often enjoy, or are subjected to, specific rights or obligations.

DDL defines data as: "any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audio-visual recording".[82] The

[73] Commission, 'A European strategy for data' COM(2020) 66 final, 12.
[74] 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 (recast) (2019) OJ L 172/56.
[75] 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) (2022) OJ L 152/1.
[76] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act), (2023) OJ L, 22.12.2023 accessible at: https://eur-lex.europa.eu/eli/reg/2023/2854/oj.
[77] At the time of writing the AI Act has not yet been officially published, however, the latest version available, the EP Adopted text of 13 March 2024 is available at this link: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf.
[78] Other instruments, general (e.g., Trade Secret Directive) or sector specific (the proposal for a Health Data Space Regulation), will not be covered in this overview.
[79] See for instance Recital 3 DGA.
[80] Margoni/Strowel, Contractual freedom and fairness in EU data sharing agreements, in de Werra & Calboli (Eds), Research Handbook on Intellectual Property Licensing; 2024, Edward Elgar, forthcoming.
[81] Art. 5(7) DGA: "Re-use of data shall be allowed only in compliance with intellectual property rights. The right of the maker of a database as provided for in Article 7(1) of Directive 96/9/EC shall not be exercised by public sector bodies in order to prevent the re-use of data or to restrict re-use beyond the limits set by this Regulation".
[82] Art. 2(1) DA.

DDL also offers definitions of personal data (as defined in the GDPR[83]) and of non-personal data (as any other data[84]). Interestingly, the DA includes a definition of metadata as: "a structured description of the contents or the use of data facilitating the discovery or use of that data".[85] Within this broad and horizontal definition(s) of "data", the DDL offers additional sector-specific definitions, which often include both personal, non-personal and meta-data. A brief overview of these sector-specific data categories provided for by various DDL instruments follows.

### 4.5.3. Data examples

Proceeding in temporal order, the first act of secondary legislation creating specific data (sub-)categories is the ODD. Interesting for present purposes is the concepts of High value datasets (HVD) and research data. The ODD defines High-Value Datasets (HVD) "as documents the re-use of which is associated with important benefits for society, the environment and the economy" (ODD, Art. 2(10)). HVD must be made available free of charge, in machine-readable format, via APIs and where appropriate, as a bulk download. In January 2023, the Commission's Implementing Regulation laying down a list of specific high-value datasets and the arrangements for their publication and re-use[86] entered into force, giving Public Sector Bodies (PSBs) 16 months to make the relevant HVD available. HVD should be found in thematic categories such postcodes, national and local maps (geospatial), energy consumption and satellite images (earth observation and environment), in situ data from instruments and weather forecasts (meteorological), demographic and economic indicators (statistics), business registers and registration identifiers (companies and company ownership), road signs and inland waterways (mobility). Music metadata are not present. However, it seems quite clear that the function played by the thematic datasets, particularly in the field of statistics and registration identifiers (e.g., company ownership), could offer an interesting *de lege ferenda* inspiration for a potential future inclusion of music metadata, particularly, if they were collected or managed by a Public Sector Body. Alternatively, a similar effect could be perhaps even conceavedoutside the ODD, in a future dedicated intervention.

A second specific category of data is identified in the ODD: research data. They are defined as "documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process or are commonly accepted in the research community as necessary to validate research findings and results" (ODD, Art. 2(9)). Art. 10 clarifies that research data must be re-usable for both commercial and non-commercial purposes when researchers have already made it publicly available through an institutional or subject-based repository.[87] It seems less clear whether research data could become a relevant category in the field of music metadata management. Nonetheless, the reuse requirements of this category remain an interesting example of how data access and reuse rules can be construed.

HVD and research data are two instances of specifically identified types of data subjected to specific access and reuse obligations. However, it should not be overlooked that the ODD regulates more broadly Public Sector Information, a category for which it has introduced a general framework of "reusability" by default. For those types of data (or better, following the ODD taxonomy, information)

[83] Art. 2(3) DA.
[84] Art. 2(4) DA.
[85] Art. 2(2) DA.
[86] Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use [2023] OJ L 345/12.
[87] Van Eechoud, 'FAIR, FRAND and open - The institutionalization of research data sharing under the EU data strategy,' in Susy Frankel and others (Eds), *Improving Intellectual Property* 2023, 320.

that does not fit into the remit of the ODD, the DGA, among other goals, has advanced a set of conditions for the re-use public sector data. Once again, there does not seem to be a clear link with music metadata in the reuse by default provisions of the ODD and DGA. Yet, the general approach rooted in the facilitation of reuse of data could represent a regulatory model for music metadata.

However, whereas the ODD and DGA set some interesting access and reuse rules, a common denominator of these two legislative interventions is that they refer generally to publicly held data. Muisc metadata may nevertheless be found more easily within a private data exchange enviroment. From this perspective the DA embodies an element of novelty as it introduces some interesting access and reuse provisions with regard to (mainly) privately held data.

The first context triggering new data access rights involves the data generated by connected devices and services. Internet of Things (IoT), or more precisely *connected products and related services,*[88] is a broad category of products (and related services) that span from everyday home or consumer appliances (smart fridges, smart TVs, etc) to professional devices in ambits as varied as precision agriculture and industrial machinery. Chapter 2 of the DA, which regulates business-to-consumer (B2C) and business-to-business (B2B) data sharing, establishes that IoT products and related services must be designed in such a way that the generated data, including metadata necessary for their interpretation, are directly accessible to the users (Art. 3(1) DA). Art. 4, further establishes that, when IoT data cannot be directly accessed by users, data holders are obligated to make such data, including metadata, accessible to the user, without delay and free of charge in a comprehensive, structured and commonly used machine-readable format (Art. 4(1) DA). This provision, despite being variously limited,[89] creates a general data access provision for users of IoT data, which can include audio, vidual or audiovisual recording and related metadata. What is more, Art. 5 DA establishes a "right of the user to share data with third parties". That said, it seems that the provisions on IoT devices will be of limited impact on music metadata, in the light of the fact that IoT products "obtain, generate or collect ... data concerning their performance, use or environment" and that "data that such sensor-equipped connected products generate when the user records, transmits, displays or plays content, as well as the content itself, which is often covered by intellectual property rights, inter alia for use by an online service, should not be covered by this Regulation" (Rec. 16 DA).

Nevertheless, and despite these limitations, Arts. 4 and 5 afford users of IoT products not only a right of access to (meta)data but also a right of portability of eligible (non-personal) data vis-à-vis data holders. This access and portability right[90] is conceptually not too distant from its personal data counterpart found in the GDPR.[91]

IoT data access and portability is a first interesting example of how the DA intervenes in data transactions between private actors with very specific contractual and precontractual obligations. The

[88] They are defined respectively as "items that obtain, generate or collect data concerning their use or environment and that are able to communicate product data via an electronic communications service and whose primary function is not the storing, processing or transmission of data on behalf of any party other than the user" and as "digital services, other than an electronic communications service, including software, which are connected with the product at the time of the purchase, rent or lease in such a way that their absence would prevent the connected product from performing one or more of its functions" (Art. 2(5) and (6) DA).
[89] Limitations of users' access rights include the preservation of the security of the IoT products and of trade-secrets that may be part of the IoT data (Art. 4 DA).
[90] Art. 7(2) DA clarifies that "Any contractual term which, to the detriment of the user, excludes the application of, derogates from or varies the effect of the user's rights under this Chapter shall not be binding on the user".
[91] The DA preamble (e.g., Recs. 7 and 35) clarifies that the Data Act "complements" Regulations (EU) 2016/679 in the field of non-personal data.

declared objectives of this course of action are multifaceted. Particularly relevant for the present analysis, there is the evident ambition to rebalance the asymmetries that characterise a market where the developer of a product or services (in this case IoT devices) has a *de facto* power to control co-generated data thanks to their ability to design the technology.[92] Additionally, the DA clarifies a broader set of objectives aimed at ensuring fairness in the allocation of data value among the actors in the data economy and at fostering fair access and use of data in order to contribute to establishing a genuine internal market for data.[93] In the specific case of IoT these data are potentially valuable to the user, to support innovation and to develop secondary markets.[94]

These interventions, which are all functional to the achievement of a working market for data, through the establishment of relevant data infrastructures and governance frameworks[95] (also known as Common European Data Spaces), are not specifically devised, nor arguably apply in the case of IoT, to music (meta)data. Yet, many of the problems that the DA, and DDL more broadly, attempt to address via a semi-regulated data market seem *prima facie* to also characterise the field of music metadata. Furthermore, the dynamic nature of the Common European Data Spaces allows for the development of sector-specific governance models, including those based on the specificities of music metadata.[96]

### 4.5.4. Other data relevant provisions

IoT data is a first and interesting case as it creates a rather novel framework for the access and sharing of data and metadata with a view of enhancing efficiency and fairness in secondary markets. Yet, these obligations are functionally limited to the data that is generated in the IoT field. It should be noted that there are numerous other provisions of both within the DA and in the broader DDL that deserve closer scrutiny. The follwoing is a first overview of the potentially relevant provisions in the variagete DDL frameowrk. Their actual relevance will depend in part on the specific legal qualification of the transactions relating to music meta data and on the final design that a music metadata repository or registry will adopt.

