

MAPPING OF EU DATABASES AND METADATA STANDARDS PROVIDING INFORMATION ON COPYRIGHT- PROTECTED WORKS



Executive Summary

Part 1 – Background

Purpose and context

The study *Mapping of EU Databases and Metadata Standards Providing Information on Copyright-Protected Works* (the study) examines what is widely recognised as a structural weakness in the European (and global) copyright system: the absence of a coherent, interoperable and trustworthy infrastructure for copyright metadata (or copyright information).

The study covers metadata relating not only to copyright-protected works, but also to subject matter protected by related rights (such as performances and audiovisual fixations). Data relating to such works and subject matter, as well as to rights holders and usage conditions, is fragmented across national, sectoral, public and private databases, which are often based on heterogeneous and incompatible metadata standards.

This fragmentation generates legal uncertainty, inefficiencies in licensing and remuneration, and increased transaction costs for creators, users, intermediaries and public institutions. The importance of metadata and interoperability for the discoverability and circulation of European cultural content in digital environments has also been highlighted in recent EU research ⁽¹⁾.

The study maps existing databases, metadata practices and standards across creative sectors. Its findings will inform the way the EUIPO explores the feasibility of an EU-level advanced search service ('CopyrightView'), which would aim to improve transparency and access to information on copyright status, authorship, ownership and permitted uses of works across the EU.

⁽¹⁾ European Commission: Directorate-General for Education, Youth, Sport and Culture, Panteia, KEA European Affairs, Vrije Universiteit Brussel, DELab – University of Warsaw, Erasmus University Rotterdam, IDEA Consult, Clarke, M., Vroonhof, P., Byrne, C., Petrov, L., Le Gall, A., Antonucci, F., Albertelli, A., Rolando, E., Ranaivoson, H., Salganik, R., Wiratama, V., Afilipoaie, A., Hardy, W., Paliński, M., Rozynek, S., Berkers, , Calkins, T., Wijngaarden, Y., Vuijsteke, C., Valintelyte, E.Janssens, J., *Study on the discoverability of diverse European cultural content in the digital environment – Executive summary*, Publications Office of the European Union, 2026, <https://data.europa.eu/doi/10.2766/2913845>

Beyond enhanced transparency, such a service could generate significant efficiency gains by reducing the time, costs and legal uncertainty associated with identifying rights holders and clearing rights, thereby facilitating licensing transactions and the exploitation of works across borders. It could also support the effective exercise and enforcement of rights, including enabling rights holders to express and communicate reservations of rights (such as TDM opt-outs), streamline licensing workflows, and facilitate actions such as notice-and-takedown requests to online platforms. More broadly, it could contribute to more accurate attribution, improved remuneration flows, and greater trust in the digital copyright ecosystem.

Problem statement

The study identifies two structural problems:

First, although copyright metadata is essential for attribution, licensing, enforcement and remuneration, it is often created inconsistently, lacks sufficient standardisation and is rarely interoperable across systems.

Second, existing databases are seldom interconnected, even when they share similar objectives such as voluntary copyright registration, collective management, legal deposit or cultural heritage access.

These issues are amplified by technological and regulatory developments. Digitisation and global online distribution have increased the volume and circulation of creative content, while rights information systems have not evolved at the same pace. Artificial intelligence (AI), and in particular generative AI (GenAI), intensifies the need for reliable and machine-readable copyright information, especially to identify protected works, to express and enforce text and data mining (TDM) opt-outs, to document the provenance of training data, and to support the fair and transparent remuneration of creators where applicable.

From a legal perspective, EU legislation increasingly assumes the availability of reliable copyright metadata. Instruments such as the InfoSoc Directive, the Collective Rights Management (CRM) Directive, the Copyright in the Digital Single Market (CDSM) Directive, the Open Data Directive, the Data Governance Act, the Data Act and the AI Act all impose obligations whose effective

implementation relies, to some extent, on the availability of accurate, interoperable and trustworthy rights management information, as well as on the ability to exchange this information across borders and sectors.

Objectives and scope

The study has three main objectives:

- to map the existing copyright databases and metadata infrastructure in the EU;
- to analyse the types of information and metadata standards used by the operators of these databases in a European context; and
- to identify the benefits and lay the groundwork on the feasibility of CopyrightView, an advanced search service to be created by the EUIPO to interconnect existing national and EU-level copyright databases.

The scope covers the major creative sectors: film and television, music, publishing (books, journals and press), photography, video games, software (including AI-related content) and cultural heritage institutions. The study identifies challenges, opportunities and design considerations that should guide any future work on the possible development of a CopyrightView service.

Methodology and structure of the study

The study combines desk and field research that consisted of interviews with 31 organisations (28 structured interviews and 3 unstructured interviews), including public authorities, collective management organisations, cultural institutions, standard-setting bodies and private stakeholders across multiple Member States and sectors.

The analysis shifts from legal and conceptual foundations to empirical findings, before exploring opportunities and feasibility considerations.

Part 2 – Legal framework and concept of copyright metadata

The study reviews the main categories of copyright metadata used by stakeholders. At the core lies rights management information (RMI), as defined in international treaties and EU law. RMI is legally protected and comprises:

- identification metadata, which uniquely identifies works, manifestations and stakeholders (including standard identifiers such as ISBN, ISWC, ISAN and ISNI); and
- rights and usage metadata, which records authorship, ownership, licensing terms, permitted uses and restrictions.

Beyond RMI, the study highlights the importance of complementary metadata categories, including descriptive metadata, usage metadata, administrative and provenance metadata, and authenticity and provenance metadata, which has become increasingly relevant in the context of AI-generated and AI-assisted content. Together, these layers form a metadata system necessary for legal compliance, market efficiency and trust.

