MINISTRY OF
EDUCATION AND CULTURE
FINLAND

Interoperable, trustworthy, and machine-readable copyright data in the AI era

Report of the CITF First Project

Niko Partanen, Philippe Rixhon, Karīna Bandere, Jānis Ziediņš, Pawan Kumar Dutt, Matīss Bolšteins, Matias Frosterus, Mona Lehtinen, Inta Miklūna-Žukeviča, Deniss Ozerskis, Päivi Maria Pihlaja, Jogita Sauka, Katerina Sornova, Aija Uzula

CULTURE | Publications of the Ministry of Education and Culture: 2025:23

Publications of the Ministry of Education and Culture, Finland 2025:23

Interoperable, trustworthy, and machine-readable copyright data in the AI era

Report of the CITF First Project

Niko Partanen, Philippe Rixhon, Karīna Bandere, Jānis Ziediņš, Pawan Kumar Dutt, Matīss Bolšteins, Matias Frosterus, Mona Lehtinen, Inta Miklūna-Žukeviča, Deniss Ozerskis, Päivi Maria Pihlaja, Jogita Sauka, Katerina Sornova, Aija Uzula

Ministry of Education and Culture Helsinki 2025

Publication distribution

Institutional Repository for the Government of Finland Valto

julkaisut.valtioneuvosto.fi

ISBN pdf: 978-952-415-143-6
ISSN pdf: 1799-0351

Layout: Government Administration Department, Publications

Helsinki 2025 Finland

Description sheet
3 October 2025

Interoperable, trustworthy, and machine-readable copyright data in the AI era
Report of the CITF First Project

Publications of the Ministry of Education and Culture, Finland 2025:23		**Subject**	Culture
Publisher	Ministry of Education and Culture		

Author(s)	Niko Partanen, Philippe Rixhon, Karīna Bandere, Jānis Ziediņš, Pawan Kumar Dutt, Matīss Bolšteins, Matias Frosterus, Mona Lehtinen, Inta Miklūna-Žukeviča, Deniss Ozerskis, Päivi Maria Pihlaja, Jogita Sauka, Katerina Sornova, Aija Uzula
Group author	National Library of Finland, National Library of Latvia, Culture Information Systems Centre (Latvia), Tallinn University of Technology, Valunode OÜ

Language	English	**Pages**	109

Abstract

The First Project of the Copyright Infrastructure Task Force (CITF) analysed requirements for a common copyright infrastructure in the Member States of the European Union. It was coordinated by the National Library of Finland and carried out in collaboration with the National Library of Latvia, the Culture Information Systems Centre (Latvia), Tallinn University of Technology (Estonia), and Valunode OÜ (Estonia) in a consultant role. The project defined measures to support the copyright ecosystem in the digital and AI era, including the use of open identifiers, standard metadata schemas, and emerging technologies as a basis for open dialogue with all relevant stakeholders.

The National Libraries hold in their collections materials and metadata that are among the largest that exist in each country, and that are mostly protected by copyright law. They also administer various identifiers. This positions them as part of the wider copyright infrastructure. Besides the National Libraries' mandates, the project also considered commercial licensing and searches across existing repositories.

The project discussed the life cycles of five use cases – a doctoral thesis, a scientific article, an eBook, an image, and an article in a periodical (Annex 2) – and analysed how they are impacted by new developments in opt-out expressions and generative AI. The life cycles were used to define and comment on requirements (Annex 3). The terminology used in the project is explained in Terms and Definitions (Annex 1).

Keywords	copyright, infrastructures, digital media, intellectual capital, metadata, identifiers, artificial intelligence, creative sectors		
ISBN PDF	978-952-415-143-6	**ISSN PDF**	1799-0351
URN address	https://urn.fi/URN:ISBN:978-952-415-143-6		

Kuvailulehti

3.10.2025

Yhteentoimiva, luotettava ja koneluettava tekijänoikeustieto tekoälyn aikakaudella
Tekijänoikeusinfrastruktuuriverkosto CITF:n ensimmäisen hankkeen raportti

Opetus- ja kulttuuriministeriön julkaisuja 2025:23		**Teema**	Kulttuuri
Julkaisija	Opetus- ja kulttuuriministeriö		

Tekijä/t	Niko Partanen, Philippe Rixhon, Karīna Bandere, Jānis Ziediņš, Pawan Kumar Dutt, Matīss Bolšteins, Matias Frosterus, Mona Lehtinen, Inta Miklūna-Žukeviča, Deniss Ozerskis, Päivi Maria Pihlaja, Jogita Sauka, Katerina Sornova, Aija Uzula
Yhteisötekijä	Kansalliskirjasto, Latvian kansalliskirjasto, Culture Information Systems Centre (Latvia), Tallinnan teknillinen yliopisto, Valunode OÜ

Kieli	englanti	**Sivumäärä**	109

Tiivistelmä

Tekijänoikeusinfrastruktuurin kehittämisverkoston (CITF) ensimmäisessä hankkeessa analysoitiin vaatimuksia yhteiselle tekijänoikeusinfrastruktuurille Euroopan unionin jäsenvaltioissa. Hanketta koordinoi Kansalliskirjasto ja se toteutettiin yhteistyössä Latvian kansalliskirjaston, Culture Information Systems Centre (Latvia), Tallinnan teknillisen yliopiston (Viro) ja Valunode OÜ:n (Viro) kanssa. Hankkeessa määriteltiin toimenpiteitä, joilla tuetaan digitaalisen tekijänoikeusekosysteemin kehittämistä tekoälyn aikakaudella, kuten avointen tunnisteiden, metadatastandardien ja kehittyvien teknologioiden käyttöä pohjustamaan avointa vuoropuhelua kaikkien relevanttien sidosryhmien kanssa.

Kansalliskirjastoilla on kokoelmissaan aineistoja ja metatietoja, jotka ovat kussakin maassa suurimpien joukossa, ja pääosin tekijänoikeuslainsäädännön suojaamia. Ne hallinnoivat myös erilaisia tunnisteita. Tätä kautta Kansalliskirjastot kytkeytyvät laajempaan tekijänoikeusinfrastruktuuriin. Kansalliskirjastojen tehtävien lisäksi hankkeessa tarkasteltiin myös kaupallista lisensointia ja olemassa olevien tietokantojen välillä tehtäviä hakuja.

Hankkeessa käsiteltiin viiden käyttötapauksen – väitöskirjan, tieteellisen artikkelin, sähkökirjan, valokuvan ja kausijulkaisun artikkelin (liite 2) – elinkaaria ja analysoitiin miten opt-out-ilmaisuja ja generatiivista tekoälyä koskevat uudet kehityskulut vaikuttavat niihin. Elinkaaria käytettiin vaatimusten määrittelyssä ja kommentoinnissa (liite 3). Termit ja määritelmät (liite 1) selventää käytettyjä käsitteitä.

Asiasanat	tekijänoikeus, infrastruktuuri, digitaalinen media, aineeton pääoma, metadata, tunnisteet, tekoäly, luovat toimialat
ISBN PDF	978-952-415-143-6
ISSN PDF	1799-0351
Julkaisun osoite	https://urn.fi/URN:ISBN:978-952-415-143-6

Presentationsblad

3.10.2025

Interoperabel, pålitlig och maskinläsbar upphovsrättsdata för AI eran
Rapport från det första CITF-projektet

Undervisnings- och kulturministeriets publikationer 2025:23		**Tema**	Kultur
Utgivare	Undervisnings- och kulturministeriet		
Författare	Niko Partanen, Philippe Rixhon, Karīna Bandere, Jānis Ziediņš, Pawan Kumar Dutt, Matīss Bolšteins, Matias Frosterus, Mona Lehtinen, Inta Miklūna-Žukeviča, Deniss Ozerskis, Päivi Maria Pihlaja, Jogita Sauka, Katerina Sornova, Aija Uzula		
Utarbetad av	Nationalbibliotek, Lettlands Nationalbibliotek, Culture Information Systems Centre (Latvia), Tallinn University of Technology, Valunode OÜ		
Språk	engelska	**Sidantal**	109

Referat

Det första projektet inom Nätverket för utveckling av upphovsrättsinfrastrukturen (CITF) analyserade kraven på en gemensam upphovsrättsinfrastruktur i Europeiska unionens medlemsstater. Projektet koordinerades av Nationalbiblioteket och genomfördes i samarbete med Lettlands Nationalbibliotek, Culture Information Systems Centre (Lettland), Tallinns tekniska universitet (Estland) och företaget Valunode OÜ (Estland) som medverkade som konsult. Projektet identifierade åtgärder för att stödja utvecklingen av ett upphovsrättsekosystem för AI-eran. Användning av öppna identifierare, metadatastandarder och ny teknologi lyftes fram som sätt att bana väg för en öppen dialog med alla relevanta intressentgrupper.

Nationalbiblioteken har i sina samlingar material och metadata, som hör till de största i varje land, och som i huvudsak skyddas av upphovsrättslagstiftningen. Biblioteken administrerar också olika typer av identifierare, vilket gör dem till en del av den mer omfattande upphovsrättsinfrastrukturen. Utöver de nationella bibliotekens uppdrag granskade projektet också kommersiell licensiering och sökning mellan befintliga databaser.

Projektet behandlade livscykeln för fem användningsfall – en doktorsavhandling, en vetenskaplig artikel, en e-bok, ett fotografi och en artikel i en tidskrift (bilaga 2) – och analyserade hur de påverkas av nya utvecklingar inom opt-out-uttryck och generativ artificiell intelligens. Livscykeln användes för att definiera och kommentera kraven (bilaga 3). Termer och definitioner (bilaga 1) förklarar begreppen som används.

Nyckelord	upphovsrätt, infrastrukturer, digital media, immateriellt kapital, metadata, identifikatorer, artificiell intelligens, kreativa sektorer		
ISBN PDF	978-952-415-143-6	**ISSN PDF**	1799-0351
URN-adress	https://urn.fi/URN:ISBN:978-952-415-143-6		

Contents

FOREWORD

The copyright infrastructure is generally understood to mean a set of rules, technologies and institutions that frame the management of copyright data in the creative sectors. In the AI era modernizing the copyright infrastructure is a crucial component to make copyright law work in practise.

The First Project of the Copyright Infrastructure Task Force (CITF) was funded in 2024 through the Ministry of Education and Culture's grant programme for developing the copyright system. This report is therefore published in the Ministry's official publication series.

The initiative builds on work started during Finland's EU Council Presidency in 2019, which led – at the invitation of the Estonian State Secretary in 2022 – to the creation of the multi-country Copyright Infrastructure Task Force. Another publication in this series is Impact of the General Data Protection Regulation and National Personal Data Protection Legislation on the Copyright Infrastructure (2022:38)

In September 2024, the Ministry announced a call for proposals to prepare a use case on requirements relevant for copyright and AI. In November 2024, the National Library of Finland was selected to lead the project and was awarded a grant of EUR 164,400 for this purpose. Participating organisations from Estonia and Latvia were not required to contribute financially but their roles were defined in a Project Collaboration Agreement supporting the CITF work plan. The content of this report is the responsibility of the authors and does not necessarily represent the views of the Ministry of Education and Culture.

The project's results especially relating to technical requirements for machine readable reservations of rights and for identification of AI-generated content, as detailed in the report and in the annexes, are intended to serve as a basis for further discussion and feedback among stakeholders as well as EU and international institutions.

Anna Vuopala
Senior Ministerial Adviser
Ministry of Education and Culture

1 Objectives

This First Project of the Copyright Infrastructure Task Force (CITF) defines requirements to be considered for a future joint project to develop a common copyright infrastructure in the Member States of the European Union. The Project defines measures to promote the functioning of the copyright ecosystem in the digital environment, including the use of open identifiers and new technologies.

The Project also aims to support information and training activities related to generative artificial intelligence, in particular the testing of relevant copyright provisions, such as the exception concerning text and data mining and mechanisms to reserve rights in a machine-readable manner (the "opt-out declaration").

The scope – **AI and copyright challenges** – has been chosen in consultation with the European Commission, discussed during several CITF monthly meetings, and presented at the Copyright Conference organised by the Belgian Presidency of the EU in April 2024.

Considering the following legislation:

- EC directive 2001/29 on the harmonisation of certain aspects of copyright and related rights in the information society;
- EU directive 2019/790 on copyright, Articles 3 and 4 (text and data mining), and corresponding provisions in Finnish, Estonian, and Latvian copyright laws, and
- the Artificial Intelligence Act, requiring respect of copyright law – specifically reservations of rights, transparency on how copyright-protected content is used for training AI models, identification of AI-generated content, and protection against disinformation.

Three clusters of requirements have been looked at:

- **Information:** standardisation of interoperable identifiers and opt-out declarations;
- **Traceability:** identification of content that has been generated by AI, and

- **Transparency:** documentation of the training and generation algorithms used to produce AI-generated content.

In this report priority is given to the section on Information, while Traceability and Transparency are of intrinsic importance to the objectives and activities of the CITF.

The project has two main goals:

- To define terms, use cases, and requirements related to the trustworthiness, interoperability, and machine-readability of rights management information – the object of the present **Report**, and
- To assess and validate the modus operandi of the project activities of the CITF as described in the Proposition[1] to establish the CITF at EU level submitted to the Cabinet of Ms Henna Virkkunen, Executive Vice-President for Tech Sovereignty, Security and Democracy at the European Commission – the object of separate **Working Papers**.

The project builds on Finland's Digital Compass (2023), the EU Commission's Study on Copyright and New Technologies (2022), the Finnish Government's Decision of Principle on the National Intellectual Property (IP) Strategy (2022), Finland's Sustainable Growth Programme (2021), and the EU Commission's Intellectual Property Rights Action Plan (2020). The development work was launched as part of the Finnish EU Presidency initiative to develop copyright infrastructure at EU level (2019).

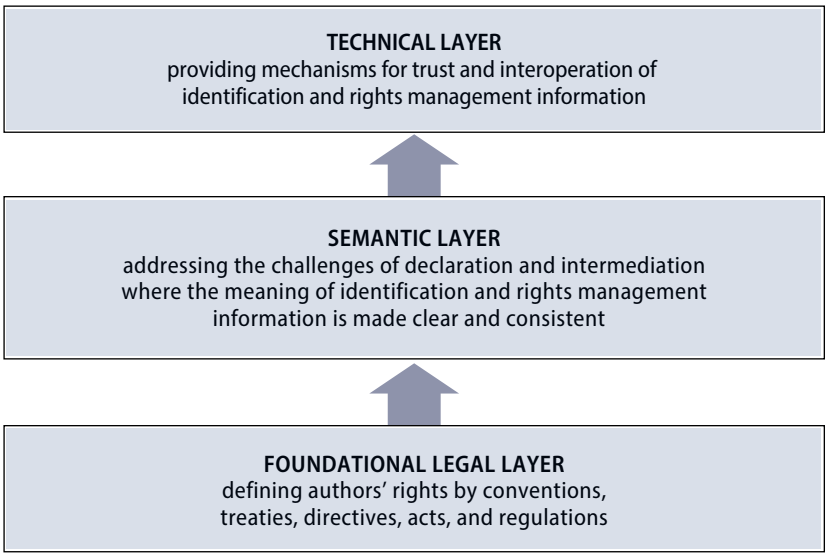
1 https://api.hankeikkuna.fi/asiakirjat/40decb0b-df3d-4002-b1ff-5c5e1b4de509/f240e554-1338-43d9-9d42-5bd56eaefc53/LIITE_20240312184041.pdf

2 Context

It is estimated that the EU media market earns € 671.5 B p.a. and represents 3.95% of EU GDP. It enjoys fast yearly growth of 5%. It involves 8 M creators, 1.2 M small and medium enterprises, and 4 M consumers. It is a pillar of democracy and culture and, at the same time, an extensively disrupted market[2].

A copyright infrastructure would facilitate the implementation and enforcement of the Acquis Communautaire. It would enable creators to protect and monetise their rights. And it would pave the way for a true Digital Single Market where open flows of data would benefit all stakeholders.

Figure 1. The three layers of a copyright infrastructure



2 Sources: https://single-market-economy.ec.europa.eu/sectors/cultural-and-creative-industries_en and https://european-union.europa.eu/principles-countries-history/facts-and-figures-european-union_en

This Project is not about copyright law, the normative framework that provides exclusive rights to authors and other rightsholders. It is about metadata – data about other data – i.e., the semantic layer of a copyright infrastructure. The Project focuses on "rights metadata" – elements related to identification, attribution, and terms and conditions. Within the intellectual property ecosystem, expressing usage conditions (e.g., opting out from text and data mining) is an exercise of authors' exclusive rights.

The cultural and creative industries will benefit from a well-functioning copyright infrastructure as described in the Stocktaking document of the Working Party on Intellectual Property published by the Council of the European Union in 2019, and from an open rights data framework as mentioned in the Study on Copyright and New Technologies published by the European Commission in 2022.

Opening and integrating the rights data framework in close collaboration with all relevant parties will boost the value of the content sectors significantly over time. The task is to make the framework trustworthy, interoperable, automated, and as accessible and comprehensive as possible. Copyright infrastructure development takes into account the diversity of business models and market segments and focuses on areas of common interest to all sectors.

2.1 The Copyright Infrastructure Task Force

In November 2023, building on the work of the previous five years[3], the Copyright Infrastructure Task Force (CITF) was formed by four EU Member States (Estonia, Finland, Latvia, and Lithuania)[4]. Today, the CITF also includes participants from other EU and EFTA Member States (Belgium, Czechia, France, Germany, Hungary, Italy, Spain, Greece, Romania, Norway, and Iceland). The CITF acts as a forum to promote the interoperability and trustworthiness of copyright data, i.e., the open rights data framework (ORDF). The CITF meets online once a month under the leadership of

3 DSM directive 2019/790, the Council's document 15016/19 on the Copyright Infrastructure summarising the progress made during the Finnish Presidency in 2019, the IP Action Plan 2020, and the recommendation of the 2022 Study on Copyright and New Technologies to build an Open Rights Data Framework (ORDF).

4 The Member States had worked since 2022 under the guidance of the Commission's multi-country project unit to form a European Digital Infrastructure Consortium on Copyright Infrastructure development (CI EDIC) at EU level. The Commission's copyright unit acted as an observer of the process.

Finland. The Commission has been observing and supporting the process since 2022. Philippe Rixhon (Valunode OÜ) is participating in the CITF as a technical expert and consultant.

At the end of 2023, the Commission's copyright unit recommended that the CITF start working on practical issues[5] related to the development of the copyright infrastructure. The Commission suggested developing a use case for a machine-readable "opt-out" expression based on Article 4 (Exception or limitation for text and data mining) of the DSM Directive. This use case would be very important considering the obligations of providers of General Purpose Artificial Intelligence (GPAI) models detailed in Article 53 of the EU AI Act ((EU) 2024/1689). Furthermore, the Commission suggested that the use case be developed in close co-operation with Europeum EDIC[6], tasked with deploying the European Blockchain Service Infrastructure (EBSI) in the EU.

By early summer 2024, the document "Defining the AI & Copyright Use Case" had been developed by the CITF and posted on the openly accessible Copyright Infrastructure portal[7]. This document also presented some exemplary use cases to pilot the opt-out mechanism in certain creative sectors.

The CITF was acknowledged in the programme of the Hungarian Presidency as a relevant tool for Member States to cooperate towards the development of a Copyright Infrastructure. On 19 September 2024, the CITF met in-person with the copyright unit of the Commission (DG CNCT) in Brussels to discuss the current state of play and ways to impact the future work plan of the new Commission.

In November 2024, the Finnish Ministry of Education and Culture granted funding to coordinate the work of public institutions in Estonia, Latvia, and Finland to plan the details of the use cases. This First Project ran until June 2025.

5 The recommendation was based on the assessment that instead of a new legal body, there was a need for more focused and concrete copyright infrastructure outcomes in the short term.

6 The Europeum EDIC was established in May 2024 on application by Croatia, Cyprus, Greece, Italy, Luxembourg, Portugal, Romania, and Slovenia.

7 The Copyright Infrastructure portal, founded in 2020, supports awareness and coordination on national, EU, and global level industry-specific initiatives to promote the proper functioning of the copyright system, in particular the formation of an Open Rights Data Framework (ORDF) for new technologies.

During the Polish presidency, the CITF reported on the progress of the First Project of the CITF at the Council working party meetings, presented the preliminary outcomes of the Project at the CITF Seminar on 16 June, contributed to the Presidency policy questionnaire on CMOs and AI, and arranged a debate on 17 June 2025 about the establishment of a mandating framework for the CITF at EU level.

Alongside the aforementioned use cases, the CITF continues its work defining a strong basis for the EU level Copyright Infrastructure, and validating the results for the use cases mentioned above at the Council working party meetings on copyright, in cooperation with the incoming Danish and Cypriot Presidencies.

Several bilateral meetings were held from 2023–2025 with different industry and stakeholder representatives about the work of the CITF with the aim of ensuring coherence and support. The CITF has also invited representatives of the Trusted Media Data Space (TEMS) and EARE (the European Alliance for Research Excellence)[8] to participate. The CITF has also liaised with the ALT EDIC (the Alliance for Language Technologies)[9] which was established in February 2024 in Paris.

The CITF promotes global, regional, and national work and dialogue. The Copyright Infrastructure also featured as the main topic of the WIPO (World Intellectual Property Organization) AI Conversation on 11 April 2025. The underlying data infrastructure is recognised as an indispensable technical element to ensure that copyright works in practice. On 8 December 2025, the AIII – Artificial Intelligence Infrastructure Interchange – will be launched at global level alongside normative WIPO activities.

2.2 The role and impact of National Libraries

National Libraries are cultural heritage organisations that are open to all and provide nationwide services to citizens, scientific communities, and other societal operators. They collect, preserve, and provide access to a nation's documentary and cultural heritage, serving citizens, researchers, and other key sectors of society.

8 The Trusted Media Data Space (TEMS) aims for a Common EU Data Space for Media; EARE (the European Alliance for Research Excellence) aims to ensure synergies in data sharing, capacities, and interoperability in the common European Digital Single Market.