- Chapter 3 of the DA regulates B2B data sharing obligations and applies to “any private sector data that is subject to statutory data sharing obligations”. Art. 12 DA clarifies that this applies to business-to-business relations when a data holder is obliged under Article 5 DA (or under applicable Union law or national legislation) to make data available to a data recipient. Art. 8 stipulates that data holders are required to agree with data recipients the arrangements for making the data available and shall do so under fair, reasonable and non-discriminatory (FRAND) terms.
- Ch. 4 of the DA applies “to any private sector data access and use on the basis of a contract between enterprises” (Art. 1(2)(c) DA). Therefore, differently from Ch. 2 (that only applies to IoT data), or Ch. 3 (which is limited to situations where data holders are under an obligation to share data with a data recipient), Ch. 4 plausibly applies to any contract under EU law that has as its object data access or data uses between enterprises. Rec. 60 clarifies that not all aspects of data contracts are covered, but only those elements that “are related to making data available, that is contractual terms concerning access to and use of the data as well as liability or remedies for

[92] American Law Institute and European Law Institute, ‘ALI-ELI principles for a data economy – Data transactions and data rights, 2017-2021’ (2022) the American Law Institute and the European Law Institute.
[93] Rec. 119 DA; for a critical perspective see Kerber Wolfgang, ‘Governance of IoT Data: Why the EU Data Act Will not Fulfill Its Objectives’ (2023) GRUR International 72(2), 120–135.
[94] Rec. 15 DA.
[95] Commission, ‘Commission Staff Working Document on Common European Data Spaces’ SWD (2024) 21 final, 4.
[96] Marina Micheli and others, ‘Mapping the landscape of data intermediaries — Emerging models for more inclusive data governance’ Publications Office of the European Union (2023), 13.

breach and termination of data related obligations". In general terms, Ch. 4 establishes that unfair contractual terms, when unilaterally imposed by one enterprise on another, shall not be binding on the second enterprise.

- Ch VI DA creates a "right to switch" between data processing services. It establishes that providers of data processing services shall adopt certain measures to enable customers to switch data processing services. Given the generous definitions of data processing services and of custormers the provisions contained in Ch VI may likewise have a relevant impact on many data contracts by establishing specific rules on enabling and facilitating customers to switch, information and transparency obligations and switching fees.
- Chapter VIII of the DA stipulates that participants in data spaces that offer data or data services to other participants "shall comply with the following essential requirements to facilitate the interoperability of data, of data sharing mechanisms and services, as well as of common European data spaces which are purpose- or sector-specific or cross-sectoral interoperable frameworks for common standards and practices to share or jointly process data for, inter alia, the development of new products and services, scientific research or civil society initiatives". Ch. VIII is specific to data spaces (and arguably to all data spaces, not only to Common European Data Spaces) and sets forth a list of requirements that focus mainly on the interoperability and publicity of standards, formats, vocabularies, data structures and data processing services.
- The AI Act does not deal specifically with CEDS, however, it seems that CEDS are an essential element of the EU data strategy which will benefit, among other things, AI. Rec 68 seems to confirm this where it establishes that: "European common data spaces established by the Commission and the facilitation of data sharing between businesses and with government in the public interest will be instrumental to provide trustful, accountable and non-discriminatory access to high-quality data for the training, validation and testing of AI systems. For example, in health, the European health data space will facilitate non-discriminatory access to health data and the training of AI algorithms on those data sets, in a privacy-preserving, secure, timely, transparent and trustworthy manner, and with an appropriate institutional governance". This notion is further emphasised in the 2024 Staff Working Document which presents CEDS as a means to fuelling artificial intelligence development and a key driver for innovation in Europe.[97]
- Chapter III of the DGA regulates Data Intermediation Services (DIS). DIS are defined as services which aim to "establish commercial relationships for the purposes of data sharing between an undetermined number of data subjects and data holders on the one hand and data users on the other, through technical, legal or other means". Providers of DIS may therefore in certain circumstances apply to "orchestrators of data sharing ecosystems" such as CEDS.[98] However, specifically excluded from DIS are "services that focus on the intermediation of copyright-protected content" (Art. 2(11) DGA). Rec. 29 further clarifies that "services that focus on the intermediation of copyright-protected content, such as online content-sharing service providers as defined in Article 2, point (6), of Directive (EU) 2019/790" should not be covered by the DGA. The operations of DIS are subjected to a number of conditions (Art. 12 DGA). Whether music (meta)data exchange would fit under the DIS provisions will arguably depend on the specific infrastructure and governance model of the service. Furthermore, the DGA has established the European Data Innovation Board (EDIB), empowered to propose guidelines for CEDS.[99] Serving as an advisory body to the Commission, the EDIB addresses DGA implementation, prioritizing cross-sectoral interoperability standards, and supporting the sharing of best practices across CEDS.[100]

[97] Commission, 'Commission Staff Working Document on Common European Data Spaces' SWD (2024) 21 final, 3.

[98] Recital 28 DGA.

[99] Recital 54 and Artt. 29-30 DGA.

[100] Commission, 'Commission Staff Working Document on Common European Data Spaces' SWD (2024) 21 final, 6.

Standards remains a pressing issue, a report from the Data Spaces Support Centre collected 130 standards and specifications after surveying the technologies landscape for data spaces.[101] Efforts such as the Data Spaces Business Alliance underscore the commitment to supporting technical convergence within the community.[102]

### 4.5.5. Interim remarks

In conclusion the regulatory framework created by DDL is a complex and multifaceted system of rules focusing on data, data transactions and data intermediaries. The rules and the specific ambits of application vary, often in function of the specific goals of the relevant directive or regulation. Whereas it seems too early to establish with certainty which of these provision will apply to a music (meta)data space, and whereas this determination will largely depend on the design of that data space, the governance of which is a key feature thereof but one which remains largely unprescribed, the relevance of DDL, and its many possible interactions with the *copyright acquis*, should be further explored.

## 4.6. Bundling of metadata streams

Considering the scale of the metadata flow and the need for up-to-date information on protected works, the nature and scope of rights and the question of ownership, it becomes apparent that particular opportunities arise from the overarching rules in the CRM Directive, the specific notification and opt-out systems following from Article 4(3) and Article 17(4)(b) CDSMD, and potential fresh initiatives that may be taken in the framework of data spaces and the growing body of EU data legislation: if all copyright-related data submitted under pertinent provisions could be harmonised and merged into a central EU music data collection, the resulting accumulation of data resources could potentially lead to a data reservoir that would dwarf existing data silos of collecting societies, rightholders and distribution platforms.[103] Moreover, as several described legal mechanisms require continuous updating of rights and ownership information, it can be assumed that a centralised EU copyright data repository could have a relatively high degree of timeliness and accuracy.

However, in order to launch such an EU copyright data repository, metadata flowing from the described legal mechanisms would have to be bundled in a systematic way. Therefore, rightholders providing metadata in the described contexts should be required to submit this information in parallel to a central body managing the EU copyright data repository. The pooling and harmonisation of copyright data could (or, if CEDS rules apply, would even have to) be done in an open and interoperable format to ensure general data accessibility and general data transparency for all interested users and access to licensing deals for all repertoire holders – regardless of size and market power. This could reduce the risk of large repertoire owners, which have a wider spectrum of works and metadata, gaining a

[101] Data Spaces Support Centre, 'Collection of Standards and Technologies Landscape Version 1.0' (*Data Spaces Support Cente.* 2023).