EU law increasingly embeds metadata-related obligations into copyright regulation, while remaining technically neutral. While this neutrality encourages innovation, it has also resulted in multiple parallel standards and uneven adoption across sectors, leading to fragmentation, limited interoperability between systems, inconsistencies in metadata quality, increased costs and legal uncertainty for stakeholders when identifying rights holders, clearing rights and managing licensing and remuneration processes.

A sustainable and interoperable copyright infrastructure will depend not only on common technical standards, but also on the creation of incentives encouraging stakeholders to record and exchange copyright data in consistent and standardised ways.

Part 3 – Database landscape

The ‘mapping’ exercise reveals a highly diverse database landscape, differing in mandate, governance, territorial scope, size and technical maturity. Databases include European-level initiatives, national libraries and public authorities, collective rights management entities (Collective Management Organisations - CMOs and independent Management Entities - IMEs), and a wide range of other public and private operators (including publishers, registries, platforms and service providers). Databases perform a variety of roles (including reference systems, vaults, marketplaces,

distributors, exchanges and collectors), and they commonly rely on permissioned access models that differentiate administrators, affiliates and public users.

The study finds that databases primarily rely on internal identifiers followed, where available, by sectoral or international identifiers; fingerprints and watermarks are used by fewer operators. Operators generally store information on authors, creators or other rights holders associated with a work (or agent information) on a need-to-know basis and rely on declarer trust for authenticity. While sector-specific metadata schemas prevail, international schemas remain less widely adopted.

Interconnection practices remain fragmented and sector specific. Operators exchange data through methods ranging from email and comma-separated values (CSV) transfers to application programming interfaces (APIs); most interconnections follow star-topology architectures, while fewer adopt networked models, with patterns varying significantly across sectors. The study highlights how metadata gaps and poor data linking contribute to inefficiencies (including remuneration challenges in music), and how the lack of standard identifiers in certain sectors (e.g. images) constrains interoperability and traceability.

Overall, while existing interconnections facilitate basic data exchange, they do not support seamless interoperability or consistent reuse of metadata across sectors and jurisdictions, resulting in data silos, duplication of records, inconsistencies in rights information, and increased operational costs and delays in processes such as rights clearance, licensing and remuneration.

Part 4 – Opportunities for the ECKC and CopyrightView

Building on the diversity of database mandates, metadata practices and the interconnection models identified in Part 3, stakeholders identified opportunities for the EUIPO Copyright Knowledge Centre (ECKC), the new EUIPO platform to support copyright knowledge, stakeholder dialogue and policy support, as well as for the development of CopyrightView.

For the ECKC, interviewees highlighted the value of maintaining an up-to-date overview (or 'mapping') of the different databases, identifiers and metadata schemas, to promote standardisation and improved workflows, facilitate cross-sector knowledge transfer and support broader communication on copyright as an enabler of access to quality content.

For CopyrightView, stakeholders emphasised potential benefits, including more flexible and interoperable access to copyright information, support for contact and licensing workflows (in particular by facilitating contact with relevant rights holders rather than aiming to provide a complete, authoritative record of all rights), improved handling of TDM reservations including by facilitating the effective exercise and communication of rights reservations (opt-out), and EU-level licence knowledge, increased visibility of national library databases beyond national borders, and support for deduplication and reciprocity between interconnected organisations.

Based on interview feedback, the study identifies two initial use cases for further discussion: finding information ('Where will I find information about that content?') and matching information (reconciling and validating RMI across databases, including completion, deduplication, monitoring/remuneration and conflict-resolution contexts). These use cases could significantly improve the discoverability of copyright information, facilitate contact with relevant rights holders, and reduce the time and costs associated with rights clearance and licensing. They could also enhance data quality and consistency across databases, supporting more accurate attribution, remuneration and dispute resolution.

Part 5 – Feasibility

The study identifies the potential benefits associated with improved access to reliable and interoperable copyright information, including evidence of authorship and ownership (supporting conflict resolution and transactional contexts such as licences and transfers), improved rights clearance, distribution of revenue, and accurate crediting of authors and performers. At the same time, desk and field research identify organisational obstructions that may hinder the delivery of these benefits, including the limited budget and human resources among database operators and other stakeholders, lengthy decision-making processes, limited operator scope, legacy systems unsuited to interconnection, and formal or legal obstacles to data sharing.

Feasibility considerations are addressed at a high level across four key dimensions: governance (roles, responsibilities and sustainability); legal (data sharing conditions, mandates and compliance); operational (controlled access, misuse risks and dispute handling); and technical (heterogeneous systems, scalability, standardisation and quality controls). In this context, the study presents three technical approaches for further consideration regarding their adaptation, combination or extension: a watermark-based lookup mechanism, a layered architecture for federated access, and a

fingerprint-based discovery protocol. Stakeholder feedback also suggests that referencing might be preferable to full integration, and that lighter designs could facilitate deployment, particularly when focused on upstream data close to the point of creation.

Overall, the study concludes that a rich but fragmented ecosystem of databases, standards and initiatives already exists. It provides a structured overview of the European copyright metadata landscape, highlighting key gaps and opportunities to improve interoperability and access to rights information. These findings will inform the next phases of the CopyrightView project, including the prioritisation of use cases and the development of a dedicated feasibility analysis in close cooperation with stakeholders. The key challenge remains to connect these elements in a way that is adapted to sectoral diversity and stakeholders' constraints, anchored in clear use cases, and supported by governance arrangements that enable trust, controlled access and sustainable interoperability.