9 ALT EDIC seeks, together with its 16 members and observers, to improve European competitiveness, increase the availability of European language data, and uphold Europe's linguistic diversity and cultural richness. The availability of language data is relevant from a copyright perspective as it involves content protected by copyright.

National Libraries were included in this First Project as they hold materials in their collections that are relevant for research, public use, and AI development and which are to a large degree protected by copyright law. They also have expertise in metadata and metadata management as they have developed highly sophisticated management systems for their collections. National Libraries are not only metadata experts, but also have deep bibliographical expertise – they understand the structure, classification, and historical development of publications, which is essential for managing copyrighted works and supporting discovery, research, and rights identification.

The role of a National Library varies from country to country, but there are some general similarities. They are primarily cultural heritage institutions that store and catalogue the national published cultural heritage. The scope of materials covered may differ, but usually various types of printed publications are included. Depending on the national legislation, some National Libraries also store collections of data harvested from the internet, and may also include various audio, audio-visual, and image collections. In the Finnish and Latvian context, the National Libraries secure the availability of cultural heritage published in or relating to their respective countries, and transmit and produce information content for research, studies, citizens, and society. Both the National Library of Finland and the National Library of Latvia develop services in cooperation with libraries, archives, museums, and other operators. These services and related metadata intersect in different ways with the larger copyright infrastructure. This position is the point of view taken in this report.

The metadata of the National Libraries is as complete as possible, following several standards and practices of the field. This report, together with the use cases and requirements presented in the appendices, presents the current practices of the participating institutions related to selected examples of works. The new legislative demands concerning opt-out expressions and indications of generative AI use, which came into force in August 2025, may require that in some cases the author is strongly identified. The current infrastructure does not have immediate solutions to these demands. In environments not directly connected to the National Libraries and their metadata, such as when works are made commercially available by the relevant industries, the emerging requirements appear even more extensive.

As cultural heritage organisations, National Libraries also have the obligation to make their collections available for research use. For example, both the National Library of Finland and the National Library of Latvia offer their legal deposit collections to researchers and other users, in line with the provisions of the Legal Deposit Act, within the scope of the exceptions and limitations defined in the

respective national legislation. The National Libraries offer their data to the public to fulfil their obligations under a defined role in the cultural heritage sector. Commercial use of these materials takes place in markets, and the terms of use are negotiated between authors, publishers, and other parties; they will need to address the issues of concern raised in this report.

The metadata that National Libraries receive may contain diverse rights management information (see the footnote to Art. 7 of the Information Society Directive). However, license information is not in itself necessarily part of the publicly available bibliographic record, which is made openly available and can be accessed by anyone. For it to be possible to query and combine more extensive information about these works, various technical interconnections between infrastructures need to be envisioned, while taking into account the roles and responsibilities of different organisations and sectors. Available persistent identifiers are used broadly to make it possible to connect different items to one another. The library sector has been rapidly moving towards a linked data framework, and it can be envisioned that e.g. author and work identifiers in library metadata will be used in future in different applications, including in commercial contexts outside their traditional context. In the Latvian context, the Culture Information Systems Centre, another partner in this report, has implemented a solution whereby copyright and license information is more extensively incorporated into the cultural heritage sector's metadata.

Recently, the National Libraries have begun assigning ISNIs to authors represented in their collections, often in cooperation with the collective management organisations (CMOs) that represent those authors. While this development demonstrates the potential of ISNI to enhance identity management and data interoperability, it also highlights a critical limitation: access to ISNI infrastructure typically requires a financial investment, which means that not all National Libraries can participate or offer this service systematically.

The catalogues of National Libraries are the most extensive metadata collections concerning the published cultural heritage in a given country. The metadata is openly available, and the interoperability of this data is a high priority. Standardised identifiers such as ISBN, ISSN mostly, and ISNI in some cases, are assigned and used to ensure that data are interoperable and reusable. This data is based on the information the authors and publishers provide, and what is indicated in the publications themselves. The National Library is not responsible for the accuracy of this data as it is based on information that is delivered, and the publicly available metadata cannot be directly connected to real persons. In the context of the National Library of Finland, the processing of personal data is based, above all, on

Article 6 of the EU General Data Protection Regulation (2016/679) and the following paragraph of Section 4 of the Data Protection Act (1050/2018): *the processing of research material and cultural heritage material containing personal data and the processing of personal data included in their metadata for archiving purposes is necessary and proportionate to the aim of public interest pursued and to the rights of the data subject*.

Based on the exceptions and limitations in the copyright legislation, works in the collections of National Libraries that are still under copyright can be made available to the public under certain specific conditions. Additional contracts are often entered into with collective management organisations (CMOs) that increase the amount of digitised material that can be displayed to the public on the different online platforms of the National Libraries. At the same time there have been projects in Latvia that increase access to content on the basis of license agreements with the publishers. Under these initiatives works are usually available to the public, but not for uses such as AI development.

It must be recognised that many current developments may impact how metadata is managed in the future. For example, born-digital works in the future could contain signed and timestamped metadata about the authors and the exact modifications they have made, and this information should be kept within the metadata of those items. At the same time, it does not seem necessary that this information be part of the bibliographic record in itself. Although in the case of an electronic work both the file and the bibliographic record are in the National Library collection, and the latter is openly available, the availability of the former is controlled by the conditions outlined above.

Several identifiers, such as ISNI, are designed to bridge metadata and resources between different databases, and the National Library of Finland has initiated several projects with collective management organisations to increase the use of ISNI to connect the databases of different organisations. This effectively illustrates the possible role National Libraries could have in developing the EU's copyright infrastructure: their catalogues are intended to be used and to be interoperable with other databases, but the structural limitations to the information that the publicly available open catalogue can contain must be also taken into account.

2.2.1 National Libraries and opt-out information

National Libraries have already started receiving electronic publications with opt-out information, meaning the author explicitly prohibits the use of their material for generative AI development purposes. The current international library infrastructure is not able to store such opt-out statements in the standardised library metadata. Similarly, the relation between the opt-out and the public bibliographic record may be similar to that of license information and the bibliographic record, where it is not always part of the bibliographic record, and future developments need to be monitored carefully. Although the materials of National Libraries are generally not available for the commercial use that the opt-out expression specifically aims to exclude, it can be reasonably assumed that national metadata catalogues would need to be able to interoperate with opt-out mechanisms in different circumstances.

For instance, if under some future agreement some materials are made available to the public in new ways, in certain instances the availability of opt-out information would be crucial to specify which restrictions apply to which works. At the same time, whether National Libraries need to store opt-out data in their databases also depends on larger questions of how the opt-out will eventually be implemented and how the copyright infrastructure will be formed overall. If there is some database where opt-out information can be queried, then increasing interoperability in this direction would be a beneficial objective as well. Also, in this scenario the identifiers managed by the National Libraries would inevitably have a significant role within the emerging infrastructure.

The National Library of Latvia operates within the legal and ethical framework of cultural heritage preservation, while also respecting the rights of authors and rightsholders. In accordance with national legislation and international best practices, the Library applies an opt-out approach in certain areas of digital access and reuse. This means that, for materials considered to be in the public interest or part of the national cultural record – particularly digitised heritage and bibliographic data – the Library may make content publicly available unless the rightsholder explicitly requests its removal or restriction.

2.2.2 Distribution services of the National Libraries

Especially in the Finnish and Latvian context, the National Libraries also offer services for the wider national cultural heritage sector. These services, Finna[10] and the Digital Library of Latvia[11], make available millions of cultural heritage items from archives, museums, libraries, and other organisations. Unlike the collections of the National Libraries, these materials are provided by many different organisations with varying metadata formats, licenses, and terms and conditions. It seems certain that these collections contain opt-out statements made by authors that would need to be carried into the metadata of the responsible organisations and displayed on the distribution platforms of the National Libraries. It must be noted that these services focus strongly on open access materials.

The emergence of generative AI is another dimension that the National Libraries must take into account. Standardised practices regarding ways to indicate which parts of a work have been produced by an AI system are also important for the preservation of the cultural heritage, so that works expressing human creativity and heritage can be kept distinct from automatically generated content, while taking into account that AI-generated publications may also need to be preserved. This also concerns services such as Finna and the Digital Library of Latvia. Knowing whether material was made by a human or generated by AI is important for the trustworthiness of the media and information in general.

When materials are distributed by National Libraries in the manner described above, it is crucial that the human- and machine-readability of the data (content) and metadata is ensured. If an AI system queries the data through an API, the conditions of use must be automatically processable, whether the information derives from the metadata or other databases where this information can be reliably queried.

10 https://www.finna.fi/

11 https://digitalabiblioteka.lv/

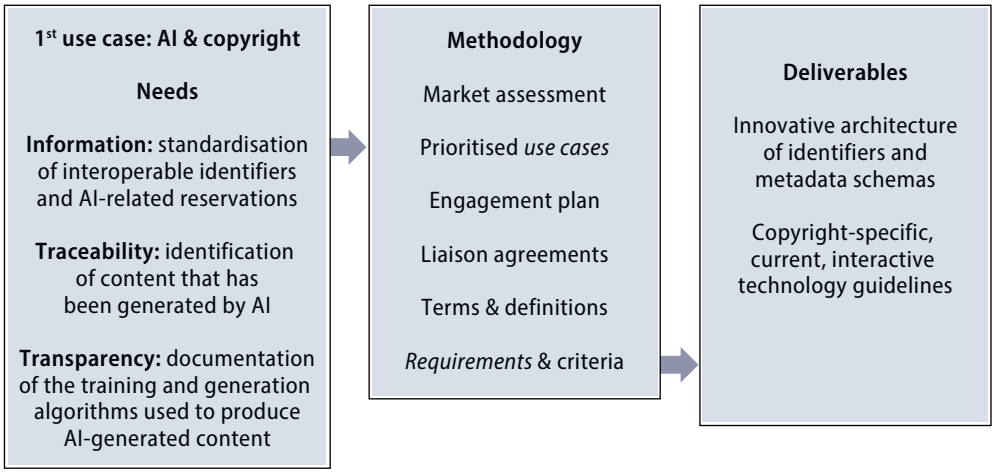
3 Methodology

The CITF is a forum dedicated to identifying and promoting the standards and technologies needed to increase the trustworthiness, interoperability, and machine-readability of rights data.

In the future, after validating the results of the First Project, a working group on identifiers and schemas will deliver an innovative architecture of identifiers and schemas. Another working group on technologies for interoperability, searchability, and trustworthiness will deliver copyright-specific, current, and interactive developer guidelines.

The CITF is a pre-standardisation forum rather than a standards developing organisation (SDO). It is open to anyone interested – representing Member States or stakeholders – and liaises with SDOs. During this First Project, the CITF liaised with CEN-CENELEC, ISO / JPEG Trust, and W3C. It also presented its objectives, methodology, and this First Project at the European Multi-Stakeholder Platform on ICT Standardisation.

Figure 2. CITF methodology and deliverables



The First Project of the CITF prioritises the need for **Information** over the need for **Traceability** and **Transparency**.

The CITF adopts and promotes best standardisation practices. A **market assessment** – here the emergence of Generative AI and its impacts on rightsholders – leads to prioritised **use cases** – here the consideration of information, traceability, and transparency.

Requirements and criteria are defined to address, i.e., solve, the use cases. A pre-standardisation forum, such as an SDO, collaborates with related stakeholders, which necessitates an **engagement plan** and **liaison agreements**.

The project uses templates from:

- The ISO/JPEG Trust for use cases, terms & definitions, and requirements, and
- The Metaverse Standards Forum for use cases, engagement plans, terms & definitions, and requirements.

The research team includes:

- The National Library of Finland, project coordinator
- The National Library of Latvia
- The Culture Information Systems Centre (Latvia)
- Tallinn University of Technology (Estonia), IP law
- Valunode (Estonia), technical consultant

The Ministry of Education and Culture did not participate in the project but was consulted as needed. The participants of the CITF were regularly informed of the progress of the First Project at the monthly CITF meetings.

3.1 Process of the First Project

Step 1: Identify, select, and document five media objects in terms of copyright infrastructure, asset traceability, and potential expression of AI reservations, and collect available metadata.

Step 2: Describe the 2025 life cycles (from creation of the work to consumption of its digital manifestations) of the selected objects, and the current (digital) infrastructure of the National Library of Finland and the Digital Cultural Heritage Platform in Latvia.

Step 3: Develop potential scenarios considering that in 2030:

- All content will be digitised and somehow available online;
- Only machines will deal with rights and royalty management;
- AI applications will grow and multiply, and
- Creators of original content must be fairly, appropriately, proportionally, and transparently remunerated.

Step 4: Define concise requirements needed to meet criteria for a system and/or protocols that support secure management of digital assets which may not be limited to the following topics:

- Interrelations between internal author IDs used by libraries, an international identifier such as ISNI, and the upcoming EU implementation of eIDAS 2.0;
- Interrelations between work IDs used by libraries and international standards such as DOI, ISBN, ISSN, and ISNI;
- Expression, storage, retrieval, and interpretation of AI-related reservations;
- Interconnection of the registers of National Libraries with a focus on their data model and high-level architecture.

During step 4, the First Project of the CITF organised four **workshops** on the following topics:

- **Agent identifiers,** facilitated by Nicolas Hauw (WIPO) and Will Kreth (Hand);
- **Asset identifiers,** facilitated by Paola Mazzucchi (mEDRA, the multilingual European DOI Registration Agency) and Titusz Pan (ISCC Foundation);
- **Assertions,** facilitated by Joshua Cornejo (MarketData) and Leonard Rosenthol (C2PA), and
- **Interconnection,** facilitated by Regis Flad (ISAN International Agency), Marko Turpeinen (1001 Lakes), and Lucille Verbaere (European Broadcasting Union).

The facilitating experts presented the latest developments in their domain of activity and answered the questions of the research team. They are herewith thanked for their important contributions.

Step 5: Present the results of the project at a seminar in Brussels with the following agenda:

- **Report on the First Project of the CITF:** Interoperable and trustworthy copyright data: Exploring technical requirements for copyright infrastructure in a 2030 scenario
- **Fireside chat:** Trust in media – a societal challenge for Europe
- **Panel discussion:** Give up, opt out, or license? Options for creators in the era of Generative AI

The presentation of the report and the link to the video recording of the seminar were forwarded to the participants and are available upon request.

Step 6: Publish the report and CITF working documents, which include among others the current CITF charter and the engagement plan for the First Project.

4 Terms and definitions

To ensure a correct understanding of this report of the First Project of the CITF, **Annex 1** defines terms and concepts as they are used in this context. In the following sections and annexes defined terms are signalled using small capitals.

The terms and definitions come from different sources:

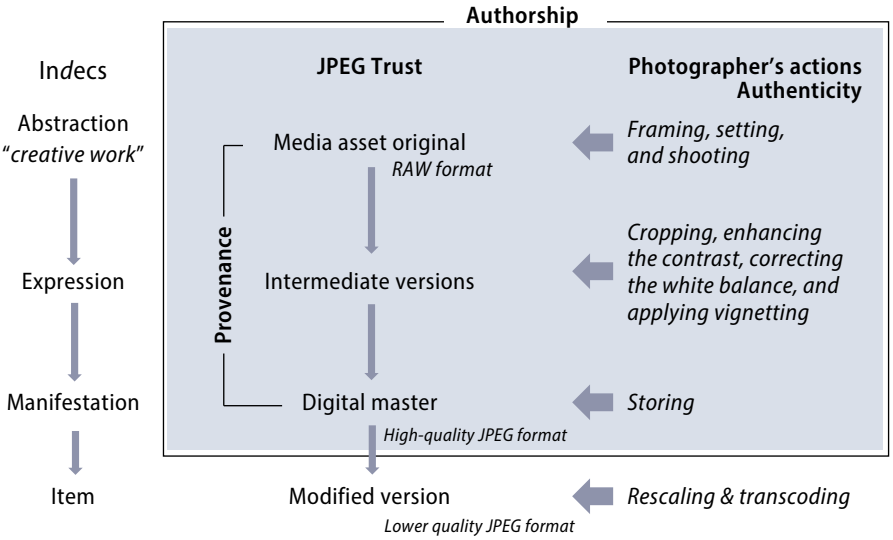
- [C2PA]: C2PA Technical Specifications
- [TRUST]: ISO/DIS 21617-1:2024(E) Core Foundation
- [DUBLIN]: Dublin Core™ Metadata Element Set
- [ODRL]: W3C® ODRL terminology
- [CITF]: Additional terms and definitions that this project has discussed and listed for future usage

Note: Copyright laws do not contain definitions of terms like "author" and "work". They are concepts that are starting to become harmonised by rulings of the European Court of Justice but may be used differently in different creative sectors. For the purpose of this report, the "author" is the first holder of the exclusive right that is provided for a creative "work" that is the original creation of the author.

Who is an "author" and what is a "work" is well defined in the music and book publishing sectors, but is only an emerging concept in photography where the "work" resulting from the creative actions of the "author(s)" – photographer and editor(s) – is the Digital master, considered as the creation intended by the "author(s)".

SDOs such as JPEG Trust define authorship in photography as below:

Figure 3. Defining authorship in photography – A working document from JPEG Trust



For the purpose of the copyright infrastructure, as it continues to develop over the years to come, it will be important to define under which conditions authorship can be determined and hence attributed in the various creative sectors.

4.1 Guidelines on the definition of an artificial intelligence system

The definition comprises seven main elements:

1. *a machine-based system,* [the term 'machine' can be understood to include both the hardware and software components that enable the AI system to function]
2. *that is designed to operate with varying levels of autonomy*
3. *and that may exhibit adaptiveness after deployment,*
4. *and that, for explicit or implicit objectives,* [to be distinguished from the "intended purposes"[12], which are the object of the sixth element]

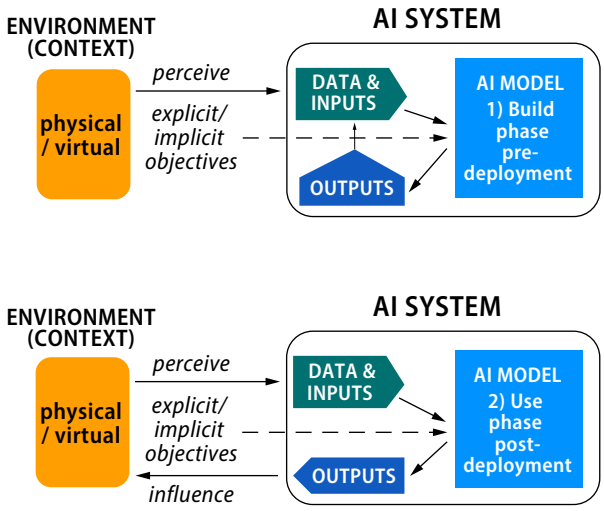
12 Art. 3 (12): 'intended purpose' means the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation.

5. **infers, from the input it receives, how to generate outputs** [through:
   a. machine learning approaches that learn from data how to achieve certain objectives, including supervised learning, unsupervised learning, self-supervised learning, reinforcement learning, and deep learning;
   b. logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved]
6. **such as**
   a. **predictions,**
   b. **content**, [The AI Act's risk-based approach means that only those systems giving rise to the most significant risks to fundamental rights3 and freedoms will be subject to its prohibitions laid down in Article 5 AI Act, its regulatory regime for high-risk AI systems covered by Article 6 AI Act and its transparency requirements for a limited number of pre-defined AI systems laid down in Article 50 AI Act, see also Art 7 (2) e) to i)]
   c. **recommendations,** or
   d. **decisions** [e.g., AI agents]
7. **that can influence physical or virtual environments.**

The definition of an AI system adopts a lifecycle-based perspective encompassing two main phases:

1. the pre-deployment or 'building' phase of the system and
2. the post-deployment or 'use' phase of the system.

Figure 4. Illustrative, simplified overview of an AI system © 2024 OECD



From Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market:

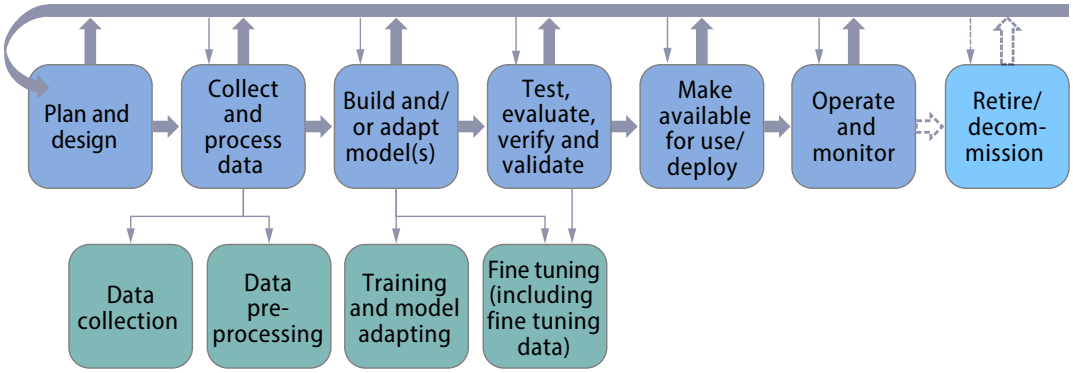
> *"Article 2 (2): 'text and data mining' means any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlation."*

From OECD – IP issues in AI trained on scraped data:

> *"'Data mining' generally refers to computational processes used to identify patterns and correlations in large datasets. There is some issue about whether and how the terms 'data mining' and 'data scraping" relate to legal definitions. The legal term 'text and data mining' (TDM) is commonly used by many countries in copyright laws to create limited exceptions to copyright protection. Unlike technical definitions, legal definitions may vary from jurisdiction to jurisdiction. Legal definitions may be open to interpretation by judicial or regulatory authorities."*

Text and data mining is defined in Directive (EU) 2019/790, Article 2, and in national law. Key copyright concepts such as "work" and "author" have been harmonised by decisions of the European Court of Justice. There may be differences in how restrictions on exclusive rights have been implemented nationally. However, Directive (EU) 2019/790, Article 4 is binding on Member States, meaning that it has been implemented in all Member States.

Figure 5. The AI model development cycle © 2024 OECD



Data scraping can involve an array of different activities, including: (1) data collection, (2) data pre-processing, (3) using the data for model training, model improvement, and/or (4) fine-tuning based on testing, evaluation, verification, and validation.