[102] Data Spaces Business Alliance, 'Technical Convergence: Discussion Document' (*Data Spaces Business Alliance*, 2023) < https://data-spaces-business-alliance.eu/wp-content/uploads/dlm_uploads/Data-Spaces-Business-Alliance-Technical-Convergence-V2.pdf>.

[103] In the legislative process leading to the adoption of Article 17 of the DSM Directive, Germany has already proposed in this sense to introduce "public, transparent notification procedures" to counteract a de facto copyright register in the hands of dominant platforms. See Council of the European Union, Opinion of Germany, 5 April 2019, point 5, p. 4, available at: https://data.consilium.europa.eu/doc/document/ST-7986-2019-ADD-1-REV-2/en/pdf.

competitive advantage with regard to new licensing opportunities, for instance in the field of AI training.

A template for legislation that would ensure this redirection of copyright data to a central data collection point can already be found in Article 3(6) of the Orphan Works Directive 2012/28 (in relation to information on the use of orphan works) and in Article 10(1) CDSMD (in relation to information on out-of-print works). Interestingly, these provisions also mention the institution that could take care of the central EU copyright database: the European Union Intellectual Property Office (EUIPO).

In order to achieve the desired interoperability of the transmitted data, the legal obligation to transmit music metadata to the EUIPO could be complemented by the additional obligation to provide the data in a specific, standardised format. In this way, the law could be used as a tool to address not only issues of data accuracy and timeliness, but also the problem of data interoperability and data harmonisation. The obligation to submit data in parallel to EUIPO would have the advantage for rightholders of creating a generally accepted data submission standard that would pave the way for the universal applicability of submitted metadata. If other addressees of metadata information, such as OCSSPs and AI trainers, would also be bound to accept the information in this standardised format, music rightholders would no longer have to deal with individual data transmission standards, which may differ from one data user to the other.

## 4.7. International ban on formalities

The international prohibition of formalities arising from Article 5(2) of the Berne Convention for the Protection of Literary and Artistic Works (BC) need not be an insurmountable obstacle with regard to the outlined system of data transmission in interoperable form. According to Article 5(2) BC, “the enjoyment and exercise” of the rights granted in Article 5(1) BC are not subject to any formal requirement. Article 5(1) includes the rights which the laws of the Berne Union countries "presently grant or may in the future grant to domestic authors, as well as the rights specifically granted in this Convention." As Van Gompel explains in his in-depth analysis of the scope of the prohibition of formalities under Article 5(2) BC, the prohibition covers:

> formalities relating to the coming into existence, the maintenance and the enforcement of copyright. The Berne prohibition on formalities does not extend to formalities that regulate the extent of protection or the means of redress afforded to authors to protect their rights. This suggests that formalities are allowed if they establish the manner of exercising copyright, but not if their non-compliance renders the exercise of rights completely impossible.[104]

Within this matrix, the concrete legal tools discussed above, such as the notification system following from Article 17(4)(b) CDSMD and the opt-out mechanism under Article 4(3) CDSMD, fall into the category of permissible formalities concerning the manner in which copyright is exercised and the regulation of the scope of protection. The content moderation rules in Article 17 CDSMD can serve as an example: by providing that platform providers carry out an act of communication to the public or an act of making available to the public when they grant public access to protected works uploaded by users, Article 17(1) CDSMD establishes direct, primary liability of online platforms[105] in an area that has

[104] Van Gompel, Formalities in Copyright Law: An Analysis of Their History, Rationales and Possible Future, 2011, 212.

[105] For a more detailed discussion of the legal nature of the right granted in Article 17 DSM-RL, see Husovec/Quintais, GRUR Int. 2021, 325 (325-348).

traditionally been regulated from the perspective of secondary liability for the uploading of infringing content by users.[106] The detailed design of this exclusive right, including the possibility to avoid liability by purchasing licences and applying content filters (Article 17(4)(a) and (b) CDSMD), clearly regulates the scope of protection.[107] The fact that rightholders are required to provide 'relevant and necessary' information under Article 17(4)(b) CDSMD shows that the provision establishes a specific way of exercising copyright.[108] In any case, rightholders can enforce their rights against individual uploaders in situations where platform providers have not been granted a licence. Thus, rather than making the exercise of copyright impossible, Article 17(4)(b) CDSMD provides rightholders with an additional means of ensuring the unavailability of their works on online platforms. The same conclusion can be drawn with regard to the opt-out mechanism in Article 4(3) CDSMD. In this case, it is also noteworthy that in Article 10bis(1) BC, the Berne Convention itself contains a longstanding opt-out system that concerns 'articles published in newspapers or periodicals on current economic, political or religious topics.' Considering this prototype in the Berne Convention itself, it can hardly be concluded that opt-out mechanisms, such as Article 4(3) CDSMD, impose impermissible copyright formalities in the sense of Article 5(2) BC.[109]

All in all, the notification system that follows from Article 17(4)(b) CDSMD and the opt-out mechanism established in Article 4(3) CDSMD, thus, constitute permissible formalities that increase the scope of protection and regulate the way copyright is exercised in the specific contexts of content moderation on online platforms and the creation of datasets for AI training. Against this background, it also seems possible to extend these notification and opt-out mechanisms and include an obligation to submit relevant metadata to a central EU data collection point that could be established at the EUIPO. The prohibition of formalities in Article 5(2) BC does not stand in the way of this approach.

# 5. Conclusion

In order to improve the visibility and accessibility of the European repertoire of works and to enable the creative industries to benefit from licensing opportunities in the field of new technologies, it is important to establish a comprehensive copyright data infrastructure focusing on European content, including smaller and lesser-known repertoires and reflecting the full cultural diversity of the EU Member States. In this context, the general framework for copyright data following from the rules in the CRM Directive plays an important role. More concrete obligations to provide work-related data can follow from provisions in the EU copyright acquis. For instance, the notification of "relevant and necessary" information for the purpose of blocking infringing content under Article 17(4)(b) CDSMD provides an important starting point. If Article 17(4)(b) notifications sent to online platform providers are collected and pooled in a central EU copyright data repository, the resulting accumulation of EU copyright data could lead to a data reservoir that dwarfs existing data silos of collecting societies,

[106] See Leistner, Zeitschrift für Geistiges Eigentum 2020, 123-214; Husovec, Injunctions Against Intermediaries in the European Union - Accountable But Not Liable? 2017; Angelopoulos, European Intermediary Liability in Copyright: A Tort-Based Analysis, 2016; Senftleben, JIPITEC 2013, 87 (87-90 and 94-95); Hoeren/Yankova, IIC 2012, 501; Matulionyte/Nérisson, IIC 2011, 55; Peguera, Columbia Journal of Law and the Arts 32 (2009), 481; Leistner, GRUR 2006, 801.

[107] Cf. van Gompel, id. at 212.

[108] Cf. van Gompel, id. at 212.

[109] For a more detailed discussion of these opt-out mechanisms, see M.R.F. Senftleben, 'Generative AI and Author Remuneration', *International Review of Intellectual Property and Competition Law* 54 (2023), 1535 (1544-1546).

rightholders and distribution platforms. However, notifications under Article 17(4)(b) would have to be detailed and comprehensive enough to allow an EU data repository to enhance the visibility of the European repertoire in a meaningful way and to expand licensing opportunities for copyright holders. On the one hand, this objective requires notifications covering a wide range of exclusive rights. In addition to the right of communication to the public and making available to the public, the reproduction right should also be covered, for example. On the other hand, it would be desirable to enrich notifications under Article 17(4)(b) with descriptive metadata that provide information on the nature and content of the works notified.

Luckily, Article 17(4)(b) CDSMD is not the only copyright norm that can provide important impulses for the creation and maintenance of accurate copyright metadata. The opt-out mechanism in Article 4(3) CDSMD – concerning a ban on the use of works for TDM purposes – offers a further example of an existing provision that could generate a continuous flow of copyright metadata. In this case, sufficiently rich metadata can be expected when rightholders use the opt-out mechanism to indicate that they are willing to grant a licence in exchange for the payment of remuneration. Seeking to pave the way for a licence agreement, it makes sense to go beyond the mere opt-out statement and provide additional work-related information and, thus, relevant metadata.

Complementing the described data improvement options in the EU copyright acquis, the legislative and policy framework set forth by CEDS and DDL could offer the regulatory support necessary to achieve the described copyright data creation, harmonisation and improvement goals. However, the relationship between the copyright *acquis* and data and digital legislation, and in particular the carve-outs present in the latter with regard to specific copyright-related activities, need to be properly assessed. Nevertheless, despite the actual applicability of the DDL legislative framework to music metadata, it is undeniable that many of the obstacles that DDL is set to solve, have an at least *prima facie* parallel in the field of music metadata. From this point of view, the development of a data space for music metadata (whether a CEDS or not), could certainly offer an interesting model, in line with the EU priorities for the data economy, that should be explored further.

# References

American Law Institute and European Law Institute. (2022). ALI-ELI principles for a data economy – Data transactions and data rights, 2017-2021

Angelopoulos. (2016). European Intermediary Liability in Copyright: A Tort-Based Analysis

Antal. (2020). Feasibility Study On Promoting Slovak Music In Slovakia & Abroad. Retrieved from https://reprex.nl/publication/listen_local_2020/

AppleWorld. (2019). APPLE, SPOTIFY TO FUND NEW MUSIC ROYALTIES COLLECTIVE. Retrieved from https://appleworld.today/apple-spotify-to-fund-new-music-royalties-collective/

Azzi/El Hage. (2021). Les métadonnées liées aux images fixes.

Bechtold. (2016). Information Society Directive in Dreier/Hugenholtz (eds), *Concise European Copyright Law* (Kluwer Law International), 483

Borghi/Karapapa. (2011). Non-display uses of copyright works: Google Books and beyond. *Queen Mary Journal of Intellectual Property*, *1*(1), 21-52.), 44-45

Burk. (2020). Thirty-Six Views of Copyright Authorship, by Jackson Pollock, Houston Law Review 58 (2020), 263 (270-321)

C-666/18 *IT Development SAS* v *Free Mobile SAS*

Cadavid. (2023). *The Origin and Purpose of Legal Protection for the Integrity of Copyright Metadata* (2023) IIC 54 1179-1202.

Carroll. (2019). Copyright and the progress of science: Why text and data mining is lawful. UC Davis L. Rev., 53, 893.

CISAC. (2024). CIAGP Technical Operations. Retrieved from https://www.cisac.org/services/creator-relations/ciagp-technical-operations

CISAC. (2024). CIS-Net. Retrieved from https://www.cisac.org/services/information-services/cis-net

Council of the European Union. (2019). Opinion of Germany, 5 April 2019, point 5, p. 4. Retrieved from https://data.consilium.europa.eu/doc/document/ST-7986-2019-ADD-1-REV-2/en/pdf

Covington/Adams/Sargin. (2019). In: Proceedings of the 10th Acm Conference on Recommender Systems, RecSys '16, 2016, 191-198. Retrieved from https://doi.org/10.1145/2959100.2959190

Dang-Nguyen/Sjøen/Le et al. (2023). Practical Analyses of How Common Social Media Platforms and Photo Storage Services Handle Uploaded Images. In: Dang-Nguyen, DT., et al. MultiMedia Modeling. MMM 2023. Lecture Notes in Computer Science, vol 13834. Springer, Cham. Retrieved from https://link.springer.com/chapter/10.1007/978-3-031-27818-1_14

DOI. (2024). Retrieved from https://www.doi.org/

Ducato/Strowel. (2021). Ensuring text and data mining: remaining issues with the EU copyright exceptions and possible ways out. *European Intellectual Property Review*.), p. 334

EDItEUR. (2024). About. Retrieved from https://www.editeur.org/2/About/#Intro

EDItEUR. (2024). ONIX. Retrieved from https://www.editeur.org/8/ONIX/

EIDR. (2024). Retrieved from https://www.eidr.org/

European Commission, Directorate-General for Communications Networks, Content and Technology, (2022). *Study on copyright and new technologies : copyright data management and artificial intelligence*, Publications Office of the European Union. https://data.europa.eu/doi/10.2759/570559.

European Commission. (2019). Ethics guidelines for trustworthy AI. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

European Commission. (2020). High-level expert group on artificial intelligence. Retrieved from https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

European Composer and Songwriter Alliance/European Writers' Council et al. (2023), *Joint Statement from Authors' and Performers' Organisations on Artificial Intelligence and the AI Act – True Culture Needs Originals: Transparency and Consent are Key to the Ethical Use of AI*. Retrieved from https://screendirectors.eu/joint-statement-from-authors-and-performers-organisations-on-artificial-intelligence-and-the-ai-act/

European Guild for Artificial Intelligence Regulation (2023), *Manifesto for AI Companies Regulation in Europe*. Retrieved from https://www.egair.eu/#manifesto

Frosio (2024) Should we ban generative AI, incentivise it or make it a medium for inclusive creativity? In: Bonadio E, Sganga C (eds) A research agenda for EU copyright law. Edward Elgar, Cheltenham (forthcoming). Retrieved from https://ssrn.com/abstract=4527461

Geiger/Iaia (2023) The forgotten creator: towards a statutory remuneration right for machine learning of generative AI. Computer Law and Security Review, forthcoming. Retrieved from https://ssrn.com/abstract=4594873

Geiger/Schönherr/Stamatoudi/Torremans/Karapapa (2021). The Information Society Directive. In Stamatoudi/Torremans (eds), *EU Copyright Law: A Commentary* (2nd ed, Edward Elagar), 361

Ginsburg/Budiardjo. (2019). Authors and Machines. Berkeley Technology Law Journal 34, 343 (395-396)

Gronau/Schaefer. (2021). EIPR 2021, 488-494; Schaefer, Kluwer Copyright Blog, 27 November 2020. Retrieved from http://copyrightblog.kluweriplaw.com/2020/11/27/why-metadata-matter-for-the-future-of-copyright/

Helberger/Guibault et al. (2019). Legal Aspects of User Created Content, 2009

Hoeren/Yankova. (2012). The liability of internet intermediaries-the German perspective. *IIC-INTERNATIONAL REVIEW OF INTELLECTUAL PROPERTY AND COMPETITION LAW*, *43*(5), 501 (501-531)

House Reports 1551, Public Law 115-264, dated 11 October 2018

Hugenholtz/Quintais. (2021). Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output? *IIC, 52*(9), 1190-1216.

Husovec. (2017). Injunctions Against Intermediaries in the European Union - Accountable But Not Liable? (Vol. 41). Cambridge University Press

Husovec/Quintais. (2021). How to License Article 17? Exploring the Implementation Options for the New EU Rules on Content-Sharing Platforms under the Copyright in the Digital Single Market Directive. GRUR Int. 2021, 325 (325-348)

ICE. (2024). About. Retrieved from https://www.iceservices.com/about/?_gl=1*xyff4r*_up*MQ..*_ga*MTU5ODY1ODAyLjE3MTEzNjY4NDg.*_ga_QSZEMWD5QD*MTcxMTM3MDMzMS4yLjAuMTcxMTM3MDMzMS4wLjAuMA

ICE. (2024). Cube. Retrieved from https://www.iceservices.com/innovation/cube/.

IFPI. (2019). RDx, recording industry's new data exchange service, now fully operational. Retrieved from https://www.ifpi.org/rdx-recording-industrys-new-data-exchange-service-now-fully-operational/.

IFPI. (2024). Retrieved from https://www.rdx-portal.org/.