References

EU: Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act) / C(2025) 924 final

OECD:

OECD - Framework for the classification of AI systems - 2022

Explanatory memorandum on the updated OECD definition of an AI system - 2024

OECD - IP issues in AI trained on scrapped data - 2025

EUIPO: Development of Generative Artificial Intelligence from a Copyright Perspective - 2025

USCO: Copyright and Artificial Intelligence, Part 3: Generative AI Training, Pre-publication version - 2025

A technical introduction (3.5-hour video): Deep Dive into LLMs like ChatGPT

5 Use cases

The First Project of the CITF built use cases around a selection of five media objects.

The National Library of Finland (NLF) chose three: a **PhD thesis**, a **scientific article**, and an **eBook**. The National Library of Latvia (NLL) chose two: an **image** and an **article in a periodical**. These objects were selected as they represent different types of materials and undergo different creation and publishing processes. They represent materials stored in the National Libraries, but serve only as examples.

Table 1. Selected media objects and types of media

Library	Object	Type of media	Creation / publication year
NLF	Methods for Building Semantic Portals	PhD thesis	2013
NLF	Milloin väkivalta on ‘perheväkivaltaa’?	Scientific article	2024
NLF	Colonial aspects of Finnish-Namibian relations	eBook	2024
NLL	XX Vispārējie latviešu dziesmu un X deju svētki	Photograph	1990
NLL	Sadarbība - Turcijas bibliotēku attīstības virzītājspēks	Article in a periodical	2011

These five media objects are protected by three types of copyrights: **Creative Commons** (CC), expressed **terms and conditions** (T&Cs), and reserved as per the **Berne Convention** without further specification (© by author). The relationship between the copyright information and the National Library metadata collection was outlined above in section 2.2. Please also see Annex 2 – Use cases for further discussion and description of the infrastructures involved.

Table 2. Selected media objects and types of copyright

Library	Object	Type of copyright
NLF	Methods for Building Semantic Portals	© by author
NLF	Milloin väkivalta on 'perheväkivaltaa'?	T&Cs
NLF	Colonial aspects of Finnish-Namibian relations	Creative Commons
NLL	XX Vispārējie latviešu dziesmu un X deju svētki	T&Cs
NLL	Sadarbība - Turcijas bibliotēku attīstības virzītājspēks	T&Cs

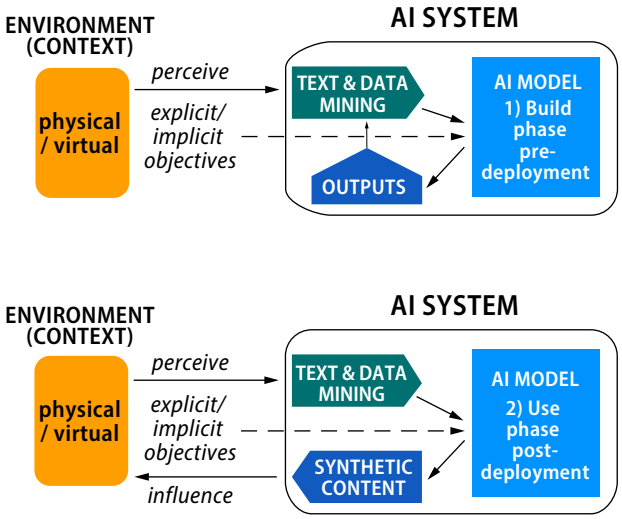
The 2030 use case scenarios were defined based on two main hypotheses:

1. **All content will be digitised and available online.** Accordingly, **Annex 2 – Use cases** describes the five media objects including their thumbnail, metadata, life cycle, and IT support as of 2025.
2. **AI applications will grow and multiply.** Therefore **Annex 2 – Use cases** presents the *Guidelines on the definition of an artificial intelligence system* published by the European Commission on the basis of OECD studies and explores – for the sake of this First Project – the potential relationship between *Artificial intelligence systems and copyright*.

As we write this report in June 2025, the multiple relationships between AI systems and copyright are being debated in numerous circles. The following paragraphs are meant to be policy-agnostic and business-model-neutral. They merely identify moments in the build and use phases of an AI system when the handling of rights management information may be necessary.

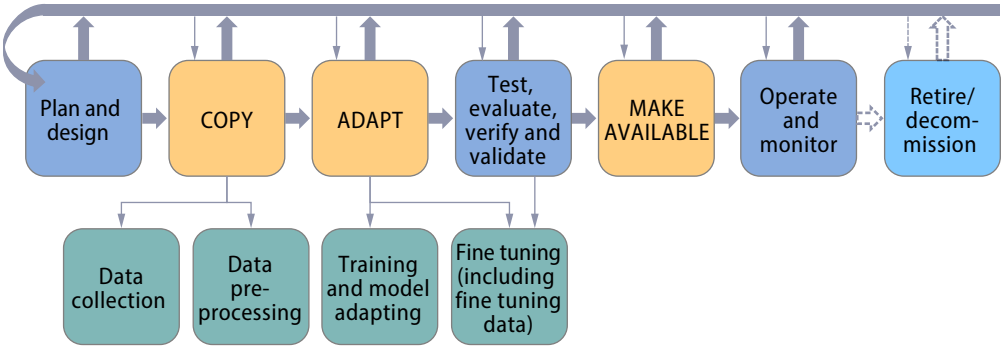
First, considering Figure 4 hereabove, we identified and localised "text and data mining" mentioned in Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market (CDSM) and "synthetic content" generated by AI as handled in the Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence:

Figure 6. Text and data mining and generative AI in the simplified overview of an AI system



Then, considering Figure 5 hereabove, we identified the rights apparently affected during the development of an AI system:

Figure 7. Rights affected during the development cycle of an AI system



It appears that:

- "text and data mining", as defined in Directive (EU) 2019/790, Art. 2, may correspond to element 5a of the definition of an artificial intelligence system presented in Section 4 – Terms and definitions;
- "synthetic content", aka content generated by AI as defined in the AI Act, may correspond to element 6b of the same definition, and
- intellectual property rights may be affected in those cases where the generative model reproduces a copyrighted work; see the abovementioned EUIPO study, pages 69-78, and the USCO report, pages 29-30.

Furthermore, it appears that:

- the pre-deployment or 'build' phase of an AI system may already affect the rights of **reproduction (copy)** (Berne Convention, Articles 1 to 6, and Infosoc Directive 2001/29/EC, Article 2), **attribution** (Berne Convention Article 6bis (1)), **adaptation** (Berne Convention Article 6bis (2)), **distribution** ("making available to the public", WIPO Copyright Treaty Article 6), and **communication to the public** (WIPO Copyright Treaty, Article 8, and Infosoc Directive 2001/29/EC, Article 3), whilst
- the post-deployment or 'use' phase of an AI system may affect the rights of **attribution, adaptation, distribution** and **communication** to the public.

Finally, this technical report understands that, for the time being, it has not been legally proven that AI development is or is not included in the TDM exception; see, for example, the article Artificial intelligence and copyright published on 2 June 2025 in IPRinfo.

One key objective of the CITF is to provide creators with more options to benefit from the AI exploitation of protected content. The debate on the use of copyrighted material to train generative AI models is evolving, shifting its focus from whether and when compensation is due to creators to determining the structure and specifics of a remuneration system. The 2030 use case scenarios also consider the need to respect authors' decisions not to allow use for training, even when remuneration would be available.

In his article Remuneration for use of works in text and data mining, Damian Flisak outlines three remuneration schemes:

- an opt-out mechanism is proposed as a way to license (or gain consent for) the use of protected works for AI training (see J.P. Quintais);
- a proposal, promoted by C. Geiger and V. Iaia, to remove the provision for commercial text and data mining (TDM) in favour of the introduction of a statutory licence to use a work to train AI models;
- an option, proposed by M. Senftleben, to make AI-generated content subject to a kind of royalty, paid to the community of creators.

In Enabling the Internet of Value, Philippe Rixhon shows how the tokenisation of media assets can bring new opportunities to the creative industries by increasing security, liquidity, and automation. Similarly, Thor Pettersen suggests reclassifying copyrights as formal financial assets. Both approaches offer a powerful, pragmatic, and forward-looking strategy for the AI era.

References

Berne Convention and WIPO Copyright Treaty and Infosoc Directive 2001/29/EC

6 Requirements

The CITF envisages the global semantic layer of a copyright infrastructure. It shall serve the cultural and creative industries in the European Union and further afield. In this sense, the semantic layer is jurisdiction-agnostic while the specific requirements emanating from EU regulations, e.g., the right to **opt-out from text and data mining**, are fully considered.

The global semantic layer is also business-model neutral. This Project considered the National Libraries mandates, commercial licensing, and future search across existing repositories of rights management information. Various business models will be supported by some or all requirements. In other words, not all requirements apply to all business models.

Metadata statements derived from a **technical** interpretation of acts and directives have no impact whatsoever on the policy or legal meanings of an article. They are only related to envisioned data flows and may be maximalised to trigger the development of robust standards addressing all possible theoretical cases.

6.1 The need for identifiers and metadata schemas

To address questions raised by the emergence of large language models and the deployment of Generative AI applications, the First Project of the CITF looked at:

- Media asset: what is what, and who can tell, and the need for **asset identifiers** (*identification metadata*);
- Parties: who is who, and who is accredited with work authorship or right ownership, and the need for **agent identifiers** (*identification metadata*);
- Attribution and terms & conditions: who did what, who owns what, what may be done with what, how authors or rightsholders can **opt-out of text and data mining**, and the need for **rights assertions** (*rights metadata*);
- Provenance and authenticity: where does that media asset come from, and the need for **provenance and authenticity assertions** (*descriptive metadata*).

The First Project distinguishes between two classes of requirements: RMI (**rights management information, 56 requirements**) and T&T (**traceability and transparency, 22 requirements**).

6.2 Different purposes, one CITF First Project

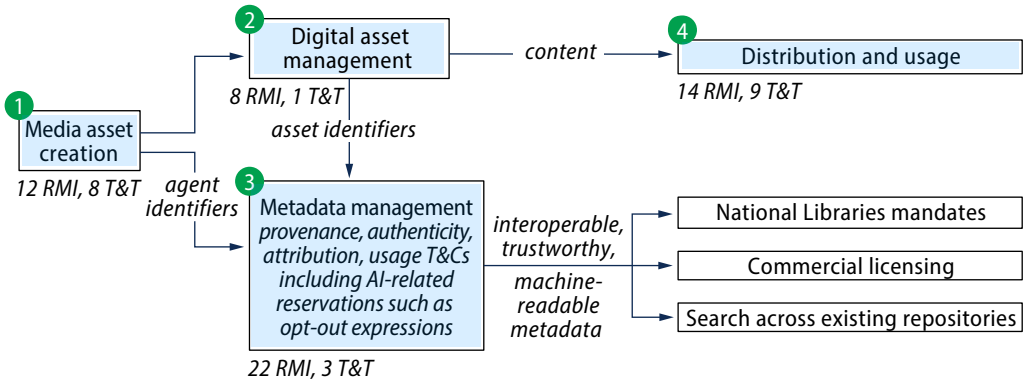
National Libraries generally have the mandate to secure the availability of cultural heritage published in or related to their countries as well as to transmit and produce information for research, studies, citizens, and society. They cooperate with libraries, archives, museums, and other operators as well as working with metadata and identifiers as mentioned previously.

Commercial licensing relies on real-time access to reliable, exhaustive, current, and interoperable rights data that supports rights management, licensing, enforcement, and remuneration. Libraries could support this level of activity as well.

Interconnection between and search across existing data repositories require **interoperable, trustworthy** and **machine-readable** metadata.

This report presents the metadata to be collected and processed in four phases (creation, asset management, metadata management, and distribution & usage). This amounts to more than 75 requirements for three main purposes (National Libraries mandates, commercial licensing, and search across existing repositories) considering four clusters of metadata (agent identifiers, asset identifiers, rights management information, and interconnection metadata).

Figure 8. Context of the requirements



Requirements are applicable for up to three purposes: NL for **National Libraries**, CL for **commercial licensing**, and IR for **interconnecting repositories**.

There is a gradual distinction, or hierarchy, between generic and specific requirements. For example, going from specific to generic, it can be said that an opt-out expression is an AI-related expression, an AI-expression is a text-and-data-mining-related expression, and that a text-and-data-mining expression is a terms-&-conditions expression.

Requirements pave the way for standardisation and use significant modal verbs: **shall** for mandatory statements, **should** for informative recommendations, and **may** for informative options.

The First Project also considered **benefits, challenges** and **technology considerations** related to some requirements.

Note: It is generally expected that technology will need to alleviate rightsholders' and rights users' metadata burdens. It will indeed be very beneficial to **simplify** rights and royalty management and have machines **generate** metadata and maintain it throughout the life cycle of the work. For example, one can imagine that a Bluetooth connection between a camera and a photographer's smartphone, which may include an EU Digital Identity wallet, could be sufficient to identify an image and photographer, attribute the image to the photographer, capture standard terms and conditions, and share the metadata on a data marketplace.

The requirements are listed and described in **Annex 3 – Requirements**.

Annex 1: Terms and Definitions

1 Introduction

To ensure correct understanding of the final report of the First Project of the CITF and its annexes, this document defines terms and concepts as they are used in this context.

Some definitions come from particular sources, e.g.,

- [C2PA]: C2PA Technical Specifications
- [C&NT]: Study on Copyright and New Technologies
- [DUBLIN]: Dublin Core™ Metadata Element Set
- [INDECS]: The <in*d*ecs> metadata framework
- [ODRL]: W3C® ODRL Terminology
- [OECD]: Framework for the classification of AI systems
- [TRUST]: ISO/DIS 21617-1:2024(E) Core Foundation
- [USCO]: Copyright and Artificial Intelligence, Part 3

Other definitions have been discussed during the project and listed here for future usage [CITF].

The **C2PA** (the Coalition for Content Provenance and Authenticity) is an open technical standard providing publishers, creators, and consumers the ability to trace the origin of different types of media [1].

The **Study on Copyright and New Technologies** – published by the European Commission in 2022 – analyses two issues linked to copyright and technology. The first part looks into the potential use of new technologies to improve the management of data linked to copyright-protected content by European creative industries. The second part concentrates on copyright-related challenges raised by the use of Artificial Intelligence [2].

The **Dublin Core™ Metadata Element Set** is a vocabulary of fifteen properties for use in resource description. The name "Dublin" is due to its origin at a 1995 invitational workshop in Dublin, Ohio; "core" because its elements are broad and generic, usable for describing a wide range of resources [3].

<in*d*ecs> (an acronym of "interoperability of data in e-commerce systems"; written in lower case) was a project co-funded by the European Community Info 2000 initiative and by several organisations representing the music, rights, text publishing, authors, library and other sectors in 1998-2000, which has since been used in a number of metadata activities. A final report and related documents were published; the document "Principles, model and data dictionary" is a concise summary [4].

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. The ODRL Information Model describes the underlying concepts, entities, and relationships that form the foundational basis for the semantics of the ODRL policies. Policies are used to represent permitted and prohibited actions over a certain asset, as well as the obligations required to be meet by stakeholders. In addition, policies may be limited by constraints (e.g., temporal or spatial constraints) and duties (e.g., payments) may be imposed on permissions [5].

The Global Partnership on Artificial Intelligence (GPAI) is an international initiative currently consisting of 44 member countries that promotes the responsible development and use of artificial intelligence (AI). Informed by a multistakeholder expert community that brings together governments, industry, academia, and civil society, it seeks to advance human-centric and trustworthy AI. In 2019, the **OECD** adopted the OECD AI Principles, the first intergovernmental standard on AI. These are the common foundations on which GPAI countries orient the initiative's work [6].

JPEG Trust provides a framework for establishing trust in media. The framework includes aspects of authenticity, provenance, intellectual property rights, and integrity through secure and reliable annotation of media assets throughout their life cycle [7].

Since launching an initiative in early 2023, the **US Copyright Office** has been examining the copyright law and policy issues raised by artificial intelligence (AI), including the scope of copyright in AI-generated works and the use of copyrighted materials in AI training. [8]

2 Terms and definitions

abstraction a creation which is a concept; an abstract creation whose existence and nature are inferred from one or more ***expressions*** or ***manifestations***. [INDECS]

Acquis Communautaire
the collection of common rights and obligations that constitute the body of EU law and is incorporated into the legal systems of EU Member States. [EUROPEAN COMMISSION]

action an operation on an ***asset***. [ODRL]

actor (actors) human or non-human (hardware or software) that is participating in the media ecosystem. Example: a camera (capture device), generation or editing software, cloud service or the person using such tools. [TRUST]

Note: see also ***agent, author, contributor*** [CITF]

agent a person or thing that takes an active role or produces a specified effect. [ODRL]

agreement a subclass of Policy that supports granting of Rules from assigner to assignee Parties. [ODRL]

AI-generated content (AIGC)
media asset created or modified by means of artificial intelligence (AI). [TRUST]

AI generative training
AI training aiming at content generation including generating new images, video, text, assessment and audio. This task combines forecasting and recognition tasks. However, the output often combines several existing elements such as images, text and audio to produce an object that was never seen before. This task tends to use structured learning. [OECD]

Notes:

1. Examples of content generation include machine translation, generative art, news stories – including fake news, spam emails and "deep fake" videos. [OECD]

2. Generative AI training involves the creation of perfect copies with the ability to analyse works nearly instantaneously. The result is a model that can create at superhuman speed and scale. In the words of Professor Robert Brauneis, "Generative model training transcends the human limitations that underlie the structure of the exclusive rights." [USCO]

AI inference "inference" is the process of using an AI model – trained from data or manually encoded – to derive a prediction, recommendation or other outcome based on new data that the model was not trained on. Different inference strategies can be used to derive varying results from the same model. These strategies are usually designed to optimise specific objectives and performance measures like robustness, accuracy, speed, business metrics or other criteria. [OECD]

AI training the model-building process, often called "training" or "optimisation", that is part of the system development in the lab. [OECD]

assertion data structure which represents a statement asserted by an ***actor*** concerning the ***media asset***. This data is a part of the ***trust manifest***. [C2PA]

asset a resource or a collection of resources that are the subject of a ***rule***. [ODRL]

Note: an asset is a file or stream of data containing digital content, asset metadata and optionally, a C2PA Manifest. [C2PA]

attribution the act of saying [...] that something is the result or work of a particular person or thing. [CAMBRIDGE DICTIONARY]

audit trail documents and records that show the history of [...] activities, examined by someone who is doing an audit. [CAMBRIDGE DICTIONARY]

authenticity property of digital content comprising a set of facts (such as the provenance data and hard bindings) that can be cryptographically verified as not having been tampered with. [C2PA]

author a human whose creativity led to a work being created or modified. [TRUST]

claim digitally signed and tamper-evident data structure that references one or more ***assertion*** by one or more ***actors***, concerning a ***media asset***, and the information necessary to represent the content binding. If any assertion was redacted, then a declaration to that effect is included. This data is a part of the ***trust manifest***. [C2PA]

claim signature digital signature on the ***claim*** using the private key of an ***actor***. The claim signature is a part of the ***trust manifest***. [C2PA]

constraint a boolean/logical expression that refines an ***action*** and ***party/asset*** collection or the conditions applicable to a ***rule***. [ODRL]

content credentials also known as a C2PA Manifest, these contain one or more assertions

which are statements about the asset, such as its origin (i.e., when and where it was created), modifications (i.e., what happened using what tools) and use of AI (i.e., how it was authored). [C2PA]

contributor an entity responsible for making contributions to the resource. [DUBLIN]

Notes:

1. ***Contributor*** is not a legal term in copyright law; it is used in the industry and standardisation developments to distinguish between main and assisting authors, a distinction which can – for example – impact royalty splits.

2. Examples of a ***contributor*** – in the industry and standardisation developments – include a person, an organisation, or a service. Typically, the name of a ***contributor*** should be used to indicate the entity. [DUBLIN]

copyright registration

copyright protection is obtained automatically upon creation of a work, without the need for registering or complying with any other formality. The fundamental principle of formality-free protection is reflected in Art 5(2) of the Berne Convention and in many other international treaties in the field of copyright and related rights. While respecting the principle of formality-free protection, voluntary registration systems are an integral part of many national copyright systems. Currently national registration and recordation systems are very diverse in terms of functioning and legal effects. [WIPO]

creative work a creation of the mind, such as literary and artistic works. [WIPO]

Note: Exhaustive lists of works covered by copyright are usually not to be found in legislation. Nonetheless, broadly speaking, works commonly protected by copyright throughout the world include:
- literary works such as novels, poems, plays, reference works, newspaper articles;
- computer programs, databases;
- films, musical compositions, and choreography;
- artistic works such as paintings, drawings, photographs, and sculpture;
- architecture; and
- advertisements, maps, and technical drawings.

Copyright protection extends only to expressions, and not to ideas, procedures, methods of operation or mathematical concepts as such. [WIPO]

creator entity primarily responsible for making the resource. Note 1 to entry: Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity. [DUBLIN]

Note: see also ***agent***, ***author***, ***contributor***. [CITF]

CRUD (create, read, update, delete)
this computer science acronym refers to the four fundamental operations that can be performed on data, especially in databases and persistent storage systems. [CITF]

data mining the process of using special software to look at large amounts of computer data in order to find out useful information. [CAMBRIDGE DICTIONARY]

date a point or period of time associated with an event in the lifecycle of the resource. Note 1 to entry: Date may be used to express temporal information at any level of granularity. Recommended best practice is to use an encoding scheme, such as the W3CDTF profile of ISO 8601 [W3CDTF]. [DUBLIN]

description an account of the resource. [DUBLIN]

Note: description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource. [DUBLIN]

digital content
the portion of an asset that represents the actual content, such as the pixels of an image, along with any additional technical metadata required to understand the content (e.g., a colour profile or encoding parameters). [C2PA]

digital signature
a mathematical scheme for verifying the authenticity of digital messages or documents. [CITF]

distributed ledger
a database that spans many physical locations. [TRUST]

distributed ledger technology (DLT)
a distributed system that enables a tamper-evident, append-only data store (ledger). [TRUST]

duty the obligation to exercise an agreed action. [ODRL]

embedded existing or firmly attached within something or under a surface. [CAMBRIDGE DICTIONARY]

expression an event which is a creation. [INDECS]

EU Digital Identity Wallet
: provides secure and versatile storage for digital credentials. It aims to simplify digital interactions across borders while ensuring interoperability and user control.

file
: information stored on a computer as one unit with one name. [CAMBRIDGE DICTIONARY]

fixation
: a requirement of copyright law for a 'work' to be recorded in some permanent form to give rise to protection, rather than simply existing as an idea or concept. [LEXIS NEXIS]

format
: the file format, physical medium, or dimensions of the resource.