Initiative Urheberrecht (2023), *Joint Statement: Authors and Performers Call for Safeguards Around Generative AI in the European AI Act*, 19 April 2023. Retrieved from https://urheber.info/diskurs/call-for-safeguards-around-generative-ai

IPTC (2024). Quick guide to IPTC Photo Metadata on Google Images. Retrieved from https://iptc.org/standards/photo-metadata/quick-guide-to-iptc-photo-metadata-and-google-images/

ISCC. (2024). ISCC - Codec & Algorithms. Retrieved from https://core.iscc.codes/

Isherwood. (2011). Global Repertoire Database, presentation at the WIPO Meeting wipo_cr_doc_ge_11, 13-14 October 2011, Geneva: WIPO 2011. Retrieved from: https://www.wipo.int/meetings/en/2011/wipo_cr_doc_ge_11/prov_program.html

ISO. (2024). International Standard Content Code (ISCC). Retrieved from https://www.iso.org/standard/77899.html

Jacobson/Murali/Newett et al. (2016). In: Proceedings of the 10th Acm Conference on Recommender Systems, RecSys '16, 2016, 373. Retrieved from https://doi.org/10.1145/2959100.2959120

Janssens/Gotzen. (2018-2019). Kunstmatige Kunst. Bedenkingen bij de toepassing van het auteursrecht op Artificiële Intelligentie. Auteurs en Media 2018-2019, 323 (325-327)

Katz. (2005). COMPARING ANTITRUST ENFORCEMENT IN THE UNITED STATES AND EUROPE. Journal of Competition Law and Economics 1, p. 276

Keller. (2023). Protecting creatives or impeding progress? Machine learning and the EU copyright framework. Retrieved from https://copyrightblog.kluweriplaw.com/2023/02/20/protecting-creatives-or-impeding-progress-machine-learning-and-the-eu-copyright-framework/

Lee. (2008). Warming up to user-generated content. University of Illinois Law Review 2008, 1459

Leistner. (2020). European copyright licensing and infringement liability under Art. 17 DSM-Directive compared to secondary liability of content platforms in the US–Can we make the new European system a global opportunity instead of a local challenge?. Zeitschrift für Geistiges Eigentum 2020, 123-214

Lyons/Sun/Collopy et al. (2019). Music 2025 - The Music Data Dilemma: issues facing the music industry in improving data management. Retrieved from https://www.gov.uk/government/publications/music-2025-the-music-data-dilemma

Margoni, in: Calboli/Montagnani, Handbook on Intellectual Property Research, 2021, 487-505

Margoni/Strowel. (Forthcoming). Contractual freedom and fairness in EU data sharing agreements, in: de Werra & Calboli (Eds), Research Handbook on Intellectual Property Licensing; 2024, Edward Elgar, forthcoming

Matulionyte/Nerisson. (2011). The French route to an ISP safe harbor, compared to German and US ways. *IIC International Review of Intellectual Property and Competition Law*, *42*(1), 55 (55-73)

Mo, Kang, Hu, Zhou, Li And Gu. (2023). Towards Trustworthy Digital Media In The Aigc Era: An Introduction To The Upcoming IsoJpegTrust Standard, in *IEEE Communications Standards Magazine*, vol. 7, no. 4, pp. 2-5. Retrieved from https://ieeexplore.ieee.org/abstract/document/10353009

OECD. (2012). Participative Web: User-Created Content, Doc. DSTI/ICCP/IE(2006)7/Final, 2007. Retrieved from https://web-archive.oecd.org/2012-06-15/135484-38393115.pdf

Pearlman. (2018). Recognizing Artificial Intelligence (AI) as Authors and Inventors Under U.S. Intellectual Property Law. Richmond Journal of Law and Technology 24, 1 (4)

Peguera (2009). The DMCA safe harbors and their European counterparts: a comparative analysis of some common problems. *Colum. JL & Arts*, *32*, 481.

Posner. (2005). TRANSACTION COSTS AND ANTITRUST CONCERNS IN THE LICENSING OF INTELLECTUAL PROPERTY. John Marshall Review of Intellectual Property Law 4, p. 325

Resnikoff. (2014). Global Repertoire Database Declared a Global Failure…. Retrieved from: https://www.digitalmusicnews.com/2014/07/10/global-repertoire-database-declared-global-failure/

Sanchez. (2019). Pan European Collecting Societies Agree on Standards for Digital Royalty Processes. Retrieved from https://www.digitalmusicnews.com/2019/07/26/towge-digital-royalty-group/

Schwemer. (2019). Licensing and Access to Content in the European Union. In Licensing and Access to Content in the European Union: Regulation between Copyright and Competition Law, 2019, 68-73.

Senftleben. (2020). In Aplin, Research Handbook on IP and Digital Technologies, 2020, 136-162

Senftleben. (2013). Breathing space for cloud-based business models: exploring the matrix of copyright limitations, safe harbours and injunctions, JIPITEC, 4, 87

Senftleben. (2023). Generative AI and Author Remuneration, *International Review of Intellectual Property and Competition Law*, 54, 1535-1560

Senftleben/Buijtelaar. (2020). Robot Creativity: An Incentive-Based Neighbouring Rights Approach. EIPR 2020, 797 (804-808). Retrieved from https://doi.org/10.2139/ssrn.3707741

Senftleben/Margoni/Antal et al. (2022). Ensuring the Visibility and Accessibility of European Creative Content on the World Market - The Need for Copyright Data Improvement in the Light of New Technologies and the Opportunity Arising from Article 17 of the CDSM Directive. JIPITEC Vol. 13, No. 1, pp. 67-86, 2022. Retrieved from https://ssrn.com/abstract=3785272

The MLC. (2021). The Mechanical Licensing Collective Begins Full Operations as Envisioned by The Music Modernization Act of 2018. Retrieved from https://blog.themlc.com/press/mechanical-licensing-collective-begins-full-operations-envisioned-music-modernization-act

Triaille/Dusollier et al. (2013). Study on the Application of Directive 2001/29/EC on Copyright and Related Rights in the Information Society, 2013, 457-510

U.S. Copyright Office Library of Congress. (2018). MLC Comments in Reply to the Designation Proposal of the American Music Licensing Collective, Inc, Docket No. 2018-11, 21. Retrieved from https://bw-98d8a23fd60826a2a474c5b4f5811707-bwcore.s3.amazonaws.com/photos/Proposed_MLC_-_Reply_Comments.pdf

Ueno. (2021). The Flexible Copyright Exception for 'Non-Enjoyment'Purposes–Recent Amendment in Japan and Its Implication. *GRUR International*, *70*(2), 145-152, 150-151

Van Eechoud. (2023). 'FAIR, FRAND and open - The institutionalization of research data sharing under the EU data strategy,' in Susy Frankel and others (Eds), Improving Intellectual Property 2023, 320

Van Gompel. (2011). Formalities in Copyright Law: An Analysis of Their History, Rationales and Possible Future, 212

WIPO. (2021). Discontinuation of the WIPO Proof Token Generation Service. Retrieved from https://www.wipo.int/wipoproof/en/news/2021/news_0003.html.

WIPO. (2024). WIPO PROOF (Discontinued). Retrieved from https://www.wipo.int/wipoproof/en.

Wong. (2009). "Transformative" User-Generated Content in Copyright Law: Infringing Derivative Works or Fair Use?. Vanderbilt Journal of Entertainment and Technology Law 11, 1075

# Annex

Table 1: Overview of mostly ongoing initiatives in metadata interoperability within the content rights infrastructure (2024)[110]

| Overview of mostly ongoing initiatives in metadata interoperability within the content rights infrastructure | | | | | |
|---|---|---|---|---|---|
| **Nr.** | **Initiative** | **Summary (mainly based on the initiatives' website information)** | | | |
| **1** | **Working groups and reports** | **Groups, consultations, and reports pursuing solutions for rights infrastructure** | **Content /media of focus** | **Sector of focus** | **Data scope** |
| **1.1** | **Developing the Copyright infrastructure** *By: Finnish Ministry of Culture and Education* | Since 2020, the Finnish Ministry of Culture and Education has been working together with several Finnish CMOs (Gramex, Kopiosto, Kuvasto, Sanasto and Teosto) to adopt International Standard Name Identifiers (ISNIs) for rightholders. Through the construction of interfaces, the ISNI should (more) effectively serve as an identifier shared between different organisations that use contributor identifiers. This should lead to data enrichment among CMOs and their affiliates.[111] | Any | Any | Parties Content Rights |
| **1.2** | **Listen Local** *By: SOZA and various stakeholders from the music industry* | Supported by the Slovak Arts Council, a collaboration between the collecting society SOZA and various stakeholders from the music industry has led to the creation of a prototype for a comprehensive data and metadatabase of the Slovak music repertoire. On this basis, the consortium created the prototype of the "Listen Local" recommendation system, which meets the requirements of the recommendations on the topic of trustworthy artificial intelligence (AI) of the European High-Level Expert Group on Artificial Intelligence. The accompanying feasibility study showed and quantified the problems arising from incomplete copyright data in existing databases and commercial AI solutions. | Music | Any | Any Awareness and Understanding |
| **1.3** | **Music 2025 UK** *By: British Intellectual Property Office* | The British Intellectual Property Office (IPO) commissioned a report in 2019 regarding the music data dilemma, based on the views and opinions of around 50 interviews from high profile music industry representatives. Suggested improvements were in the realm of (i) education and awareness, (ii) collaboration, (iii) Interoperability and (iv) governance. | Music | Any | Any Awareness and Understanding |

[110] Based in part on "*Study on copyright and new technologies: copyright data management and artificial intelligence*, SMART 2019/0038, Annex 5.3 – List of current and ongoing initiatives in data interoperability within the content rights infrastructure."