Notes:

1. Examples of dimensions include size and duration. Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types [MIME]. [DUBLIN]

2. Format is an artefact on which an expression may be infixed to create a **manifestation**. [INDECS]

General-purpose AI
: an AI system which is based on a general-purpose AI model and which has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems. [AI Act EU/2024/1689]

generative AI media asset
: a media asset created by means of artificial intelligence (AI) and machine learning (ML). [TRUST]

GLAM institutions
: Galleries, libraries, archives and museums. [TRUST]

global ID
: a single ID for a natural person or a legal entity in the form of an alphanumeric token or Digital ID to be accepted within the global IP community. [WIPO]

hard binding
: one or more cryptographic hashes that uniquely identifies either the entire asset or a portion thereof. [C2PA]

identifier
: an unambiguous reference to the resource within a given context.

Notes:

1. Recommended best practice is to identify the resource by means of a string conforming to a formal identification system. [DUBLIN]

2. An identifier is a value that uniquely refers to something, such as an asset or actor. [TRUST]

intellectual property rights (IPR)
: exclusive right of the actor to the intellectual work of their creation. [TRUST]

interoperability
: the ability to work together with other systems or pieces of equipment. [CAMBRIDGE DICTIONARY]

item
: a single instance of an artefact. [INDECS]

language
: a language of the resource. Note 1 to entry: Recommended best practice is to use a controlled vocabulary such as RFC 4646 [RFC4646]. [DUBLIN]

ledger
: a ledger is a book of accounts that contains the records of transactions. [TRUST]

legal deposit
: a system whereby publishers are legally required to deposit copies of their published works with designated repositories, like National Libraries. This ensures the preservation and accessibility of a country's cultural heritage. [CITF]

machine-readable
: able to be understood and used by a computer. [CAMBRIDGE DICTIONARY]

manifestation
: an artefact containing an infixion of an expression. [INDECS]

media assets
: digital assets including images, videos, audio or text. [TRUST]

media asset metadata
: portion of a media asset that represents non-technical information about the media asset or its content, such as location, creator, annotations or IPR information. [TRUST]

media type
: a standard description of the type and/or format of the data. [IETF RFC 2046]

metadata
: information that is given to describe or help you use other information. [CAMBRIDGE DICTIONARY]

  Note the different types:

  1. Identification metadata. These metadata make up the smallest set of data that uniquely defines an identified entity – works, related subject matters, and stakeholders. Although different sets of data can be identified, there is usually community consensus on the appropriate set to use in particular circumstances.

  2. Rights metadata. This metadata or attribution metadata are the information relating to the authorship and ownership, information about the terms and conditions of use and reuse of a protected

subject matter. It links the answers to who did what, who owns what, and what can we do with that.

3. Descriptive metadata. These metadata or rich metadata or discovery metadata can include information that is helpful in searching for a particular entity (the genre of a novel or musical composition), or aid its enjoyment or understanding (the influences on the creator or group of creators) or for general interest (the location of the first performance of a concerto).

4. Usage metadata. This is a category of metadata that is becoming more and more important because of the emergence of new business models, for example advertisement- or subscription-based digital platforms, mashups and prosumerism. This category comprises information on where, when, and how frequently content has been used, and how it has been monetized: it is essential to monitor usage of copyright-protected content and remuneration for its use.

5. Administrative metadata. The metadata in all the previous categories have provenance – who asserted the information and when. This fifth category of metadata, often called administrative metadata, is needed to assess the reliability, value, and trustworthiness of the other metadata. [C&NT]

micro-licensing
a legal phrase to describe the difference between a standard licence and a limited licence in terms of time, purpose, channels, territories, etc. [TRUST]

obligation the obligation to exercise an agreed action. [TRUST]

opt-in to choose to be part of an activity, arrangement, etc. [CAMBRIDGE DICTIONARY]

opt-out to choose not to be part of an activity or to stop being involved in it. [CAMBRIDGE DICTIONARY]

orphan works works that are still protected by copyright, but whose authors or other right holders are not known or cannot be located. Music, books, newspaper and magazine articles and films can be orphan. [EUIPO]

out-of-commerce works
works that are still protected by copyright but are not available commercially, such as literary works, audio-visual works, phonograms, photographs and unique works of art. [EUIPO]

party an entity or a collection of entities that undertake Roles in a ***rule***. [ODRL]

permission the ability to exercise an action over an ***asset***. [ODRL]

persistent record
data that is stored in such a way that it remains accessible and unchanged even after the process that created it has ended. [CITF]

persistent identifier (PID)
a long-lasting, unique code that reliably identifies and locates a digital resource, even when its online location changes. [CITF]

playback interface
a device that enables a computer to send [audio] signals to various output devices for playback. It is essentially a digital-to-analogue converter, translating digital [audio] data from a computer into analogue signals that can be amplified and played. [CITF]

playback session
the act of playing back previously recorded data, such as audio, video, or other digital media. It allows you to revisit or analyse the recorded content. [CITF]

policy a group of one or more rules, i.e., a non-empty group of Permissions (via the permission property) and/or Prohibitions (via the prohibition property) and/or Duties (via the obligation property). [ODRL]

preliminary bibliographic record
also known as a pre-bibliographic record, contains standard MARC fields (machine-readable cataloguing format) with appropriate content designation and sufficient information to permit the determination of a bibliographic match manually or by machine processing. [YALE]

prohibition the inability to exercise an action over an ***asset***. [ODRL]

provenance the logical concept of understanding the history of an asset and its interaction with actors and other assets, as represented by the provenance data. [C2PA]

publisher an entity responsible for making the resource available. Note 1 to entry: Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity. [DUBLIN]

qualified timestamp
a digital process that ensures the integrity of electronic documents by guaranteeing their legal validity, security and traceability. It involves the use of a trusted third party who employs advanced

cryptographic techniques to generate a unique digital fingerprint for an electronic document, which is then encrypted and securely stored in a tamper-proof container known as a "timestamp token." [CITF]

Note: by using eIDAS qualified timestamping, legal and technology professionals can ensure the authenticity and integrity of electronic documents while complying with EU regulations.

qualified trust service provider

a provider who operates within a framework of established regulations, such as the European Union's eIDAS regulation. This guarantees adherence to stringent legal and security standards. It raises confidence in electronic transactions for users, clients and stakeholders alike, fostering a secure and reliable environment for conducting business and interacting digitally. [CITF]

recordation of transfers

in the context of copyright, refers to the process of filing documents related to copyright ownership. This includes transfers of copyright ownership, assignments of rights, copyright licenses and other documents pertaining to a copyright. [CITF]

registrar the actor that performs a registration. [TRUST]

registration process of storing information (e.g. media asset, metadata or provenance) about a media asset, separate from the media asset itself. [TRUST]

relation a related resource. Note 1 to entry: Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system. [DUBLIN]

rights information about rights held in and over the resource. Note 1 to entry: Typically, rights information includes a statement about various property rights associated with the resource, including intellectual property rights. [DUBLIN]

rights management information

any information provided by rightsholders which identifies the work or other subject matter […], the author or any other rightsholder, or information about the terms and conditions of use of the work or other subject-matter, and any numbers or codes that represent such information. [This] shall apply when any of these items of information is associated with a copy of, or appears in connection with the communication to the public of, a work or other subject matter. [Infosoc directive 2001/29/EC]

Note: rights management information includes identification metadata and rights metadata, see ***metadata***. [C&NT]

rule an abstract concept that represents the common characteristics of ***permissions, prohibitions*** and ***duties***. [ODRL]

signer actor who digitally signs a ***media asset***. [TRUST]

signing process that establishes the relation between an ***actor*** and a ***media asset*** in a tamper-evident manner. [TRUST]

soft binding a content identifier that is either (a) not statistically unique, such as a fingerprint, or (b) embedded as an invisible watermark in the identified digital content. [C2PA]

source a related resource from which the described resource is derived. Note 1 to entry: The described resource may be derived from the related resource in whole or in part. Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system. [DUBLIN]

statement something that someone says or writes officially, or an action done to express an opinion. [ADD]

statutory licence
a licence or certificate of registration issued under statute, statutory instrument or by a Government or Local Authority provided that this licence or certificate is necessary to conduct core business activities. [CITF]

subject the topic of the resource. Note 1 to entry: Typically, the subject will be represented using keywords, key phrases or classification codes. Recommended best practice is to use a controlled vocabulary. [DUBLIN]

synthetic media asset
media asset generated at least partially by a computer program. [TRUST]

talent ID a global identifier standard providing cultural industries the ability to resolve, verify and automate talent identity. [HAND]

text and data mining
any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations. [CDSM Directive EU/2019/790]

three-step test
based on Article 5.5. of the Infosoc directive EC/2001/29, Hugenholtz and Okediji formulate the three steps as follows:

	- Limitations and exceptions cannot be "overly broad" [from "certain special cases"], - Limitations and exceptions cannot "rob right holders of a real or potential source of income that is substantive" [from "do not conflict with normal exploitation of the work"], - Limitations and exceptions cannot "do disproportional harm to the rights holders" [from "prejudice the legitimate interests"].
title	a name given to the resource. Note 1 to entry: Typically, a Title will be a name by which the resource is formally known. [DUBLIN]
timestamp	a record in printed or digital form that shows the time at which something happened or was done. [CAMBRIDGE DICTIONARY] Note: timestamping provides a secure method to ensure that electronic documents are not tampered with or altered after they have been signed.
tokenization	the process of encapsulating an asset's rights into a digital token. [TRUST]
trust credentials	the set of ***trust indicators*** that are derived from a ***media asset*** and its ***trust record***. [TRUST]
trust declaration	specific type of ***trust manifest*** that, when present, is always first in the trust record. It represents the ***actor*** that created the ***media asset*** and contains only mandatory assertions. [TRUST]
trust indicator	information derived from a combination of the ***media asset*** and the ***trust record*** that indicate the level of trustworthiness of a media asset in a given context. [TRUST]
trust manifest	set of information about the ***provenance*** of a ***media asset***. A trust manifest is part of a ***trust record***. [TRUST]
trust profile	set of expressions that are used to evaluate each ***trust indicator*** in a given ***trust credential*** to indicate the level of trustworthiness of a given ***media asset***. [TRUST]
trust record	collection of one or more ***trust manifests*** that can either be embedded into a ***media asset*** or be external to it. [TRUST]
trust report	result of evaluating a ***trust credential*** against a trust ***profile***. [TRUST]
trustworthy	able to be relied on as being what it is asserted to be. [TRUST]
type	the nature or genre of the resource. Note 1 to entry: Recommended best practice is to use a controlled vocabulary such as the DCMI Type Vocabulary [DCMITYPE]. To describe the file format, physical

medium, or dimensions of the resource, use the Format element. [DUBLIN]

URI (abbreviated term)
uniform resource identifier.

user an actor who uses a computer or network service. A user often has a user account and is identified to the system by a username. [CITF]

verifiable able to be checked. [TRUST]

wizard a step-by-step guide or tool that helps users perform complex tasks or configurations without requiring extensive technical knowledge. [CITF]

work see creative work.

3 References

[1] C2PA Technical Specifications v2.2: https://c2pa.org/specifications/specifications/2.2/index.html
[2] Study on Copyright and New Technologies: https://op.europa.eu/en/publication-detail/-/publication/cc293085-a4da-11ec-83e1-01aa75ed71a1/language-en
[3] Dublin CoreTM Metadata Element Set v1.1 Reference Description: https://www.dublincore.org/specifications/dublin-core/dces/
[4] The <in*d*ecs> metadata framework – Principles, model and data dictionary: https://www.doi.org/resources/indecs_framework_2000.pdf
[5] W3C® Open Digital Rights Language Information Model v2.2: https://www.w3.org/TR/odrl-model/#terminology
[6] OECD Framework for the Classification of AI systems: https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html
[7] JPEG Trust Terms and Definitions v2.0: https://ds.jpeg.org/documents/jpegtrust/wg1n100972-105-REQ-Terms_and_definitions_for_JPEG_Trust_v2_0.pdf
[8] USCO, Copyright and Artificial Intelligence, Part 3: Generative AI Training – 2025: https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf

Annex 2: Use cases, aka scenarios 2030

1 Introduction

This Annex details the use cases, aka scenarios 2030 described for the First Project of the Copyright Infrastructure Task Force (CITF) on Interoperable, trustworthy and machine-readable copyright data. Written from the perspective of the National Library infrastructures, the use cases analyse the life cycle of five different objects deposited with the National Libraries of Finland and Latvia. The life cycle refers to the different steps from creation to deposition with the National Library and further uses.

These objects predate contemporary artificial intelligence technology, especially the widespread use of generative AI, as well as the opt-out reservations that have recently been introduced in the European Union. Therefore, the impact of these current developments has been reflected on the current life cycles of the objects. For example, regarding the expression of opt-out reservations, we discuss whether and how authors could and would need to express these reservations in relation to current publication and metadata management processes.

The National Library infrastructures and data flows in Finland and Latvia are presented at the beginning of this Annex to familiarise the reader with their context.

The 2030 use cases are defined based on two main hypotheses:

Hypothesis 1: all content will be digitised and available online.

Descriptions of five objects with their thumbnail, metadata, life cycles, and IT support as per 2025:

- First object – a doctoral thesis
- Second object – a scientific article
- Third object – an eBook
- Fourth object – an image
- Fifth object – an article

Hypothesis 2: AI applications will grow and multiply.

Guidelines on the definition of an artificial intelligence system.

1.1 Systems and data flows of the National Library of Finland (2025)

This section of the use cases appendix describes the data flows of the National Library of Finland and the National Library of Latvia according to data type: where materials are processed, where metadata are collected and used, and where both are involved. In addition, the systems are described from the perspective of integration: which systems can be connected to new components, which systems are maintained by the National Library, and which were acquired from elsewhere. Concerning the National Library of Finland, the text is a modified and translated version of a report published in 2021: Lehtinen, M., Niininen, S., Inkinen, J., & Lappalainen, M. (2021). *Automaattisen kuvailun palvelun integroiminen Kansalliskirjaston järjestelmäkokonaisuuteen – tietovirrat ja prosessit.* https://urn.fi/URN:ISBN:978-951-51-6986-0

The terms and descriptions may not align with their interpretations in other contexts, like the copyright environment. Nonetheless, they are defined and explained for the purpose of the use cases. In the context of the terms and definitions the notes include the relevance of the term for each use case.

1.1.1 Sources

- **Electronic legal deposits**
  - Online form
    - A form to submit electronic legal deposit copies and related metadata.
    - Both metadata and full text are available.

  - Mass donations
    - The digital publications of major publishers can be obtained as mass donations from the publication intermediaries directly to the reception server. The disclosures include both the publication and the related ONIX product information.
    - Both metadata and full text are available.

- **Preliminary information**
  - Book brokerage / Storia
    - The author provides preliminary information on new publications and updated product information in ONIX format directly to the receiving server.
    - Only metadata available.
  - ISBN Centre
    - The ISBN Centre produces MARC datasets in Melinda in conjunction with the issuance of ISBN identifiers.
    - Only metadata available.

- **Digitisation**
  - Digitisation services produce productions based on the annual digitisation plan of the National Library of Finland.
  - Both the full text and metadata are available.

- **Publication archives**
  - Publication archives have input forms that users can use to import and describe material. Automatic import from other systems is also possible.
  - Both the full text and metadata are available in the publication archives.

- **Licensed reference works and databases**
  - E-resources subject to a fee in the Faculty of Arts discipline collection.
  - Only metadata available.

- **Licensed books, journals, and their packages**
  - E-resources subject to a fee in the Faculty of Arts discipline collection.
  - Only metadata available.

1.1.2 Background systems

- **Varsta** – An archive for the release of legal deposit copies.
- **Melinda** – A shared description environment and metadata pool for libraries. Melinda uses the library system Aleph developed by Ex Libris and Clarivate.
- **Fikka** – the National Library of Finland collection database, works in the open-source KOHA library system.
- **Melinda batch import system** – Imports refer to the import of records into the Melinda database through interfaces in connection with the Melinda service. With the help of the batch import system, records are

imported from, for example, commercial operators, open publication archives of various organisations, and the Helmet database of public libraries in the Helsinki Metropolitan Area. Records imported by batch import are often converted from another format to the MARC 21 format.
- **Digi** – The publication system for the National Library of Finland's digitisation.

1.1.3 End-user systems

- **National Library of Finland search service (Finna)** – A Finnish joint search service and platform for accessing library, archive, and museum materials. Maintained and developed by the National Library of Finland, materials are provided by a consortium of Finnish libraries, archives, and museums. Finna.fi is the national search service and UI for all these materials. Re-use services like programmable interfaces (APIs) for the same data are available at Finna.fi, as well as platform services for member organisations to build their own search and usage UIs focused on their materials on top of common Finna index and UI templates. The Finna platform is based on open-source software (VuFind, Solr) and harvests metadata through the standard OAI-PMH interfaces of the individual collection systems of the material provider organisations. It supports the most typical standard metadata formats used in the library and cultural heritage fields (MARC21, DC, LIDO, EAD, EAD3, Forward). Finna is a discovery platform and search service for the materials, but for those materials containing a direct link to the digital object it is also a UI. If no direct link is available to the digital object, the user is provided with a link to the system where it is available.
- **Varia** – DSpace-based publication archive for electronic legal deposit copies. Available only at legal deposit workstations.

1.1.4 Electronic legal deposit copies and their processing

The task of the National Library of Finland is to make Finnish publications available to researchers and other people who need them. Its operations are regulated by the Act on the Storage and Preservation of Cultural Materials 1433/2007 (Cultural Materials Act). In the case of online publications, the National Library of Finland strives to make a representative and diverse sample of material available to the public at various times through its data networks.

The digital resources are described, as appropriate, in the Fennica National Bibliography and Viola National Discography. The stored resources are available locally at the National Library of Finland and other legal deposit libraries as well as at the legal deposit workstations at the National Audiovisual Institute and the Library of Finland.

The electronic legal deposit copies of major publishers are obtained as bulk donations, whereby the National Library is provided with both the publication itself and the related ONIX product information. The ONIX information is converted into MARC 21 by the Melinda batch import system and exported to Melinda. The aim is that, if a previous description of the publication is already available in Melinda, the converted data will be combined with it in accordance with the Melinda merger rules. Finally, the purpose is to introduce Fennica's content description of the corresponding printed version of the record, so that the content descriptions of the electronic and printed versions are consistent in the local database and its search service. The use of the content description of the printed version is also justified by the fact that the Library Service rapidly provides Fennica with comprehensive metadata on the printed material of large publishers, often even before the electronic version is released. The combination of records and the identification of the corresponding printed version have not yet been implemented, but after their completion, the processing of e-library copies by major publishers can be fully automated.

Legal deposit copies of small publisher and self-published works can be obtained as individual donations through the donation form. The publications are manually described in Melinda using both the metadata provided by the donor and the metadata already available in Melinda. If the preliminary information is already in Melinda, the actual description is incorporated on top of it. The donation form also provides some material not included in the description, such as ephemera and series, whose individual parts are not described as monographs. In the future, automation tools will hopefully also be utilised to make these datasets available, if the bibliographical description of the datasets can be automated.

Released publications are stored in the publication archive Varia, and the publication description information is always added to the URN that leads to Varia. Publications can be searched through the National Library's search service, but they are only available at legal deposit workstations.

1.1.5 Preliminary information process

Preliminary information on the publications of large commercial publishers can be obtained from the book logistics / brokerage company Storia before the publication of the work. At the moment, pre-information records are obtained in MARC 21 format, but in the future, they will be submitted by the Authorising Officer as an ONIX message, which will be converted by the Melinda batch import system and then exported to Melinda and Fennica. Annif is used by book brokers in their own systems, which means that the preliminary information also includes concepts from the General Finnish Ontology. PLC (Public Libraries Classification System) classifications and other keywords can also be included.

In addition to the preliminary information, Storia also submits revised product information, which they add to the preliminary information after the publication of the work. This is processed by the Melinda batch import system along with the preliminary data.

The ISBN Centre manually produces preliminary information on publications by small publishers when it issues an ISBN identifier.

The preliminary information does not cover all Finnish publications. In particular, medium-sized publishers are left behind. However, the National Library is in the process of redesigning its identifier system, and in the future, the retrieval and distribution of ISBN, ISSN, and ISMN identifiers, as well as the registration of publishers and the maintenance of publisher data, will be carried out through the new identifier register. If implemented, the identifier register would improve the quality and coverage of the preliminary information obtained from publishers, including medium-sized publishers.

1.1.6 Publication archives

The National Library's repository services provide their customer organisations with an information system that enables them to manage and publish digital documents from theses and other texts to images and videos. Customer organisations include universities, higher education institutions, institutions, agencies, and associations. The client describes the datasets and loads them into the system using the input form. The data can also be imported automatically from other systems. Some of the metadata in the publication archives will be converted by the Melinda batch import system using the Dublin Core MARC 21 format and exported to Melinda. The entire text material related to the converted records will also be exported to Varia.

The National Library of Finland maintains its own repository Doria (www.doria.fi), as well as other repositories that are used by several organisations, including:

- **Jukuri**, the Natural Resources Institute Finland (Luke) open repository: http://jukuri.luke.fi
- **Julkari**, the publication archive for the administrative branches of the Ministry of Social Affairs and Health: http://www.julkari.fi
- **Lauda**, the University of Lapland electronic publication archive: http://lauda.ulapland.fi
- **LutPub:** http://lutpub.lut.fi
- **Osuva**, the University of Vaasa's repository: http://osuva.uwasa.fi
- **Trepo**, the University of Tampere's repository: https://trepo.tuni.fi/
- **Theseus**, the University of Applied Sciences repository: http://www.theseus.fi
- **UTUPub**, the University of Turku's publication archive: https://www.utupub.fi/
- **Valto**, the Finnish Government's joint publication archive: http://publications.valtioneuvo.fi

1.1.7 Digitisation process

In the process of digitising catalogued monographs, an existing printed record is retrieved from the Finna API and later also from the Fikka API. A digital publication record is created on the basis of the record.

In the process of post-processing non-catalogued monographs, minimum cataloguing information is produced (name, author, publisher, time, place, material, if any) and a digital record is created. In the case of small-scale printing, digitisations also include a parent aid made by a collection, which describes the category of the collection (e.g., cycling) the work belongs to, and this dataset has a link to the records of individual ephemera.

MARC records will be extracted from Digi's OAI-PMH interface outwards, from which they will be made available through the Melinda batch import system to Fennica. In such cases, the link to the digitised publication in Digi will be transferred to both the digitised publication and the original printed records. In addition, we are working on re-harvesting the records from Finna/Melinda to make the changes available in Digi as well.

In the press material process, the data are listed in ISSN/NBN, with the number and date of publication of the journal, which are recorded in the MODS format.

1.1.8 Licensed e-resources

Only metadata are processed in the description processes of licensed e-resources. The full texts are located on the publishers' servers. Terms of use and rights have been defined for each dataset, which very rarely allow for mass loading by a machine. The new EU Copyright Directive (DSM Directive, (EU) 2019/790) aims, among other things, to allow data mining for the benefit of non-commercial scientific research. The implementation of the directive should be monitored to see whether it would also enable development of the automation of electronic datasets.

1.1.9 Identifier services

The National Library of Finland has national responsibility for ISBN, ISSN, ISMN, and URN publication identifiers as well as for ISIL organisational identifiers used by libraries and other institutions. The National Library distributes these identifiers, maintains the identifier registry, and serves as the national authority responsible for these systems.

The National Library also uses ISNIs to identify agents related to publications produced in Finland. The National Library is a member of the international ISNI consortium as well as an ISNI registration agency, meaning that it can submit requests for new ISNI assignments on behalf of other organisations. The National Library collaborates extensively with the rest of the GLAM sector as well as with collective management organisations related to identifiers.

1.2 Systems and data flow at the National Library of Latvia

This section of the use cases describes how the key information systems of the Digital Culture Heritage Platform of Latvia – ATS, DOM, *APLIS*, and the Digital Library of Latvia – participate in the processing, management, and use of data. These systems are managed by the National Library of Latvia and the Culture Information Systems Centre. It distinguishes between workflows involving full-text content, metadata, or both, and illustrates how these systems support different aspects of the data lifecycle. In addition, this section addresses system integration and interoperability.

This section also outlines the end-user systems of the National Library of Latvia that provide access to various types of content and services. The description highlights how these systems interact with backend services, retrieve metadata and content, and support different user groups, including researchers, librarians, and the public.

1.2.1 Background systems

The Digital Object Management System (DOM) serves as the central platform for the management and preservation of digital cultural heritage objects at the National Library of Latvia and 300 institutions who are culture heritage holders (museums, archives, libraries, other). It ensures unified handling, metadata standardisation, and long-term storage of digitised content in the country.

The Reference Data System (ATS) is a national authority data and knowledge organisation system developed for Latvia's cultural heritage institutions. Its primary objective is to collect, normalise, and connect reference data that previously existed in fragmented institutional databases, legacy systems, and documents.

Key features of ATS include:

- Centralised knowledge organisation: a unified platform for creating, enriching, and managing authority data for cultural heritage entities.
- Data object types:
  - Entities such as people, institutions, places, and events;
  - Controlled vocabularies;
  - Classification systems.
- Cross-institutional interoperability: ATS enables structured collaboration and metadata harmonisation across multiple cultural memory institutions in Latvia.
- Integration with DOM: ATS is linked to DOM, allowing staff to access and apply standardised reference data directly during metadata creation workflows.

The Copyright Management and Licensing System (*APLIS*) is responsible for the accurate identification, administration, and licensing of copyright and related rights associated with culture heritage objects and digital cultural heritage objects. It ensures that objects are handled in compliance with the Latvian Copyright Law and licensing agreements.

APLIS provides:

- Functionality for the determination and management of copyright and related rights;
- Tools for assigning usage rights to digital objects, either based on legal exceptions or through licensing;
- Support for accounting and rights holder remuneration, where applicable.

APLIS is also embedded within the DOM environment, enabling staff to define access and licensing conditions directly during the object description process. *APLIS* also uses data from ATS – particularly death dates of authors – to determine whether copyright is still active (e.g., applying the 70-year post-mortem rule). The rights metadata produced in *APLIS* is then displayed to end users through the Digital Library of Latvia, providing transparent information on how digital objects may be used, shared, or reused.

Aleph, the library information system used by the National Library of Latvia, serves as the primary catalogue for bibliographic and authority records. Authority records created in Aleph using the MARC21 format are automatically transferred to ATS the following day.

The National Library of Latvia's Portal for Publishers (TPS) provides a centralised application for the assignment of identifiers, including ISBN, ISMN, ISSN, UDC, and Cutter numbers. Publishers complete web-based application forms with minimal required metadata to receive identifiers. TPS also incorporates invoice creation functionality for chargeable services (ISBN and ISSN barcode generation). Payment system integration is planned for future development.

Depending on the requested identifier type, additional mandatory data are required:

- **UDC applications** – author quantity, publication topics, and annotations;
- **ISBN/ISMN applications** – provisional publication quantity projections;
- **ISSN applications** – periodicity, title page imagery, and URL addresses for e-publications.

Upon identifier assignment, TPS automatically generates short bibliographic records using application data. TPS incorporates a built-in messaging system that enables NLL staff to monitor and encourage legal deposit compliance, with automatic email notifications. TPS also serves as the primary channel for e-publication submissions except for periodicals and audiovisual publications, which are acquired through FTP servers and cloud services. Following legal deposit delivery and staff verification (de visu inspection of the publication against the record), TPS exports metadata to the ALEPH Catalogue and metadata along with e-publication files to DOM.

1.2.2 End-user systems

Digital Library of Latvia – a national discovery platform and search service for digitised Latvian cultural heritage, maintained and developed by the National Library of Latvia. It aggregates content from over 500 libraries, archives, museums, cultural organisations, and individuals, offering access to millions of items including books, periodicals, photographs, audiovisual materials, and intangible heritage. The Digital Library is accessible via digitalabiblioteka.lv, which serves as the central search and usage interface. Metadata is harvested through OAI-PMH interfaces and mapped from various formats (e.g., MARC21, DC, LIDO, EAD) to EDM. The discovery platform allows users to search, browse, and access digital objects directly and also links them to the external systems the content comes from. Full-text search is powered by Elasticsearch, with advanced capabilities including article-level search in periodicals, synonym handling, and linked data entity referencing. The service includes features such as user workspaces, digital collection creation, digitisation requests, and access to high-resolution files. The interface is available in Latvian and English and includes automated metadata translation. Data security is ensured by maintaining two independent copies using OpenStack Swift object storage within infrastructure managed by the National Library of Latvia. A catalogue of Latvian digital heritage resources is also integrated into the service.

The National Library of Latvia hosts a wide range of **digital collections**, including the Books Portal, which features around 15,000 digitised titles from the 16th to 21st centuries, available for full-text viewing and metadata-based search. The Periodicals Portal offers newspapers and magazines dating back to 1612. Other highlights include historical maps with geospatial tools, rare manuscripts, traditional music recordings, photo archives like In Search of Lost Latvia, and curated thematic cross-institutional collections like Latvian Song and Dance Celebrations.

1.2.3 Identifier services

The National Library of Latvia holds national responsibility for the assignment and administration of ISBN, ISSN, and ISMN identifiers in Latvia. It is the designated national agency for these publication and organisational identifiers, ensuring their distribution, maintaining the respective national registries, and serving as the official authority representing Latvia in international identifier systems.

Latvia is currently exploring possibilities to implement the ISNI at the national level. Discussions are ongoing regarding potential governance models, institutional responsibilities, and technical integration with existing national authority

and metadata systems. The objective is to evaluate how ISNI could enhance interoperability, improve author identification, and support cultural heritage data infrastructure in Latvia.

2 First object – Thesis / Dissertation: "Methods for Building Semantic Portals" including scientific articles

2.1 Metadata (2025)

Image 1. Cover of Osma Suominen's PhD thesis discussed below. (Cover art by Osma Suominen.)

{
  "@context" : {
      "dc" : "http://purl.org/dc/elements/1.1/",
      "dcterms" : "http://purl.org/dc/terms/",
  },
"dc:creator" : ["Osma Suominen", "https://isni.org/isni/0000000484082102"],
"dc:identifier" : ["urn:ISBN:9789526052540", "urn:ISSN:17994942"],
"dc:title" : "Methods for Building Semantic Portals",
 "dc:subject" : ["semantic web", "faceted search", "automatic subject indexing",
"vocabulary quality"],
"dc:publisher" : "https://www.aalto.fi/",

 "dc:date" : "2013-09-09T00:00:00Z",
"dc:type" : "Article-based doctoral dissertation",
"dc:format" : "pdf",

"dc:language" : "en",
"dc:rights" : "All rights belong to the author. You may download, display and print this publication for your own personal use. Commercial use is prohibited."
}

Note that the above metadata is in Dublin Core format. The library system contains more bibliographic metadata in MARC 21 format but in the interest of brevity and readability only the Dublin Core version is included here. There are also separate records for the printed version as well as the electronic publication. Please note that many metadata formats exist and different sectors use different vocabularies (see Annex 1: Terms and Definitions).

2.2 Life cycle (2025)

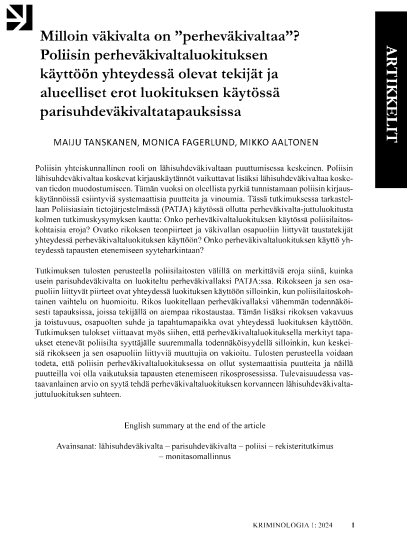
1. **Writing the thesis:** The doctoral candidate is the sole author of at least the introduction of the thesis, and copyright emerges automatically based on the criteria detailed in the law.
   However, the complete thesis may contain articles describing scientific work written with other collaborators, which may have been published in various locations. This object contains five different articles with six different authors. In this case the main author obtains the necessary information and permissions from the other authors. The thesis and its cover may also contain illustrations with their own authors.
2. **Supervisor approval and pre-examination:** The thesis supervisor(s) review the work and give approval for it to be submitted for external examination. Its scientific quality is ensured and pre-examiners provide a report recommending acceptance, rejection, or further revisions. This phase does not impact authorship and creation of copyright.
3. **Thesis defence:** If the pre-examiners approve the thesis, a public defence is arranged, where the author (doctoral candidate) defends their work publicly. This phase does not impact authorship and creation of copyright.
4. **Approval and publication:** If the thesis is approved after the previous step, it can be published by the university. In many universities, theses are published both in print and digital formats, and made available through the university library. The National Library of Finland assigns an ISBN number to both printed and electronic versions. In this example the printed version contains all articles included in the dissertation, reproduced in print with permission from the publishers, but only the introduction is included in the electronic version. This phase involves

transfer of economic rights from the creator to the publisher, while moral rights stay with the creator.

5. **Distribution:** The published thesis becomes available in the university library and may also be added to national and international databases of scientific publications/articles. A legal deposit will be provided to the National Library. The author fills in the thesis metadata when the thesis is published, and the metadata may be edited in the National Library by metadata specialists. We note that when the author completes the metadata as part of the thesis publication process, this is the moment when opt-out expressions should be added alongside other licence information. AI tools have already been integrated into the metadata completion process, and we foresee that such applications will increase in the future, raising questions about how AI-generated and manually created metadata are distinguished.
6. **Availability to readers:** Readers can find and read the thesis through university library collections, databases, or by requesting a copy directly from the creator. Many universities offer digital archives where theses are available through open online access. Due to the differences between the print and electronic versions delineated above, different content may be available in different formats.

3 Second object – Scientific journal article: “Milloin väkivalta on ‘perheväkivaltaa’?”

Image 2. Screenshot of the first page of the article discussed below



ARTIKKELIT

Milloin väkivalta on ”perheväkivaltaa”? Poliisin perheväkivaltaluokituksen käyttöön yhteydessä olevat tekijät ja alueelliset erot luokituksen käytössä parisuhdeväkivaltatapauksissa

MAIJU TANSKANEN, MONICA FAGERLUND, MIKKO AALTONEN

Poliisin yhteiskunnallinen rooli on lähisuhdeväkivaltaan puuttumisessa keskeinen. Poliisin lähisuhdeväkivaltaa koskevat kirjauskäytännöt vaikuttavat lisäksi lähisuhdeväkivaltaa koskevan tiedon muodostumiseen. Tämän vuoksi on olellista pyrkiä tunnistamaan poliisin kirjauskäytännöissä esiintyviä systemaattisia puutteita ja vinoumia. Tässä tutkimuksessa tarkastellaan Poliistasiain tietojärjestelmässä (PATJA) käytössä ollutta perheväkivalta-juttuluokitusta kolmen tutkimuskysymyksen kautta: Onko perheväkivaltaluokituksen käytössä poliisilaitoskohtaisia eroja? Ovatko rikoksen teonpiirteet ja väkivallan osapuoliin liittyvät taustatekijät yhteydessä perheväkivaltaluokituksen käyttöön? Onko perheväkivaltaluokituksen käyttö yhteydessä tapausten etenemiseen syytteeseen? Tutkimuksen tulosten perusteella poliisilaitosten välillä on merkittäviä eroja siinä, kuinka usein parisuhdeväkivalta on luokiteltu perheväkivallaksi PATJA:ssa. Rikokseen ja sen osapuoliin liittyvät piirteet ovat yhteydessä luokituksen käyttöön silloinkin, kun poliisilaitoskohtainen vaihtelu on huomioitu. Rikos luokitellaan perheväkivallaksi vähemmän todennäköisesti tapauksissa, joissa tekijällä on aiempaa rikostaustaa. Tämän lisäksi rikoksen vakavuus ja toistuvuus, osapuolten suhde ja tapahtumapaikka ovat yhteydessä luokituksen käyttöön. Tutkimuksen tulokset viittaavat myös siihen, että perheväkivaltaluokituksella merkityt tapaukset etenevät poliisilta syyttäjälle suuremmalla todennäköisyydellä silloinkin, kun keskeisiä rikokseen ja sen osapuoliin liittyviä muuttujia on vakioitu. Tulosten perusteella voidaan todeta, että poliisin perheväkivaltaluokituksessa on ollut systemaattisia puutteita ja näillä puutteilla voi olla vaikutuksia tapausten etenemiseen rikosprosessissa. Tulevaisuudessa vastaavanlainen arvio on syytä tehdä perheväkivaltaluokituksen korvanneen lähisuhdeväkivalta-juttuluokituksen suhteen.

English summary at the end of the article

Avainsanat: lähisuhdeväkivalta – parisuhdeväkivalta – poliisi – rekisteritutkimus – monitasomallinnus

KRIMINOLOGIA 1: 2024 1

3.1 Metadata (2025)


```
{
  "@context" : {
      "dc" : "http://purl.org/dc/elements/1.1/",
      "dcterms" : "http://purl.org/dc/terms/",
  },
  "dc:creator" : ["Maiju Tanskanen", "https://orcid.org/0000-0002-2328-3563"],
  "dc:identifier" : "https://doi.org/10.54332/krim.143302",
  "dc:title" : "Milloin väkivalta on “perheväkivaltaa”?",
  "dc:subject" : ["police", "intimate partner violence", "multilevel analysis", "registry
study"],
  "dc:publisher" : "https://kriminologia.journal.fi/article/view/143302",
  "dc:contributor" : ["Monica Fagerlund", "https://orcid.org/0000-0001-9588-7959"],
  "dc:contributor" : ["Mikko Aaltonen", "https://orcid.org/0000-0002-3686-2986"],
  "dc:date" : "2024-10-28T00:00:00Z”,
  "dc:type" : "Scientific article",
  "dc:format" : "pdf",
  "dc:language" : "fi",
  "dc:rights" :  "All rights reserved by Kriminalogia."
}
```

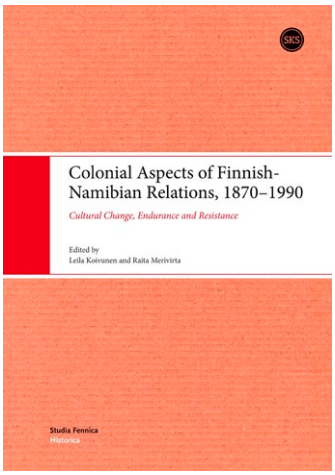

3.2 Life cycle (2025)

1. **Manuscript preparation:** Researchers carry out the research and write the manuscript. They also choose the journal in which they would like to publish their work. The researchers are the authors of the manuscript, and copyright emerges individually, applying to each author automatically based on the criteria detailed in the law. The work may be a joint or combined work.
   Journals customarily provide information about their open-access and licensing policies and conditions. Publishers may have their own opt-out considerations.
2. **Submission:** Authors submit their manuscript to the journal of their choice for consideration. Either at the submission step or upon acceptance the author signs a form wherein the rights are negotiated with the publisher. This is the step at which rights reservation should be expressed in accordance with available mechanisms, as well as indications about the use of generative AI. The publisher may have their own opt-out considerations, which will be part of the negotiation between the author(s) and the publisher.

3. **Preprints:** Some researchers may choose to share their work as a preprint, if this is allowed in the agreement signed in the previous step. Any opt-out expressions and information about use of generative AI (if applicable) should be included in the preprint versions.
4. **Peer review, revisions:** The manuscript undergoes peer review. The peer review process does not create new authorship. Authors may need to revise their manuscript during the process. Earlier versions of the work and the revised version would be considered different versions of the work in instances where, e.g., preprints have already been distributed.
5. **Acceptance and publication:** Once accepted, the article is published in the journal. The considerations for opt-out and generative AI use expressions are the same for the preprints and published version, but as these are distinct versions, they may have different information related to these aspects.
6. **Dissemination, indexing, and archiving:** Libraries and authors ensure the article and relevant data are indexed and stored in appropriate repositories and databases. When the work is digitally preserved as a legal deposit copy at the National Library, this version is available, in a relatively limited manner, as part of the national cultural heritage.

4 Third object – eBook: "Colonial aspects of Finnish-Namibian relations"

Image 3. Cover of the book Colonial Aspects of Finnish – Namibian relations, 1870-1990 discussed below.



4.1 Metadata (2025)

{
  "@context" : {
      "dc" : "http://purl.org/dc/elements/1.1/",
      "dcterms" : "http://purl.org/dc/terms/",
  },
  "dc:creator" : ["Leila Koivunen", "https://isni.org/isni/ 0000000036621889"],
  "dc:identifier" : "urn:ISBN:9789518588866",
  "dc:title" : "Colonial aspects of Finnish-Namibian relations",
  "dc:publisher" : ["Finnish Literature Society", https://kirjat.finlit.fi/sivu/tuote/colonial-aspects-of-finnish-namibian-relations-1870-1990/5124338, "https://www.jstor.org/stable/jj.16971622"],
  "dc:contributor" : ["Raita Merivirta", "https://orcid.org/0000-0002-8396-1098", "https://isni.org/isni/0000000077374212"],
  "dc:date" : "2024-01-01T00:00:00Z",
  "dc:type" : "Book with illustrations",
  "dc:format" : "ePub",
  "dc:language" : "en",
  "dc:rights" : "https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en"
}

4.2 Life cycle (2025)

1. **Manuscript preparation, authoring, and peer review:** The author(s) write(s) and edit(s) the manuscript which usually also goes through a peer review process when it is a scientific work.
2. **ISBN (International Standard Book Number) allocation:** The publisher obtains an ISBN for both the eBook and the physical copy. Each format requires a separate ISBN.
3. **Digital and print formatting:** The manuscript is formatted for print. It is also converted into eBook formats. Depending on how the opt-out will be implemented, different formats may have different opt-out expressions. If generative AI has been used in some parts of the manuscript, this should be indicated in machine readable form.
4. **Cover design:** A cover design is often added for both the physical copy and the eBook. The cover may be designed by an artist, in which case this work will have its own copyright. If the cover has been created with generative AI in a manner which requires indication of use, this would need to be indicated in machine readable form. The same applies to other images and figures in the book.

5. a) **Printing (physical copy):** The physical book is printed. (The physical copy and its metadata may also contain the relevant opt-out and generative AI use information.)
   b) **Digital conversion (eBook):** The manuscript is converted into digital formats suitable for distribution as an eBook. The eBook file may contain opt-out and generative AI use information.
6. **Legal deposit and metadata submission:** Legal deposits of electronic and physical books must be submitted to the National Library by law. Detailed bibliographic information is provided to the National Library, including author details and possibly identifiers, publication date, ISBNs, and other relevant metadata. The metadata may come to the National Library in formats such as ONIX.
7. **Distribution:** the publisher is responsible for distribution of the physical and eBooks, not the National Library.
8. a) **Cataloguing:** The National Library catalogues the book, making it searchable and accessible. In the cataloguing process ISNI identifiers are added for all agents that are within the scope of the National Bibliography.
   b) **Archiving:** Both the physical copy and digital versions are archived.
9. **Access and lending:** The National Library, among others, provides access to the physical copy and the eBook.
10. **Updates:** Depending on how the opt-out is implemented, an opt-out expression for the same work may be added or changed later. The opt-out process for new editions also needs to be addressed separately. The issues concerning update of an opt-out expression are comparable to those where the author wants to edit an existing licence.