[111] See: https://www.kansalliskirjasto.fi/en/projects/adoption-isnis-copyright-management-organisations

| | | | | | |
|---|---|---|---|---|---|
| **1.4** | **Music Copyright Explained** | The guide 'Music Copyright Explained' is commissioned by the IPO as a free, user-friendly way to explain music-makers how music copyright works in the UK and how copyright gives them control over the songs and recordings they create. | Music | Any | Any |
| **1.5** | **International Council of Creators of Graphic, Plastic and Photographic Arts (CIAGP)** *By: Visual Arts Council of Confederation of Societies of Authors and Composers (CISAC)* | The Visual Arts Council of CISAC has expanded its original work on resale rights and established an online licensing platform under the umbrella of the International Council of Creators of Graphic, Plastic and Photographic Arts (CIAGP). | Visual works | Any | Parties<br>Content<br>Rights |
| **1.6** | **Reversion Rights in the European Member States** *By: reCreating Europe* | The reCreating Europe consortium has brought forth a working paper which maps provisions allowing authors and performers to reclaim their rights (reversion rights) which are currently or were historically a part of the national laws of the EU Member States. The initiative spurred from the introduction of the right of revocation in Article 22 of the Directive EU/2019/790 on copyright in the Digital Single Market, a reversion right based on a use-it-or-lose-it logic. | Any | Legal | Rights |
| **1.7** | **Technical Online Working Group Europe (TOWGE)** | The Technical Online Working Group Europe (TOWGE), brings together a large group of European collecting societies, music publishers and rights agencies to develop a digital royalty processing system. TOWGE sets out the rights splits between performing and mechanical rights for each territory in online and mobile exploitations. Metadata details include repertoire name, the direct licensor managing the repertoire, the repertoire's exact definition, which DSPs fall under the mandate, the use types covered, and the period of the mandate. | Music | Any | Parties<br>Content<br>Rights |
| **1.8** | **Working Party on Intellectual Property** *By: Council of the European Union* | In 2019, under the Finnish presidency, the Council issued a stocktaking document on developing the Copyright Infrastructure. Its objective is clearly stated: unleashing the digital potential of Europe's creative sectors through effective metadata, improved licensing efficiency, and automated revenue distribution. | Any | Any | Parties<br>Content<br>Rights<br><br>Awarenes s and Understa nding |

| 2 | Data frameworks | Generic frameworks and schemas for interoperability for content and rights data | Content /media of focus | Sector of focus | Data scope |
|---|---|---|---|---|---|
| **2.1** | **Interoperability of Data in E-commerce Systems** *(with the support of the European Commission)* | The *indecs* project (short for "Interoperability of Data in E-commerce Systems") was an international initiative of organizations for intellectual property rights owners. It successfully developed a framework of metadata standards to support networked commerce based on intellectual property. The project was partly funded by the European Commission's *Info 2000* initiative and laid the foundations for initiatives as ONIX, DDEX, DOI or MDDF. | Any | Any | Parties<br>Content<br>Rights |
| **2.2** | **Distributed Trust Rights Framework** *By: Digiciti Networks OÜ* | Digiciti Networks OÜ is a company that gathers new, breakthrough technologies that focus on fundamental aspects of rights management, including content identifiers, stakeholder identifiers, metadata associations, authoritative assertions, as well as the use of trusted, multi-party, distributed, dynamic data management systems to create and share Rights Management Information. | Any | Any | Parties<br>Content<br>Rights<br><br>Authority |
| **2.3** | **European Blockchain Services Infrastructure (EBSI)** *By: European Blockchain Partnership* | Since 2018, the European Blockchain Services Infrastructure (EBSI) has focused on four use cases relevant to rights management information: notarisation (of digital works or assets), certification, self-sovereign identity (of authors, rightholders or other stakeholders) and trusted data exchange. The European Blockchain Partnership has received the proposals of consortia asked to build a performing EBSI. | Any | Any | Rights<br><br>Authority |
| **2.4** | **Experiments towards a copyright infrastructure** *By: Estonian government* | Through platform AccelerateEstonia, the Estonian Government, in collaboration with the Estonian music sector, built a system in 2021 enabling identified artists and managers to declare rights related to identified songs and recordings, and forward their music and licensing data to streaming services. Such aggregation of data related to recordings and songs should ensure every rightholder gets paid fairly and swiftly. | Music | Any | Parties<br>Content<br>Rights<br><br>Awareness and Understanding<br><br>Authority |
| **2.5** | **Movielabs Digital Distribution Framework (MDDF)** *By: Motion Picture Laboratories Inc.* | The MovieLabs Digital Distribution Framework (MDDF) consists of standards and technologies enabling automation, cost reduction, and improved consumer experiences across the audio-visual industry. Several aspects of online distribution are included, such as identification, metadata, avails, asset delivery, and reporting. | Audio-visual | Comm. | Content<br>Rights |

| 3 | Identifier standards | Standards for identifiers of parties or content | Content /media of focus | Sector of focus | Data scope |
|---|---|---|---|---|---|
| **3.1** | **Universal Resource Identifier (URI)** | URI (Universal Resource Identifier) is a member of a universal set of names in registered name spaces and addresses referring to registered protocols or name spaces. | Any | Any | Any |
| **3.2** | **International Standard Audio-visual Number (ISAN)** *By: ISAN Agency* | ISAN (International Standard Audiovisual Number) is a voluntary numbering system for the identification of audiovisual works. It provides a unique, internationally recognized and permanent reference number for each audiovisual work registered in the ISAN system. | Audio-visual | Any | Content |
| **3.3** | **International Standard Book Number (ISBN)** *By: ISBN Agency* | ISBN (International Standard Book Number) is a product identifier used by publishers, booksellers, libraries, internet retailers and other supply chain participants for ordering, listing, sales records and stock control purposes. The ISBN identifies the registrant as well as the specific title, edition and format. | Books | Any | Content |
| **3.4** | **International Standard Content Code (ISCC)** *By: The ISCC Foundation* | ISCC (International Standard Content Code) is an identification system for digital assets (including encodings of text, images, audio, video or other content across all media-sectors). It's currently still under development at the International Organization for Standardization (ISO). The ISCC should serve as a similarity-preserving fingerprint designed for digital content to identify content in decentralized and networked environments across the creative industries (journalism, books, music, film, etc.). Moreover, the aim for the ISCC is to be free, open-source and transparent. | Any | Any | Content Authority |
| **3.5** | **International Standard Recording Code (ISRC)** *By: ISRC Agency* | ISRC (International Standard Recording Code) enables sound recordings and music videos to be uniquely and permanently identified. | Sound record-ings | Any | Content |
| **3.6** | **International Standard Serial Number (ISSN)** *By: ISSN Agency* | ISSN (International Standard Serial Number) is an identifier used to identify newspapers, journals, magazines and periodicals of all kinds and on all media–print and electronic. | Serial public-cations | Any | Content |
| **4** | **Identifier standards with metadata** | **Standards for identifiers of parties or content for which metadata registration is required** | **Content /media of focus** | **Sector of focus** | **Data scope** |
| **4.1** | **Digital Object Identifier (DOI)** | The DOI system provides an infrastructure for persistent unique identification of objects of any type over the Internet. The governance body of | Any | Any | Content |