5 Fourth object – image: "XX Vispārējie latviešu dziesmu un X deju svētki"

Image 4. XX Vispārējie latviešu dziesmu un X deju svētki

5.1 Metadata (2025)

{
  "@context" : {
      "dc" : "http://purl.org/dc/elements/1.1/",
      "dcterms" : "http://purl.org/dc/terms/",
  },
  "dc:creator" : ["Leons Balodis", "http://www.isni.org/0000000110470290"],
  "dc:identifier" : "https://dom.lndb.lv/data/obj/496922",
  "dc:title" : "XX Vispārējie latviešu dziesmu un X deju svētki",
  "dc:subject" : ["Mūzikas instruments", "Mežrags"],
  "dc:description" : "Svētku dalībnieki Mežaparka estrādē",
  "dc:publisher" : ["Latvijas Nacionālā Bibliotēka", "https://dom.lndb.lv/data/obj/496922"],
  "dc:date" : "1990-07-08T00:00:00Z",

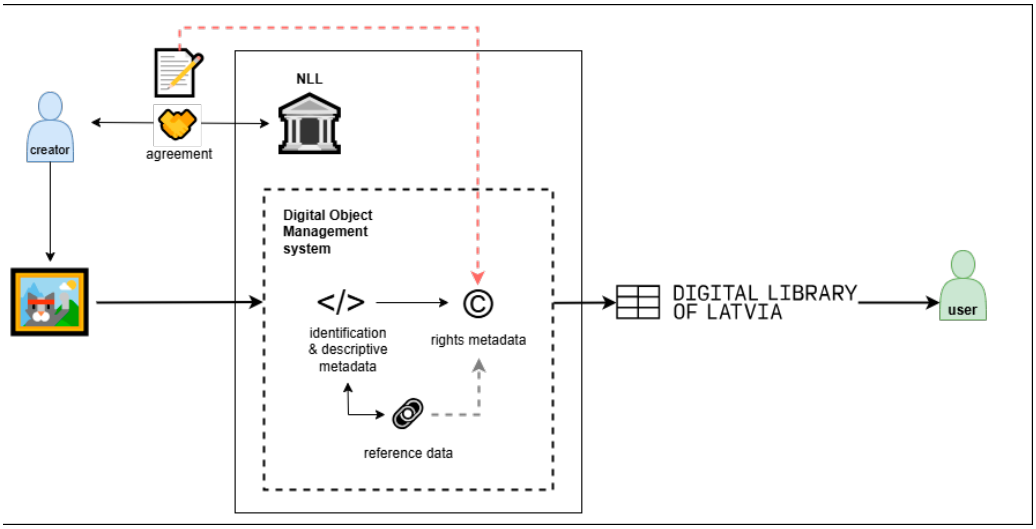
"dc:type" : "Photograph",
"dc:format" : ["image", "8.16 MB"],
"dc:language" : "lv",
"dc:rights" : ["Protected by copyright", "Available on the web"]
}

5.2 Life cycle (2025)

1. The author creates an image.
2. Terms and conditions are agreed between the National Library of Latvia (NLL), which collects images, and the creator.
3. The creator provides the image to NLL.
4. NLL creates a digital object record for the image. This includes:
   a. creating the **descriptive metadata**, enriched with **reference data** for linking to authority records and/or controlled vocabularies, and inventory/stocktaking metadata, which helps in internal tracking and resource management;
   b. uploading the image files and providing detailed descriptions of the files;
   c. indicating **rights metadata**: copyright-related values are added to the record using *APLIS*. These values determine access and use conditions according to the Terms and Conditions agreement.
5. The digital object is available for discovery and access in the Digital Library of Latvia.
6. NLL offers its collections of both digitalised and digitally created texts and images for use in computerised analysis (TDM) for research and study purposes.

5.3 Systems and data flows of the National Library of Latvia (2025)

Image 5. Typical life cycle of an image in the National Library of Latvia's Digital Workflow.



5.4 Life cycle of an image in the Digital Object Management System (DOM) (2025)

Image 6. First step of digital object creation in DOM

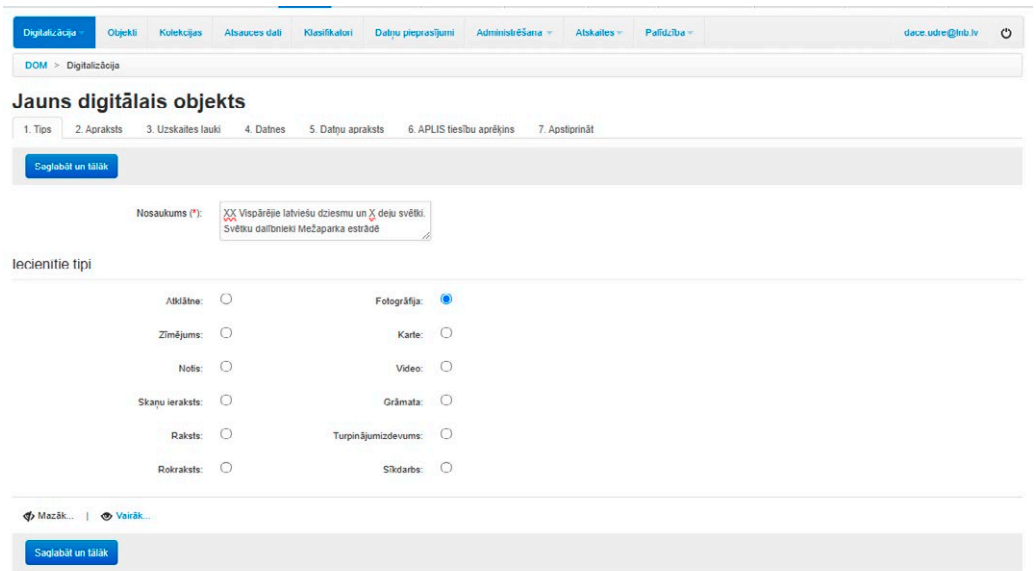


Image 6 shows the creation of a new object in DOM, title of resource, and selection of type, here "photograph".

Next steps:

2. Description (metadata with reference data linking).
3. Inventory or stocktaking fields.
4. Files added.
5. Description of added files.
6. *APLIS* calculation of access rights (see next six images).
7. Confirmation.

Image 7. *APLIS* – Initiating Access Rights – Step 1

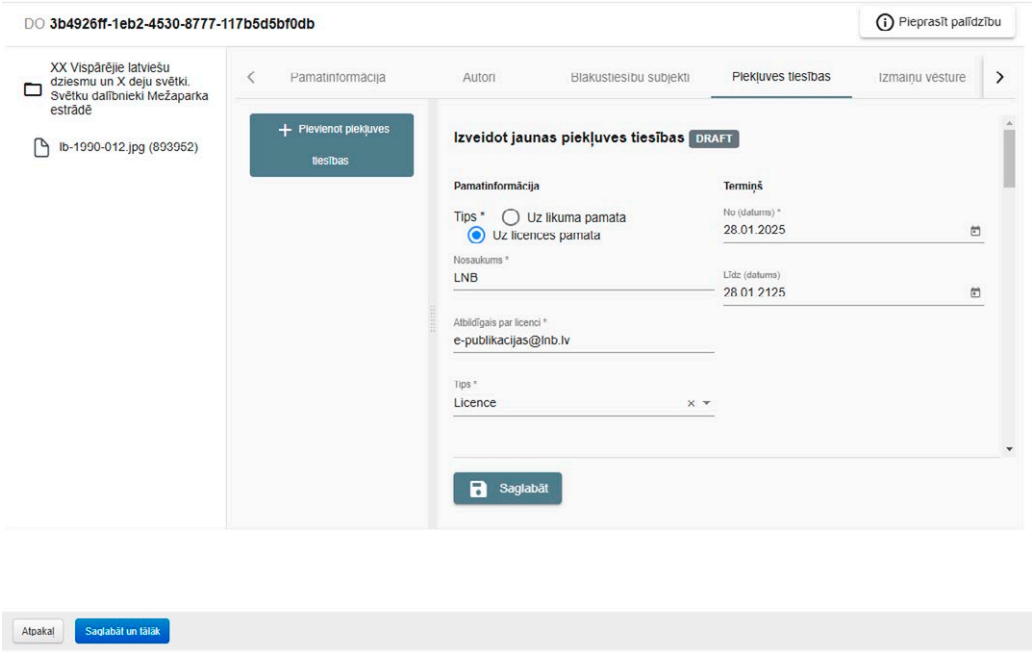


Image 7 shows the initial stage of creating access rights metadata for a digital object. *APLIS* allows the user to define whether access rights are established "on a legal basis" or "on the basis of a licence." In this case, the user selects the latter.

Key fields:

- **Type:** "*Uz licences pamata*" (On the basis of a licence)
- **Name:** Institution or entity managing the licence (e.g., "LNB" – National Library of Latvia)
- **Responsible for licence:** Contact email
- **Valid from / to:** Specifies the licence's validity period (e.g., from 28.01.2025 to 28.01.2125)
- **Licence Type:** "Licence" selected

The metadata entered here is the foundation for more detailed licence configuration in subsequent steps.

Image 8. *APLIS* – Licence Details – Step 2

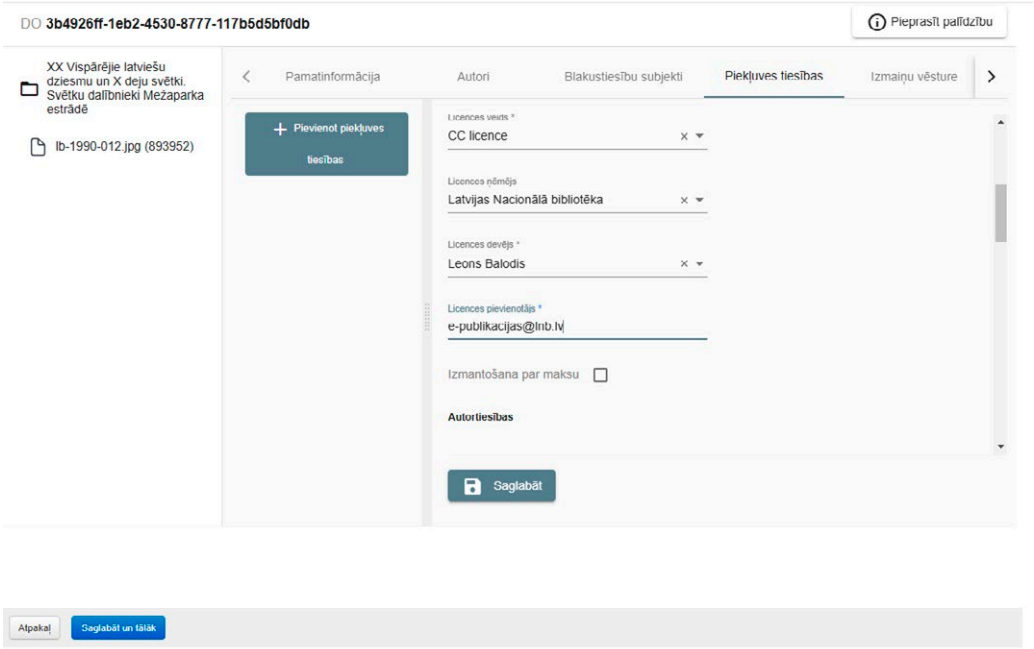


Image 8 shows the detailed licence metadata entry form. Here, the specific characteristics of the licence are documented.

Key fields:

- **Licence Type:** "*CC licence*" (Creative Commons licence)
- **Licence Recipient:** National Library of Latvia
- **Licence Grantor:** Leons Balodis
- **Licence Entered by:** Email address of person submitting the data
- **Paid Use:** A checkbox indicating whether access involves payment (not selected in this example)

These entries define the legal and organisational context in which the object can be accessed and reused, including attribution and rights transfer details.

Image 9. *APLIS* – Public Access and Use Terms – Step 3

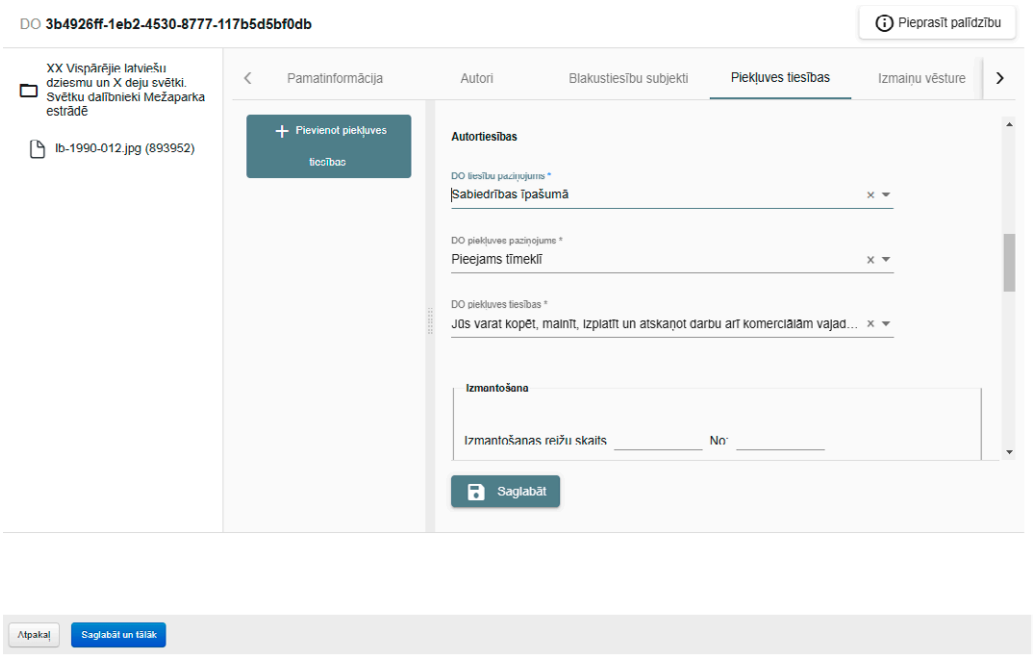


Image 9 shows how public-facing rights statements and reuse permissions are set.

Key fields:

- **Rights Statement:** "*Sabiedrības īpašumā*" (Public domain)
- **Access Statement:** "*Pieejams tīmeklī*" (Available online)
- **Access Rights:** A textual field stating reuse rights, in this case: "You may copy, modify, distribute, and perform the work, even for commercial purposes..."

Image 10. *APLIS* – Usage Limits – Step 4

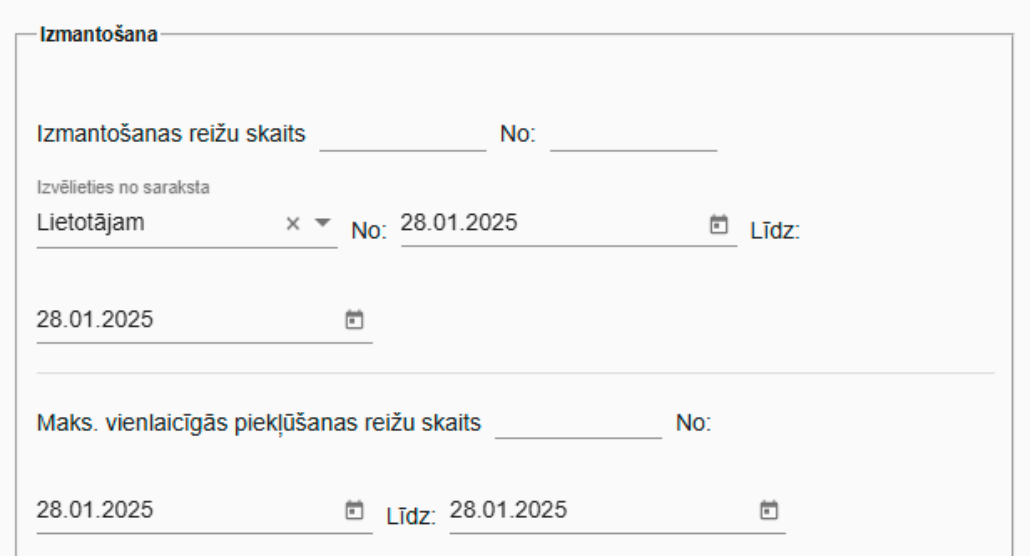


Image 10 shows how *APLIS* allows the entry of detailed usage conditions, such as:

- **Usage count** (*Izmantošanas reižu skaits*)**:** Defines the number of times an object may be accessed.
- **User-specific usage permissions:** Allow configuration of access rights for individual users or user groups.
- **Simultaneous access limit** (*Maks. vienlaicīgās piekļūšanas reižu skaits*): Limits how many users can access the object at the same time.

This ensures time-bound or user-bound restrictions for controlled dissemination of the object.

Image 11. *APLIS* – Usage Limits – Step 5

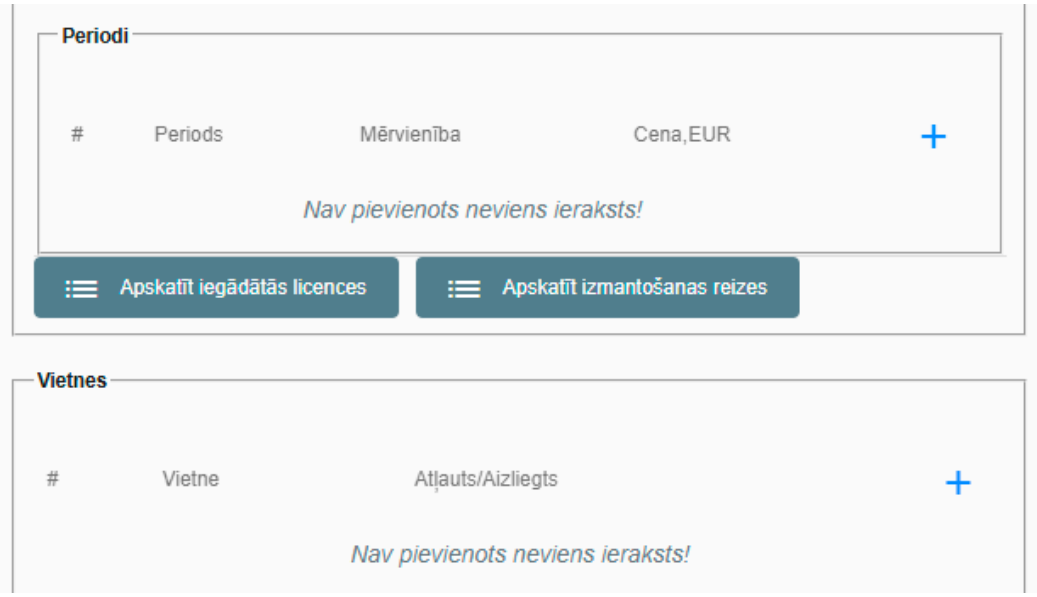


Image 11 shows how *APLIS* allows management of access periods and site-based permissions:

- **Periods:** indicates the period (in hours, days, months, or years) during which the licence is valid.
- **Sites:** Allows definition of which web domains are permitted or denied access to the object.

Image 12. *APLIS* – Usage Limits – Step 6

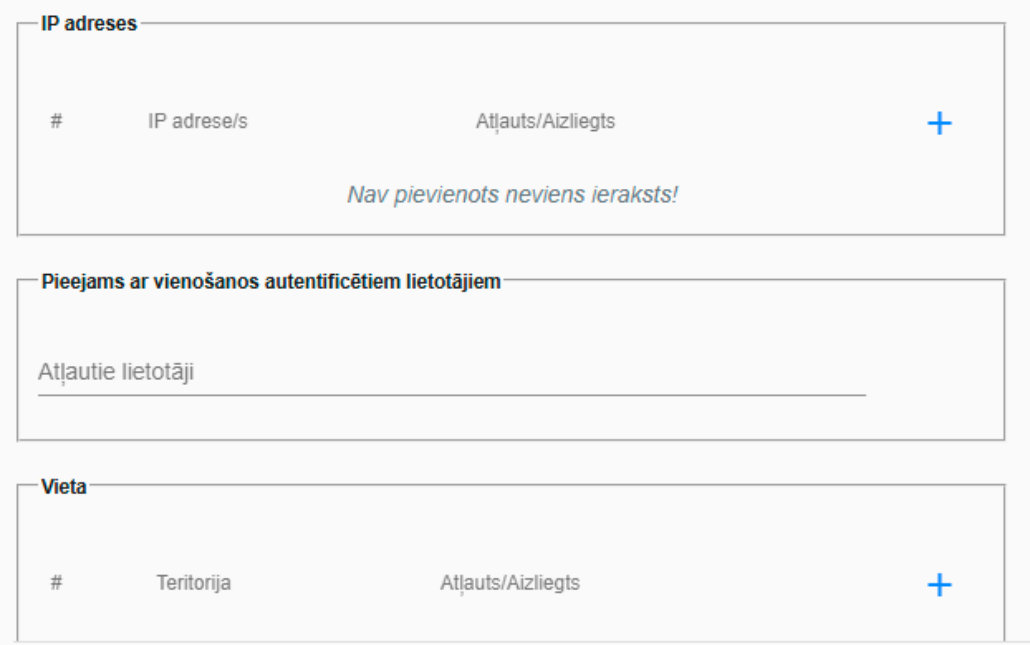


Image 12 shows:

- **IP address permissions:** Access can be allowed or denied based on specific IP addresses or address ranges.
- **Authenticated user access:** Only listed users or user groups may access the object after login.
- **Geographic access rights:** Specific territories or regions may be assigned allowed or forbidden status.

No opt-out mechanisms have been implemented or announced on the technical side. Regarding the accessibility of digital material in the Digital Library of Latvia, we have encountered cases where authors request to opt out.

The National Library of Latvia offers its collections of both digitalised and digitally created texts and images for use in computerised analysis (TDM) for research and study purposes. The library offers to prepare, for example, text corpora for a specific author or newspaper on request for cases when it is necessary to perform an in-depth full-text search and analysis. Only non-copyrighted text corpora are transferred or published outside the library.

6 Fifth object – article: "Sadarbība - Turcijas bibliotēku attīstības virzītājspēks"

Image 13. The first page of the article "Sadarbība - Turcijas bibliotēku attīstības virzītājspēks".

Sanita Burķite Ilga Rampāne Aldona Volkova

Sadarbība — Turcijas bibliotēku attīstības virzītājspēks

Projekta dalībnieces pie Stambulas Tehniskās universitātes galvenās bibliotēkas: (no kreisās) Sanita Burķite, Ilga Rampāne, Aldona Volkova

Stambulas Universitātes centrālās ieejas vārti

austrumu kultūra, senatnīgums ar mūsdienīgo, radot neatkārtojamu krāsojumu. Prakses laikā bija lieliska iespēja iepazīties ar izcilākajām Turcijas bibliotēkām, sastapt brīnišķīgus cilvēkus, kuri palīdzēja iepazīt ne tikai bibliotēku vidi, bet arī kultūru, tradīcijas un iemīlēt šo "kukurūzu pērļu" un tautu. Pārliecinājāmies, ka Turcija ir moderna valsts ar bagātu vēsturi un kultūru, viesmīlīgiem cilvēkiem. Jāatzīst, ka Latvijā ir vērtspusēji, bieži par kļūdaini priekšstati par šo valsti, minimālas zināšanas par Turcijas bibliotēkām. Autores ar vairākiem rakstiem cer paplašināt Turcijas bibliotēku sistēmu, profesionālo vidi, atsevišķām bibliotēkām, iepazīstināt ar atklājumiem, pārsteigumiem un atziņām, kā arī iedvesmot lasītājus patstāvīgam izziņas procesam. Šajā rakstā būs iespēja iepazīt Turcijas vēsturi, bibliotēku sistēmu, profesionālās organizācijas un to darbību, kas sniegs priekšstatu

Raksta autores Eiropas Savienības Mūžizglītības programmas Leonardo da Vinči apakšprogrammas Mobilitātes projekta (2010-1-LV1-LEO02-00663, projekta vadītāja Latvijas Universitātes bibliotēkas (LUB) direktore Iveta Gudakovska) ietvaros divas nedēļas pavadīja Stambulā (Turcijā — pilsētā, kurā nemanāmi saplūst rietumu un

65

6.1 Metadata (2025)


```
{
  "@context" : {
      "dc" : "http://purl.org/dc/elements/1.1/",
      "dcterms" : "http://purl.org/dc/terms/",
  },
  "dc:creator" : ["Sanita Dubra", "https://ats.lndb.lv/object-search/person/249278"],
  "dc:identifier" : "urn:ISSN:16915631",
  "dc:title" : "Sadarbība - Turcijas bibliotēku attīstības virzītājspēks",
  "dc:description" : "Raksts publicēts žurnāla "Bibliotēku Pasaule" Nr.51/52 (2011)",
  "dc:publisher" : ["Latvijas Nacionālā Bibliotēka", "https://dom.lndb.lv/data/
obj/48340"],
  "dc:contributor" : ["Ilga Rampāne", "https://isni.org/isni/0000000072811209"],
  "dc:contributor" : "Aldona Volkova",
  "dc:date" : "2011-01-01T00:00:00Z",
  "dc:type" : "Text",
  "dc:source" : ["Bibliotēku Pasaule;51/52", "https://dom.lndb.lv/data/obj/48340.
html"],
  "dc:language" : "lv",
  "dc:rights" :  ["Protected by copyright", "Available on the web"]
}
```


6.2 Life cycle (2025)

Note: The digital object chosen here is an individual article. This is an exception, since individual articles are not usually described in DOM; the standard practice is to describe entire periodical issues. Ideally, articles would be described individually, with a similar workflow applied. The workflow described below reflects the standard **issue-level workflow**.

1. The author(s) create(s) an article. The article may contain elements, such as photographs, which have their own authors.
2. Articles are then included in periodical issues, and the publisher publishes the issue.
3. The publisher supplies the issue (not the individual article) as a **legal deposit** to NLL.
    - Delivery methods: **small publishers** typically send issues via email or make them available for direct download; **large publishers** deliver issues via FTP.

4. NLL negotiates with the publisher to enable broader access and use (through Terms and Conditions agreements).
5. NLL creates a digital object record for the **issue** in DOM. This includes:
   a. entering the **descriptive metadata**, enriched with reference data for linking to authority records and/or controlled vocabularies, and inventory/stocktaking metadata, which helps in internal tracking and resource management;
   b. uploading the **issue** files and describing them;
   c. indicating **rights metadata**: copyright-related values are added to the record using *APLIS*. These values determine access and use conditions according to the Terms and Conditions agreement.

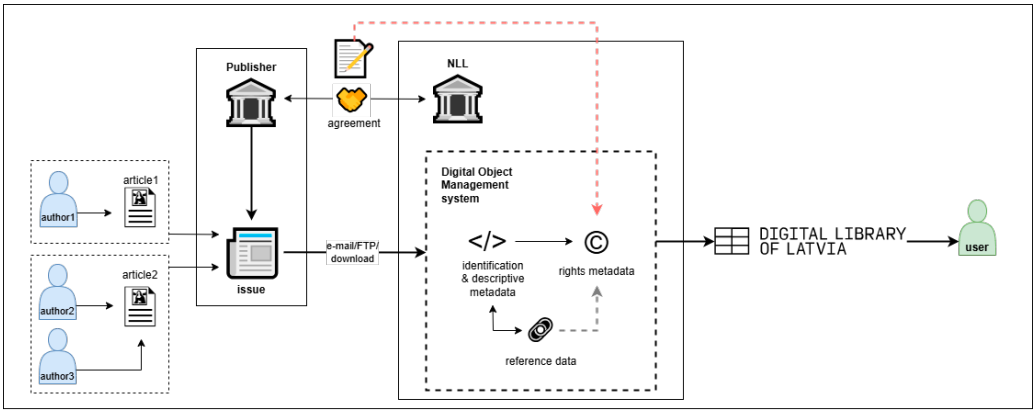
APLIS calculates the copyright status (copyright protection term) based on the digital object type and descriptive metadata. If the metadata specifies authors, the system calculates the copyright protection term for each author individually. When a digital object is created, the system determines:

– Access rights based on the law;
– Access rights based on a licence, if a relevant licence has been entered into the system.

6. The digital object is available for discovery and access in the Digital Library of Latvia.

6.3 Systems and data flows of the National Library of Latvia (2025)

Image 14. Typical life cycle of an issue of a periodical according to the National Library of Latvia's **Digital** Workflow.



Note: The lifecycle of an individual article would follow a similar process.

6.4 Life cycle of an article in the Digital Object Management System (DOM) (2025)

Image 15. Creation of a new object in DOM

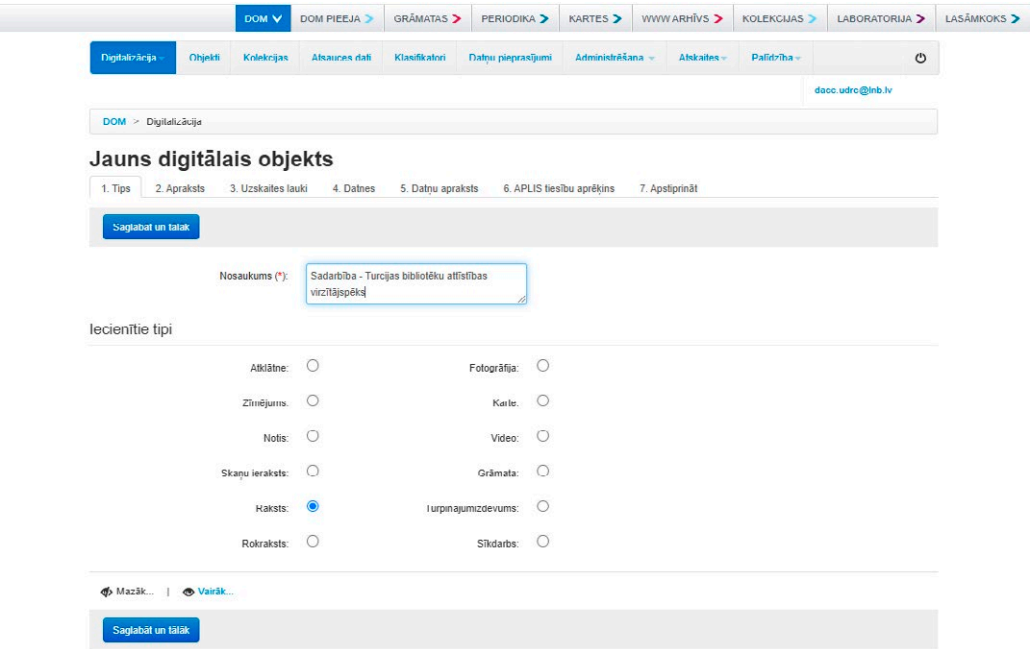
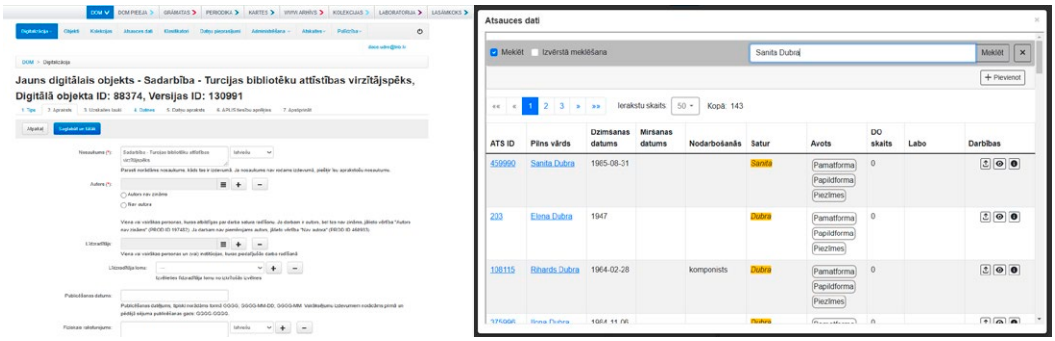


Image 15 shows creation of a new object in DOM, title of resource and selection of type, here: "article".

Image 16 shows the second step of the workflow in DOM:

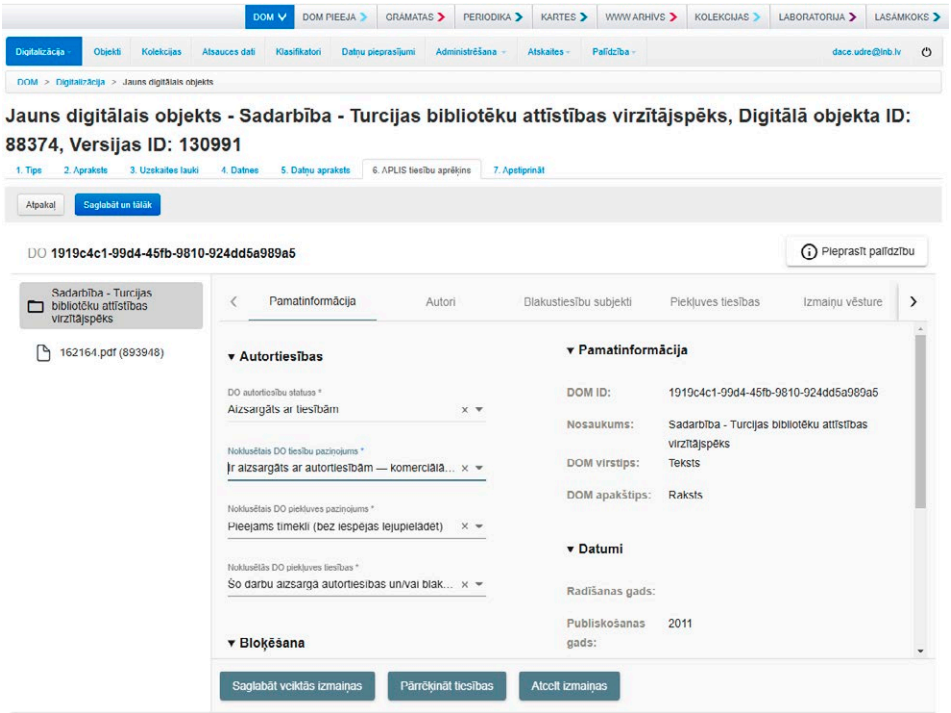
Image 16. Metadata creation view in DOM and iframe to the Reference Data System



Next steps:

3. Inventory or stocktaking fields.
4. Files added.
5. Description of added files.
6. *APLIS* calculation of rights (see below).
7. Confirmation.

Image 17. Copyright and access values in *APLIS*



Rights metadata: Copyright and access values (*APLIS* Copyright Management and Licensing System).

- Copyright status – Protected by rights.
- Digital Object Rights statement – In copyright – out of commerce.
- Digital Object Access notification – Available online (no download option).
- Digital Object Access Rights – This work is protected by copyright and/or related rights. This work may be freely used for personal use, scientific research, or self-education. Other uses must be authorised by the rightsholder(s).

The *APLIS* system contains the following copyright statuses:

- Not subject to copyright – the subject matter is not the author's work – natural objects, documents according to the Copyright Law of Latvia;
- Not copyrighted – subject matter is the author's work but is no longer protected – 70 years have passed since the death of the author or last co-author;
- Protected by copyright – subject matter is the author's work and is still protected – author is alive or 70 years have not passed;
- Copyright not clear - subject matter is author's work but author/authors unknown or no data about year of death (including when author is still alive and data of creation is unknown).

Each copyright status has its own access statement or several statements resulting from the rights status. In the case of multiple notifications, the employee must choose the most appropriate one for the specific situation.

Annex 3: Requirements

Introduction

We consider two main clusters of requirements.

Cluster 1.

The EU Acquis Communautaire shall be easily implementable and efficiently enforceable; as such we derive **data requirements** from three regulatory documents:

- Directive (EC) 2001/29 on the harmonisation of certain aspects of copyright and related rights in the information society, which states:
  - Exclusive rights in the digital environment.
  - Rights management information, such as identifiers and metadata in cases where they contain rights management information or are part of it, is protected against removal and alteration.

- Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market, which states:
  - Reproduction for text and data mining is allowed if the terms in the exception (Arts. 3 and 4) apply (legal access, three-step test).
  - Creators of original material shall be fairly, appropriately, proportionally and transparently remunerated.

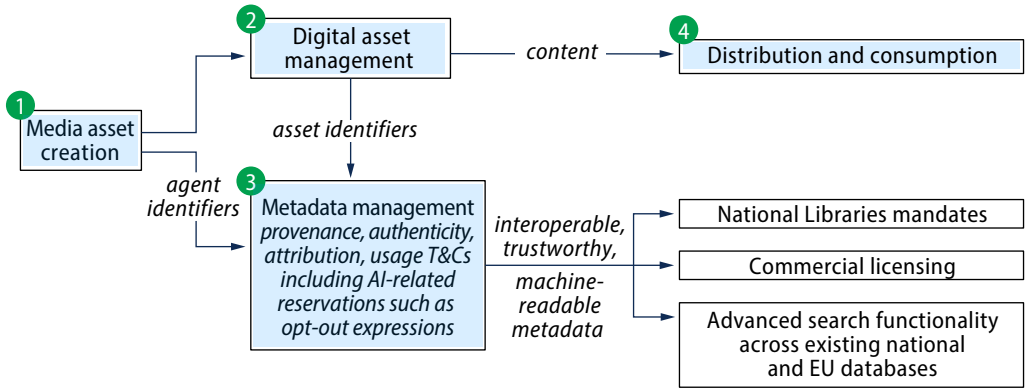
- Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence, which states:
  - Obligations to respect copyright law, including opt-out declarations, require that rights-management information, interoperable identifiers, and AI-related expressions are standardised or otherwise made clear and concise.
  - Traceability requires identification of content that has been generated by AI.
  - Transparency requires documentation of the training and generation algorithms used to produce AI-generated content.

Cluster 2.

RIGHTS-MANAGEMENT INFORMATION[13] shall be machine-readable, trustworthy, and interoperable.

We look at what standards and technologies would be required in 2030 and beyond to facilitate National Libraries mandates[14], commercial licensing, and an advanced search functionality across existing national and EU databases across the four main stages of the life cycle of any media asset:

Figure 1. Context of the requirements



Notes:

- The following requirements do not apply equally to National Libraries mandates, commercial licensing, or advanced search functionality across existing national and EU databases.
- Throughout the standards development, "shall" denotes an obligation, "should" denotes a recommendation, and "may" denotes an option.
- The applicability of requirements is signalled by:
  - [NL]: for National Libraries.
  - [CL]: for commercial licensing in the respective creative industries.
  - [IR]: for interconnected repositories

13 See section 5.1: Harmonisation of certain aspects of copyright and related rights in the information society, Article 7 (2).

14 As described in section 2.2.

Generally, identification, rights, and usage metadata shall facilitate the proactive implementation and reactive enforcement of the Acquis Communautaire related to copyright, data, and finance through open standards and effective technologies.

1. **Media asset creation**
   References to the use case lifecycles (see Annex 2):
   - Thesis: steps 1) to 4)
   - Scientific article: steps 1 and 2, then 4) and 5)
   - eBook: steps 1), then 3), 4), then 10)
   - Image: step 1)
   - Article: step 1)

   1.1. Information required at point of creation
   1.1.1. As rights are granted to authors for their works, AUTHOR and WORK shall be defined for various creative sectors.

   1.1.2. AUTHORS and CONTRIBUTORS shall be distinguished; CONTRIBUTOR ROLES should be standardised by creative sector and declared.

   Benefit: accurate remuneration of some usages based on industry practices, e.g., standardised royalty splits.

   Challenge: contributor roles should also be standardised and translated across countries.

   1.1.3. AUTHORS and CONTRIBUTORS shall be identified with standard and persistent identifiers at the highest possible levels of trust and interoperability.

   Challenges: handling anonymity, aliases, and data privacy; dealing with natural and legal persons.

   1.1.4. AGENT IDENTIFIERS for AUTHORS and CONTRIBUTORS shall be issued by trusted authorities that recognise each other.

   Benefits:
   - *Streamlined online transactions by exchanging trusted identifiers instead of lengthy descriptors prone to typos and misspellings.*
   - *Identifiers impose minimum requirements for metadata to be collected about an AGENT, thus ensuring better interoperability.*

- *Trusted identifiers can be used to distinguish AGENTS without using any sensitive personal data.*
- *Facilitation of authority and access controls.*
- *AGENT IDENTIFIERS recorded at the time of creation or publication can be maintained throughout the lifecycle of the digital asset and included in a persistent way.*

Challenge: AGENT IDENTIFIERS, e.g., ISNI, require a publication or WORK, which imposes timing requirements on the process of binding AGENT and WORK Identifiers as early as possible. It should be possible to acquire an AGENT IDENTIFIER based on preliminary metadata (pre-bibliographic record).

1.1.5. Various AGENT IDENTIFIERS shall have clearly defined roles with explicit mappings between them that define when they can be used interchangeably and when they differ in scope or usage.

Challenge: Some AGENT IDENTIFIERS, such as ISNI, are designed to disambiguate public identities, and are intended to be used as bridge identifiers between other databases if a connection to the actual identities of the authors and related information is needed. This is a challenge to interoperability within larger infrastructure.

1.1.6. AUTHORS and all CONTRIBUTORS and their IDENTIFIERS shall be unambiguously resolvable for humans and machines within and outside the system identifying them.

- *If an opt-out expression is later changed, it is necessary to know who has the right to make such a change.*
- *If there are terms and conditions within the rights expression that allow certain uses, it is necessary to know with whom any amendments on these uses should be negotiated.*
- *If remuneration models move toward the possibilities outlined in Section 5 of the report, then there will be an increasing need to identify contributors in an unambiguously resolvable manner.*

Challenge: Unambiguous resolution of a public identity is a different task from resolving the author's identity.

1.1.7. Technically derivable[15] from Art. 19 §4 of the CDSM directive EU/2019/790: attributions should include CONTRIBUTORS' names, identifiers, roles, and entitlements [CL].

Challenges:

- *Percentage of contribution is not a value that can be expressed in the metadata formats in use in the library sector. To define authors and contributors, which is the main purpose of the library metadata, this information is not necessary or possible to determine in practice. The percentages can be defined in the contracts and agreements between the creators of the work and the publishers, but this is not usually seen as part of or appropriate for the metadata.*
- *Percentage of contribution is not set by directives, but by agreements between rightsholders or by CMO practices. However, Article 19.3 states that the obligation set out in paragraph 1 shall be proportionate and effective in ensuring a high level of transparency in every sector. Member States may provide that in duly justified cases where the administrative burden resulting from the obligation set out in paragraph 1 would become disproportionate in the light of the revenues generated by the exploitation of the work or performance, the obligation is limited to the types and level of information that can reasonably be expected in such cases. As such the issue depends on the copyright law of each EU member state.*

1.1.8. A WORK (i.e., ABSTRACTION) and its EXPRESSIONS, MANIFESTATIONS, and ITEMS shall be distinguished; EXPRESSIONS, MANIFESTATIONS, and ITEMS shall be unambiguously linked to the WORK they relate to at the highest possible levels of trust and interoperability.

Challenges:

- *Tracking file compressions, e.g., higher resolution for printing and lower resolution for display on the web.*
- *Management of trust indicators to achieve the highest possible level of trust.*
- *Definition of interoperable identifiers for WORKS, EXPRESSIONS, MANIFESTATIONS, and ITEMS.*

15 Technically derivable requirements are implicit requirements which should be verified as resulting standards could contribute to a better functioning of the copyright eco-system.

1.1.9. WORK IDENTIFIERS shall be issued by trusted authorities that recognise each other.

1.1.10. Various WORK IDENTIFIERS shall have clearly defined roles with explicit mappings between them that define when they can be used interchangeably and when they differ in scope or usage.

1.1.11. Types of MEDIA ASSETS should be standardised by creative sector and expressed in metadata.

1.1.12. A minimum set of METADATA may be defined for specific creative sectors and METADATA exchanges.

1.2. Traceability and transparency at creation

1.2.1. Technically derived from Art. 50 §2 of the AI Act EU/2024/1689[16]: applying to any and all Generative AI (GenAI) systems generating any type of media, including generation and manipulation, while not applying to assistance for standard editing:

GenAI output shall be tagged when required, and the tag shall be machine-readable and detectable.