| | | | | | |
|---|---|---|---|---|---|
| | *By: International DOI Foundation* | the DOI system is the DOI Foundation, that is also the ISO Registration Authority for the standard. The DOI system is implemented through a federation of registration agencies coordinated by the DOI Foundation. | | | |
| **4.2** | **Entertainment ID Registry (EIDR)** | Entertainment ID Registry (EIDR) is a universal labelling system for film and television data based on DOI technology. | Audio-visual | Comm. | Content |
| **4.3** | *Entertainment ID Registry (EIDR)*, in cooperation with *International Standard Audio-visual Number (ISAN) Agency* | The Entertainment ID Registry Association (EIDR) together with the International Standard Audiovisual Number International Agency (ISAN-IA) started a dual registration service. It resulted from the European Commission's Audiovisual Standard ID policy aimed at boosting opportunities for audiovisual arts, entertainment, information, and archival management.[112] | Audio-visual | Comm. | Content |
| **4.4** | **Interested Parties Information System (IPI System)** *By: SUISA, the Cooperative Society of Music Authors and Publishers in Switzerland*, used by all BIEM/CISAC societies | The Interested Parties System (IPI system) is a system that facilitates the global unique identification of rightholders acting across multiple creation disciplines assuming different roles (author of literature, musical performer, film director, etc.), and owning all rights (performing right, reproduction right, radio broadcast right etc.), determined by each creation discipline they deal with. | Any | CMOs | Parties |
| **4.5** | **International Standard Name Identifier (ISNI)** *By: ISNI International Agency Limited* | ISNI (International Standard Name Identifier) is the ISO certified global standard number for identifying the contributors to creative works and those active in their distribution, including researchers, inventors, writers, artists, visual creators, performers, producers, publishers, aggregators, and more. | Any | Any | Parties Content |
| **4.6** | **International Standard Musical Work Code (ISWC)** *By: International Confederation of Societies of Authors and Composers (CISAC)* | ISWC (International Standard Musical Work Code) identifies a musical work as a unique intangible creation. It relates to the result of an intangible creation of one or more people, regardless of copyright status, distributions or agreements that cover this creation. | Musical works | Comm. | Content |
| **5** | **Metadata standards (content & rights)** | **Metadata schemas for both content and rights data.** | **Content /media of focus** | **Sector of focus** | **Data scope** |

[112] See https://www.eidr.org/eidr-isan-ia-announce-joint-registration-service/.

| | | | | | |
|---|---|---|---|---|---|
| **5.1** | **CISAC Cue Sheet standards and rules** | Implemented in a collaboration between CISAC and music publishers and producers through the Society Publisher Forum, the Cue Sheet Standards & Rules simplify the rules governing the identification of musical works used in audio-visual productions. The harmonisation aims to improve the administration of music rights, bring a new consistency to the use of cue sheets, and should lead to increased efficiencies and potentially reduced costs for rightholders and users. | Music | Audio-visual | Content Rights |
| **5.2** | **Digital Data Exchange (DDEX)** *By: Digital Data Exchange LLC* | Digital Data Exchange LLC is a standards-setting organisation which runs the DDEX project. The purpose of DDEX is, in short, to develop standards relating to metadata creation and management, identification of entities and the communication of such information in relation to media rights, and to promote global awareness and compliant implementation of those standards. | Music | Comm. | Parties Content Rights |
| **5.3** | **EBUCore** *By: European Broadcast Union* | The EBUCore is an initiative of the European Broadcast Union. It defines a set of concepts, relationships and properties that apply to media, providing a framework for descriptive and technical metadata to use in service orientated architectures. | Broad-cast | Any | Content |
| **5.4** | **IPTC Photo Metadata** | The IPTC Photo Metadata Standard is the most used standard to describe photos, because of its universal acceptance among news and photo agencies, photographers, libraries, museums, and other related industries. The standard structures and defines metadata properties, allowing users to add precise and reliable data about images. | Photos | Any | Content Rights |
| **5.5** | **Online Information Exchange (ONIX)** *By: EDItEUR Ltd* | The ONIX family includes standards for books, serials and licensing terms & rights information. All ONIX standards are designed to support computer-to-computer communication between parties involved in creating, distributing, licensing or otherwise making available intellectual property in published form, whether physical or digital. All are expressed in XML. | Books, Serials | Any | Content Rights |
| **6** | **Metadata standards (rights)** | **Metadata schemas specifically for rights data.** | **Content /media of focus** | **Sector of focus** | **Data scope** |
| **6.1** | **Creative Commons** | Creative Commons (CC) is a not-for-profit organisation that aims to help overcome legal obstacles to the sharing of knowledge and | Any | Any | Rights |

| | | | | | |
|---|---|---|---|---|---|
| | | creativity to address the world's pressing challenges.<br>CC provides Creative Commons licenses and public domain tools, available to every person and organization in the world. These should enable a free, simple, and standardized way to grant copyright permissions for creative and academic works; ensure proper attribution; and allow others to copy, distribute, and make use of those works. | | | |
| **6.2** | **Europeana Rights Statements**<br>*By: Europeana & partners* | Europeana is a web portal created by the European Union containing digitised cultural heritage collections. Europeana has published twelve different rights statements, which can be used by cultural heritage institutions to communicate the copyright and re-use status of digital objects to the public. The rights statements have been designed with both human users and machine users (such as search engines) in mind and are made available as linked data. Each rights statement is located at a unique Uniform Resource Identifier (URI). | Any | Cultural heritage | Rights |
| **6.3** | **IPTC Web Statement of Rights**<br>*By: International Press Telecommunications Council* | The IPTC Photo Metadata Standard aims to define the optimal way to fill metadata fields such as Creator, Credit Line, Copyright Notice, Web Statement of Rights, and Licensor URL. Google Images started using the IPTC Photo Metadata in 2018 in their image search results. Next to a selected photo, the image's creator, credit line, and a copyright notice are shown. | Visual | Any | Rights Awareness and Understanding |
| **6.4** | **Open Digital Rights Language (ODRL)**<br>*By: World Wide Web Consortium & W3C* | The Open Digital Rights Language (ODRL) is a policy expression language that provides a flexible and interoperable information model, vocabulary, and encoding mechanisms for representing statements about the usage of content and services. | Any | Comm. | Rights |
| **6.5** | **Text and Data Mining Reservation Protocol (TDM Protocol)**<br>*By: World Wide Web Consortium & W3C* | The goal of the W3C Community Group "Text and Data Mining Reservation Protocol" is to facilitate Text and Data Mining (TDM) Reservation Protocol in and outside of Europe, by specifying a simple and practical machine-readable solution, capable of expressing the reservation of TDM rights - following the rules set by the new European DSM Directive / Art.4 - and the availability of machine-readable licenses for TDM actors. | Any | Any | Rights |
| **7** | **Data access/ exchange** | **Systems and schemas providing access to content and rights data from multiple sources** | **Content /media of focus** | **Sector of focus** | **Data scope** |

| | | | | | |
|---|---|---|---|---|---|
| **7.1** | **CIS-Net**<br>*By: CISAC, International Confederation of Societies of Authors and Composers* | With its various nodes in several regions of the world, the CIS-Net system and its associated standards represent a global tool to facilitate music licensing and revenue distribution. | Music | CMOs | Content Rights |
| **7.2** | **Concertify**<br>*By: Teosto, Mind Your Rights* | The Finnish collecting society Teosto's collaboration with the start-up *Mind Your Rights* has resulted in the platform "Concertify", which aims to provide an efficient and transparent system for cross-border copyright licensing in addition to existing industry structures. Concertify enables artists, rightholders, including collecting societies, music publishers and event organisers, to collaborate and transmit information directly using specific modules, such as a module for the transmission of setlists. | Music | Any | Content rights |
| **7.3** | **Cube**<br>*By: ICE Services – International Copyright Enterprise Services* | Cube is a new copyright platform of ICE Services, a joint venture between the Performing Rights Organisations PRS (United Kingdom), STIM (Sweden), and GEMA (Germany). Its goal is to deliver a highly automated copyright system, which will increase the speed and accuracy with which ICE consolidates multi-territorial copyright data by harnessing cloud computing and machine learning technologies. | Music | Comm. | Content Rights |
| **7.4** | **European Data Strategy**<br>*By: European Commission* | As part of the European Strategy for Data, "the Commission intends to fund the establishment of EU-wide common, interoperable data spaces in strategic sectors. Such spaces aim at overcoming legal and technical barriers to data sharing across organisations, by combining the necessary tools and infrastructures and addressing issues of trust, for example by way of common rules developed for the space." | Any | Any | Any |
| **7.5** | **Handle system**<br>*By: Corporation for National Research Initiatives (CNRI)* | The Handle system assigns persistent digital identifiers or handles to information resources. The Handle system provides the underlying technology for the DOI and EIDR identification systems (yet works independently of those) as well as others. It enables (i) digital objects to retain their identifiers when their location URL changes and (ii) a single handle to direct users to multiple objects (for example, content or rights metadata). | Any | Any | Any |
| **7.6** | **Linked Open Data**<br>*By: Europeana* | Linked Open Data revolves around publishing structured data that allows metadata to be connected and enriched, facilitating that | Any | Cultural | Any |

| | | | | | |
|---|---|---|---|---|---|
| | | different representations of the same content can be found, and links can be made between related resources.<br>Every Europeana dataset can be explored and queried through the SPARQL API. The metadata for all the objects in the Europeana portal is open, in that it is all licensed under the CC0 Public Domain Dedication under the terms of the Data Exchange Agreement (DEA), and can be freely downloaded via the API. | | heritage | |
| **7.7** | **OnLineArt (OLA)** | OnLineArt (OLA) is a one-stop shop for acquiring licences for the online use of works of visual art worldwide, which currently includes works by 60,000 artists. | Visual | | Content Rights |
| **7.8** | **Repertoire Data Exchange (RDx)**<br>*By: International Federation of the Phonographic Industry (IFPI) & Worldwide Independent Network (WIN)* | Repertoire Data Exchange (RDx) is an industry data portal for the supply and exchange of performance rights repertoire data between multiple record companies and multiple Music Licensing Companies (MLCs). RDx is a data exchange hub, which puts the DDEX "Recording Data and Rights" (RDR) standard into practice as part of its core functionality. | Music | Comm. | Content Rights |
| **7.9** | **URights**<br>*By: Société des auteurs, compositeurs et éditeurs de musique in France (SACEM) & International Business Machines Corporation (IBM)* | URights is a platform developed by SACEM and IBM. It is open by design to allow other partners to integrate, such as other CMOs across the world, allowing them to avoid cost duplications. URights allows CMOs to address royalties for creators and publishers in areas such as music and audio-visual content. The open architecture enables other partners to use URights' processing technology while maintaining their own rights databases, guarded by standards of security and data confidentiality. | Music | Comm. | Content Rights |
| **7.10** | **European Data Strategy**<br>*By: European Commission* | As part of the European Strategy for Data, "the Commission intends to fund the establishment of EU-wide common, interoperable data spaces in strategic sectors. Such spaces aim at overcoming legal and technical barriers to data sharing across organisations, by combining the necessary tools and infrastructures and addressing issues of trust, for example by way of common rules developed for the space." | Any | Any | Any |
| **8** | **Datasets** | **Non-proprietary databases and datasets of global value** | **Content /media of focus** | **Sector of focus** | **Data scope** |
| **8.1** | **Lumière VoD**<br>*By: European Audio-visual Observatory* | Lumière VoD is a directory of European works (film and TV content) available on on-demand services in Europe. It helps find the services and countries where a film or a TV content is | Audio-visual | Comm. | Content Rights |

| | | | | | |
|---|---|---|---|---|---|
| | | released on pay-video on-demand services, transactional and subscription Video on Demand (VoD), and combines search criteria to create lists of available films by director, country or year of production and available TV content by country of production. Lumière VoD is primarily designed for audio-visual industry professionals: authors, producers, distributors, film funds and regulators in order to help them track the exploitation of works on VoD and to assess the composition of the VoD catalogues. | | | |
| **8.2** | **Orphan Works Database** *By: EU Intellectual Property Office (EUIPO)* | The Orphan Works database is a single publicly accessible database that provides the public with information related to orphan works contained in the collections of publicly accessible libraries, educational establishments and museums, as well as archives, film or audio heritage institutions and public-service broadcasting organisations established in the Member States. Information related to orphan works is recorded in the database in accordance with Article 3(6) of the Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works. | Any | Cultural heritage | Content Rights Authority |
| **9** | **Rights Platforms** | **Shared solutions for managing rights** | **Content /media of focus** | **Sector of focus** | **Data scope** |
| **9.1** | **Access to Rights Data via Identification Technology Optimisation (ARDITO)** *European Horizon 2020 research and innovation project* | The ARDITO project aimed at filling the gap in the digital content value network and connecting online contents to rights information, by building a complementary digital rights data network. In the project, tools and market-driven services were developed to support creators and SMEs in the creative content sector to find new business ideas through monetising the re-use of their content. The project implemented existing components of a rights data network by optimising several content identification technologies (e.g. DOI, watermarking and fingerprinting) to provide unified access from the identifiers to the rights and licensing information and services and integrating them into the Copyright Hub ecosystem. | Text, Images | Comm. | Content Rights |
| **9.2** | **The Copyright Hub** | The Copyright Hub is a UK-based non-profit organisation focused on lowering the transaction costs of licensing copyrighted items. The Copyright Hub is a technology platform. A copyright work on the Internet, such as an | Any | Any | Parties Content Rights |

| | | | | | |
|---|---|---|---|---|---|
| | | image or a piece of music, is given a unique identifier. Someone wishing to reuse that image or that music can connect to the computer of the rights owner or creator, who also has a unique "party" identifier. This is called "resolving". The rights owner or creator can then offer, machine to machine, standard licenses for reuse requiring payment or proper acknowledgement. | | | Awareness and Understanding<br><br>Authority |
| **9.3** | **WIPO Connect**<br>*By: World Intellectual Property Office (WIPO)* | WIPO Connect is an IT solution, developed and offered by WIPO, aiming at facilitating the collective management of copyright and related rights. It works on two levels:<br>(i) *WIPO Connect Local* is a web application used for day-to-day operations, either installed on a local server or hosted in the cloud, for registration of registration of rightholders, Management of licensing agreements, etc.<br>(ii) *WIPO Connect Shared* is a fully cloud-based solution, synchronizing WIPO Connect Local implementations and exchanging data with industry data sources. | Any | CMOs | Parties<br>Content<br>Rights |
| **10** | **Authentication** | **Shared solutions for authenticating content, data or parties** | **Content /media of focus** | **Sector of focus** | **Data scope** |
| **10.1** | Content Authenticity Initiative (CAI)<br>Adobe and others | The Content Authenticity Initiative builds systems to provide provenance for digital media, giving creators tools to express objective reality and empowering consumers to evaluate whether what they are seeing is trustworthy. The initiative is designing components and drafting standards specifications for a simple, extensible, and distributed media provenance solution. It focuses its efforts on (a) the detection of deliberately deceptive media, (b) education, and (c) content attribution. | Any | Any | Parties<br>Content |
| **10.2** | **WIPO Proof**<br>*By: World Intellectual Property Office (WIPO)* | The WIPO PROOF service was launched as a program by WIPO in 2020. The digital business service provided a date- and time-stamped digital fingerprint of any file, proving its existence at a specific point in time. The service was meant to complement WIPO's existing intellectual property systems.<br>However, the service was formally discontinued on February 1, 2022. The WIPO Secretariat has explained to member states that "since the initial feasibility studies, the market has evolved quickly, driven by the accelerated digitalization". | Any | Any | Content<br><br>Authority |

| 11 | Policing | Systems for tracking or acting on breaches | Content /media of focus | Sector of focus | Data scope |
|---|---|---|---|---|---|
| 11.1 | **Copyright Clearing House on the Internet (CUII)** *By: Selbstregulierung Informationswirtschaft e.V., a German association of rightholders and Internet service providers* | The Copyright Clearing House on the Internet (CUII) is an independent body, founded by German internet access providers and rightholders to use objective criteria to check if blocking the access to a structurally copyright-infringing website is lawful. | Any | Any | Authority |