This tagging process should be effective (tagging what should be tagged), interoperable, robust, and reliable.

Challenges:
- *The metadata formats currently in use, for example in the library sector, do not support indication that generative AI has been used.*
- *Besides content, Generative AI may also be used in the future to create metadata, which would require additional tagging mechanisms.*

16 Codes of practice may be enacted (as long as they are not, they remain recommendations).

1.2.2. Technically derived from Art. 50 §4 of the AI Act EU/2024/1689[17]: applying to generation and manipulation of deep fakes and text which is published with the purpose of informing the public on matters of public interest, while not applying when GenAI content is reviewed and/or edited by human(s) and published under editorial responsibility of a natural or legal person:

GenAI output shall be disclosed as such and the disclosure shall be human-readable and detectable.

Human-made and AI-generated content shall be distinguished and identified.

Benefits:
- *Consumers can use metadata to filter materials depending on whether they are generated by AI or not.*
- *Encourages responsible use of AI by requiring disclosure.*

Challenges:
- *Identification of hybrid assets, e.g., a book translated by AI but edited by a human.*
- *Definition of minor and major uses of AI.*
- *Definition of where human creativity begins and ends.*

1.2.3. Technically derivable from Art. 50 §5 of the AI Act EU/2024/1689[18]: tags (from Art. 50 §2) and disclosures (from Art. 50 §4) shall be clear, distinguishable, and immediate.

1.2.4. AI algorithms shall generate METADATA at each step of the training and usage processes of AI systems.

1.2.5. AI algorithms shall be identified.

Challenges:
- *Some creators or organisations may resist disclosing the use of AI algorithms due to competing requirements regarding the*

17 Codes of practice may be enacted (as long as they are not, they remain recommendations).

18 Codes of practice may be enacted (as long as they are not, they remain recommendations).

confidentiality of intellectual property and sensitive data, or it might be unclear when and to what degree to disclose AI use.
- *When AI is integrated continuously into new tools and software, it may not be clear to the discloser which algorithm has been used. We discussed whether in these cases it would be possible to* identify the *software used: this may also be problematic if software has various AI-generated functionalities.*

1.2.6. Devices such as text/writing software, cameras, mixing tables, etc. should generate METADATA at the moment of creation.

Challenges:
- *In order to enable easy integration of contributor identifiers into devices and applications used to produce content, a standardised way of including identifiers and other metadata is needed.*
- *In this case, devices – and potentially model, firmware version, serial number – should be identified.*

1.2.7. Modifications of a DIGITAL ASSET shall be identified; multiple modifications of a DIGITAL ASSET shall be tracked and traced, including multiple digitisations of the same DIGITAL ASSET.

Technology considerations:
- *Hashing may be used.*
- *Identification methods such as the ISCC may be used to track various versions of a digital file.*
- *Digital signatures and timestamps may be used to increase information trustworthiness.*

1.2.8. Standard generic and sectoral METADATA sets shall be expanded with new terms to signal the human-AI distinction.

2. **Digital asset management**

References to the use case lifecycles (see Annex 2):
- Thesis: steps 5) and 6)
- Scientific article: steps 3) and 6)
- eBook: steps 3), then 6) and 9)
- Image: step 5)
- Article: steps 2), 5.b), 6)

2.1. Information

2.1.1. Digital Asset Management Systems (DAMS) shall be identified.

Challenge: interconnections between systems.

2.1.2. Type of DAMS shall be identified, e.g., legal deposit, voluntary copyright registration, recordation of transfers, collective management, etc.

2.1.3. CREATIVE WORK, original FIXATION, aka master file or master record, EXPRESSIONS, MANIFESTATIONS, and ITEMS of the CREATIVE WORK shall be distinguished and identified.

Challenges:
- *Not all derived works involve new authorship, e.g., a photograph of a painting, a scan of an image, a non-reframed smaller version of a DIGITAL ASSET.*
- *The relationship between EXPRESSION, MANIFESTATION, and "version" should be clarified and standardised.*

Technology considerations:
- *In a Digital Asset Management System there are usually two files – an Archive file (master file or preservation file, high resolution) and a Distribution file (potentially lower resolution), e.g., both have hash values in the National Library of Latvia.*
- *Indeed, one item may have multiple images connected to it, not only an Archive file and a Distribution file, and connecting their identifications may be required.*

2.1.4. Copyright-protected WORKS, OUT-OF-COMMERCE WORKS, and ORPHAN WORKS and other related subject matters shall be distinguished and identified.

2.1.5. FIXATION, and time of FIXATION – a legal concept – shall be identified at the highest possible levels of trust and interoperability.

2.1.6. The FILE FORMAT of the DIGITAL ASSET shall be identified.

2.1.7. The DATA FORMAT of METADATA related to a DIGITAL ASSET shall be identified.

2.1.8. A minimum set of METADATA may be defined for specific creative sectors and METADATA exchanges.

2.2. Traceability and transparency

2.2.1. Machines should be able to distinguish between parts of a DIGITAL ASSET that are created by different authors and different AI systems.

3. **Metadata management**

In this section the requirements focus on attribution, i.e., assertions of author and asset identifiers and expression of terms and conditions for the use of the work, such as reservation of rights.

References to the use case lifecycles (see Annex 2) for attribution:
- Thesis: steps 4) and 5)
- Scientific article: steps 5) and 6)
- eBook: steps 2), then 6) and 9)
- Image: steps 2), 4.a), 4.b), 4.c)
- Article: steps 3), 5.a), 5.c)

References to the use case lifecycles (see Annex 2) for terms and conditions:
- Thesis: steps 4) and 5)
- Scientific article: step 5.b)
- eBook: implicitly in steps 7.b) and 8.b)
- Image: steps 2) and 4.c)
- Article: steps 4) and 5.c)

3.1. Information

3.1.1. Technically derivable from Art. 7 §1 of the Infosoc directive 2001/29/EC: RIGHTS-MANAGEMENT INFORMATION shall only be removed or altered by duly authorised people wherever the information may be stored, embedded in WORKS or other subject-matter, or stored in an external database. Therefore, the authority of a person removing or altering RIGHTS-MANAGEMENT INFORMATION shall be verified, hence also the authority of a person adding or entering RIGHTS-MANAGEMENT INFORMATION. Accordingly, the information assertion should be signed and timestamped.

Challenge: Only the authors or other actors with authorisation are allowed to edit (create, read, update, and delete information) rights-management information, referring here to terms and conditions, licenses, and opt-out expressions, etc. If these are edited, who had the right to make the changes must be known.

3.1.2. TRUST FRAMEWORKS shall be designed, developed, and deployed. They may refer to TRUST RECORDS, TRUST MANIFESTS, TRUST INDICATORS, TRUST PROFILES, and TRUST REPORTS. For example:

- signatures and timestamps should be qualified by listed QUALIFIED TRUST SERVICE PROVIDERS, and
- attribution and T&Cs assertions may be issued as QUALIFIED ELECTRONIC ASSERTIONS OF ATTRIBUTES.

Challenge: managing different levels of trust, e.g., presence of no timestamp, of a timestamp, of a qualified timestamp.

Challenge: It must be noted that QUALIFIED TRUST SERVICE PROVIDERS and QUALIFIED ELECTRONIC ASSERTIONS OF ATTRIBUTES are mechanisms that have not yet been linked to management of publication metadata and copyright infrastructure, the possible integrations and connections of which are still undefined.

3.1.3. Most attributions and terms and conditions are asserted by proxy. It is therefore necessary to be able to verify that these intermediaries, e.g., managers or publishers, are mandated to do so.

3.1.4. It should be possible to verify the granularity of such mandates.

3.1.5. Technically derivable from Art. 7 §2 of the Infosoc directive EC/2001/29: legally-protected RIGHTS-MANAGEMENT Information includes any information, numbers, and code provided by rightsholders to:

- identify AUTHORS or any other rightsholders [NL, CL, IR];
- identify WORKS and other subject matters [NL, CL, IR], or
- assert terms and conditions of use of WORKS or subject matters (T&Cs) [CL].

It is therefore necessary to:

- identify WORKS (and other subject matters), T&Cs, and stakeholders, and
- link WORKS and T&Cs, WORKS and stakeholders, and T&Cs and stakeholders.

3.1.6. Attribution and T&Cs assertions should be automated to some extent. The automated ("by machine") origin of METADATA should be signalled. Automated METADATA generation reviewed by a human should be considered manual / made by human.

3.1.7. Graphic user interfaces (GUIs) and application programming interfaces (APIs) should be intuitive, informative, and adaptive.

3.1.8. As copyright protection is limited in time, the dates of first fixation and the author's death shall be recorded.

Challenge: Legally speaking, when a person dies, their death certificate records the date of death. But if the body cannot be found, or the person disappears, then the law provides for a waiting period (sometimes up to 7-10 years) before a person can be declared legally dead. This is dependent upon the law of the EU member state; there is no EU-wide standard.

3.1.9. Attribution and T&Cs assertions should be valid as long as they are not superseded. Governance, audit, and enforcement processes should be put in place to manage the liabilities of operators of repositories of rights management information. Asset provenance and metadata provenance shall be transparent.

3.1.10. Usage policies shall be coherent, e.g., an allowed use shall not be in conflict with other permissions and licenses. In the case of earlier agreements on uses with permissions, these must be noted somehow.

3.1.11. Misattributions of works should be detectable and identified. The resolution of attribution conflicts should be automated – at least partially with human review.

3.1.12. Usage definitions should be related to both moral and economic rights and be standardised by creative sector.

3.1.13. Technically derivable from Art. 3 §1 of the CDSM directive EU/2019/790: text and data mining is authorised for reproductions and extractions (objectives) by research organisations and cultural heritage institutions (rights users) for the purposes of scientific research (purposes). Therefore, objectives, rights users' affiliation (e.g., 'research organisation' or 'cultural heritage institution'), and purposes shall be identified. This would allow "Distribution and consumption" without remuneration to the author when the conditions of the exception under the three-step test apply.

3.1.14. Technically derivable from Art. 4 §1 of the CDSM directive EU/2019/790: objectives (reproduction and/or extraction) and purpose (text and data mining) shall be identified (to authorise "Distribution and consumption") without the need to pay remuneration in accordance with CDSM Art. 19 §1. Uses that are not in line with the three-step test (under CDSM Art. 7) require a license from the rightsholders (opt-in) and must comply with CDSM Art. 19 §1).

3.1.15. Technically derivable from Art. 4 §2 of the CDSM directive EU/2019/790: Rightsholders may reserve the rights to text and data mining conducted for any other purpose than scientific research. The reservation should be expressed by machine-readable means in the case of content made lawfully, publicly available online. Therefore, the rights expression shall be discoverable and interpretable.

Challenges:

- *Would the absence of a TDM-related reservation in some asset metadata be sufficient to conclude that there is no opt-out, even if such an opt-out reservation has been expressed elsewhere? This is entirely dependent on details of how opt-out declarations are implemented and required to be set up*[19]*: it must be straightforward for AI developers to check opt-out status.*
- *When AI-related reservations are very nuanced, whether they are actually usable and actionable by machines may become a challenge.*

19 See for example the Code of Practice for GPAI model developers at https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai

3.1.16. Generic and sectoral METADATA standards shall be expanded with new terms to signal opt-out expressions and other AI-related expressions.

3.1.17. Opt-out expressions and other AI-related expressions should be human-readable and machine-readable.

3.1.18. AI-related rights expressions should be granular and relate to specific AI usages such as text and data mining, inferencing, training, training for generation, etc. Nuanced machine-readable expressions shall consider new technical developments.

Challenge: technology is moving fast, so keeping up to date with different AI usages can be a challenge. Indeed, it is currently unknown what will be relevant in the future, or how different AI usages will be conceptualised and named.

3.1.19. Due to the highly iterative nature of AI usage, expressions of AI-related rights shall be highly scalable, for example in the case of human-machine hybrid content (e.g., content produced by several human creators – authors and contributors – and several machines), and in the case of multiple generations of hybrid human-machine mixed contents.

3.1.20. The link between any manifestation of a particular WORK and METADATA pertaining to that particular WORK – especially assertions of AI-related rights expressions – shall be robust and standardised, especially in cases of dynamic assertions and amendments of the expressions.

3.1.21. Assertions of AI-related rights expressions shall not hinder the accessibility of contents as intended by their authors and rightsholders.

3.1.22. There should be coordination across platforms to ensure that METADATA is unified, coherent, and consistent, including terms and conditions, especially AI-related reservations and rights declarations for particular WORKS.

3.2. Traceability and transparency of metadata management phase

3.2.1. It shall be possible to recognise what parts of a creative WORK have been produced by one or several humans and what parts by one or several machines.

3.2.2. In some cases, e.g., the remuneration of musicians, it should be possible to identify whether they are members of a collective management organisation.

3.2.3. It should be possible to identify what organisation owns or operates a generative algorithm.

3.2.4. The provenance and authenticity of DIGITAL ASSETS should be expressed. This information shall be updateable – e.g., through signed and timestamped claims, and the infrastructure should be able to fetch the updated information reasonably easily.

4. **Distribution and use of contents by commercial and non-commercial entities**

References to the use case lifecycles (see Annex 2) for distribution:
- Thesis: step 7)
- Scientific article: steps 5) and 6)
- eBook: steps 7), then 8a) and 9)
- Image: steps 3), 5) and 6)
- Article: steps 2) and 6)

References to the use case lifecycles (see Annex 2) for consumption:
- Thesis: step 7)
- Scientific article: step 5), then 6)
- eBook: steps 7), 8a), then 9)
- Image: step 6)
- Article: step 6)

4.1. Information

4.1.1. Technically derivable from Art. 18 (right to fair and proportionate remuneration) of the CDSM directive EU/2019/790: RIGHTS MANAGEMENT INFORMATION shall be reliable, current, and exhaustive, as well as accessible and interoperable.

4.1.2. RIGHTS MANAGEMENT INFORMATION shall be either embedded in or linked to all distributed DIGITAL ASSETS and remain intact throughout the distribution process.

4.1.3. Technically derivable from Art. 19 §1 of the CDSM directive EU/2019/790: transparency requires licensees to have the technical capability to track and report:

- exploitation of works and performances;
- modes of exploitation, to be able to report on the use;
- revenues generated by online platforms and other licensees, and
- remuneration due to authors and other rightsholders.

4.1.4. Distributors shall respect crawler-exclusion tools (robots.txt and the like) and other opt-out mechanisms, and monitor AI-related traffic when so required by the rightsholders of the distributed content.

4.1.5. Distributors should coordinate across platforms to ensure that crawler-related procedures and other opt-out mechanisms support a unified, coherent, and consistent TDM[20] position, and that TDM reservations are expressed in a standardised way.

4.1.6. Exceptions and limitations require validation of the user's eligibility. The legislation often includes conditions that must be met in order for the exception or limitation to apply. There is indeed a need to understand for what purpose a copyright-relevant action is taken. Verifying usage may also require verification of users, unless the exception is mandatory for specific purposes.

Challenges:

- *The conditions are based on directives that allow for national flexibilities. Therefore, they may differ within the EU and require specific attention in due course.*
- *The data processing of exceptions and limitations shall respect the GDPR and trade secrets.*

Additional considerations

The following requirements concern primarily the commercial sector and fall outside the scope of National Libraries. The requirements are uncertain, as the infrastructure and legislation are still developing. Nevertheless, they support

20 TDM = text and data mining.

he overall goal of the report. This section encourages open discussion about the requirements that can be foreseen, in this case with commercial licensors, concerning the requirements in their respective sectors.

4.1.7. Technically derivable from Art. 19 §2 of the CDSM directive EU/2019/790 in the case of commercial licensing: transparency requires licensees and sub-licensees to have the technical capability to track and report:

- exploitation of works and performances;
- modes of exploitation, e.g., archiving, communication to / consultation by the public;
- revenues generated, and
- remuneration due.

Transparency also requires the capability to identify and report sub-licensees, e.g., customers of an image marketplace.

Therefore, rights metadata shall be embedded in or linked[21] to all media assets made available to the public, including attribution and terms and conditions such as use limitations and expiration dates.

Benefits:

- *Providing rights metadata helps distributors stay compliant with copyright laws and licensing agreements.*
- *It reduces the risk of unauthorised use or distribution, avoiding legal liability.*
- *License verification and providing metadata streamlines the management of digital collections, especially across multiple intermediaries and media asset types.*

Challenges:

- *Some media assets (e.g., older ones, particularly outside National Libraries) may lack standardised metadata formats, making it difficult to consistently embed and read rights metadata.*
- *Ensuring compatibility across different platforms, especially with older systems, can be problematic.*

21 E.g., through hard binding or soft binding; see Terms and Definitions.

4.1.8. When required to align with the scope of the license, digital service providers should plan and enforce access control.

Digital service providers shall clearly communicate licence conditions, allowed uses, and usage restrictions at the point of communication to the public. Consumer interfaces should signal the source of the content, its licence, and any attribution or citation requirements.

Digital service providers shall provide a reason for denial when content cannot be streamed (e.g., due to licence expiry or regional restriction).

Challenges:
- *Ensuring the accuracy, timeliness, and granularity of licence data is difficult, and metadata upkeep is resource-demanding.*
- *Sharing data with rightsholders may raise privacy and legal concerns, particularly under data protection laws such as the GDPR.*

4.1.9. Technically derivable from Art. 19 §3 of the CDSM directive EU/2019/790: proportionate obligation requires the identification and description of standards and technologies which allow fulfilment of obligations.

Challenge: the administrative burden should remain proportionate and economically justifiable.

Consumer interfaces should reflect any rights expressions, including exceptions, limitations, and reservations; these limitations and reservations shall be easily understood.

4.1.10. Technically derivable from Art. 20 of the CDSM directive EU/2019/790: additional remuneration requires the technical capability to track and report usages and related revenues.

4.1.11. Rights expressions, a fortiori in the case of automated commercial licensing, should be supported by the standardisation of:

- types/identifiers of users to manage standard use cases, exceptions and limitations, e.g., research organisations, cultural heritage institutions, and others;

- types of usages, referring to rights as defined in the Berne Convention, WIPO Treaty on Copyright, EU Acquis Communautaire, MS legislations, and
- industry royalty split practices, e.g., different treatment of performing rights splits among Collective Management Organisations.

4.1.12. Commercial licensing, a fortiori automated commercial licensing, should be facilitated by the availability of machine-readable:

- usage policies, i.e., there is a need for policy engines and intelligent wizards which consider regulations and general industry practices to help generate policies;
- price structures, and
- usage contexts, i.e., there is a need for context engines and intelligent wizards which consider regulations and general industry practices to help generate contexts.

4.1.13. In the case of commercial licensing, awareness and education campaigns should be made available to authors, creators, and performers to help them know their rights and benefit from them.

4.1.14. In the case of commercial licensing, services may be set up to identify and alert authors, creators, and performers who do not seem to exploit their rights.

4.2. Traceability and transparency

4.2.1. Going forward, provenance and authenticity assertions should be either embedded in or linked to distributed DIGITAL ASSETS and remain intact throughout the distribution process.

4.2.2. Distribution systems shall enable audit trails for both internal and external governance. Audit trails shall include claim signatures and timestamps.

4.2.3. Downstream usage of licensed original content by AI algorithms should be traceable and transparent.

4.2.4. In the case of commercial licensing, digital service providers should log monetised access to content, including asset identifier and other necessary information, to facilitate the proportional and transparent remuneration of authors, creators, and performers.

4.2.5. Users shall be informed when and why their activity is logged, with reference to the digital service provider's data and rights policies.

4.2.6. Logs shall be stored securely, and their access shall be controlled.

4.2.7. RIGHTS MANAGEMENT INFORMATION such as provenance and authenticity assertions should be visible to end-users before they consume the DIGITAL ASSET.

4.2.8. Consumers should be able to view opt-out declarations and any associated right expressions.

4.2.9. RIGHTS MANAGEMENT INFORMATION shall be presented to end-users in a clear, concise, and understandable format.

5. **References**

5.1. Harmonisation of certain aspects of copyright and related rights in the information society

Source: DIRECTIVE 2001/29/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society

Article 7 – Obligations concerning rights-management information

1. Member States shall provide for adequate legal protection against any person knowingly performing without authority any of the following acts:
   a. the removal or alteration of any electronic rights-management information;
   b. the distribution, importation for distribution, broadcasting, communication or making available to the public of works or other subject-matter protected under this Directive or under Chapter III of Directive 96/9/EC from which electronic rights-management information has been removed or altered without authority,

if such person knows, or has reasonable grounds to know, that by so doing he is inducing, enabling, facilitating or concealing an infringement of any copyright or any rights related to copyright as provided by law, or of the sui generis right provided for in Chapter III of Directive 96/9/EC.

2. For the purposes of this Directive, the expression 'rights management information' means any information provided by rightsholders which identifies the work or other subject-matter referred to in this Directive or covered by the sui generis right provided for in Chapter III of Directive 96/9/EC, the author or any other rightsholder, or information about the terms and conditions of use of the work or other subject-matter, and any numbers or codes that represent such information.

The first subparagraph shall apply when any of these items of information is associated with a copy of, or appears in connection with the communication to the public of, a work or other subject matter referred to in this Directive or covered by the sui generis right provided for in Chapter III of Directive 96/9/EC.

5.2. Copyright and related rights in the Digital Single Market

Source: DIRECTIVE (EU) 2019/790 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC

Article 2 Definitions (excerpt)

For the purposes of this Directive, the following definitions apply:

1. 'research organisation' means a university, including its libraries, a research institute or any other entity, the primary goal of which is to conduct scientific research or to carry out educational activities involving also the conduct of scientific research: (a) on a not-for-profit basis or by reinvesting all the profits in its scientific research; or (b) pursuant to a public interest mission recognised by a Member State; in such a way that the access to the results generated by such scientific research cannot be enjoyed on a preferential basis by an undertaking that exercises a decisive influence upon such organisation;
2. 'text and data mining' means any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations;

3. 'cultural heritage institution' means a publicly accessible library or museum, an archive or a film or audio heritage institution.

Article 3 – Text and data mining for the purposes of scientific research

1. Member States shall provide for an exception to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, and Article 15(1) of this Directive for reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access.
2. Copies of works or other subject matter made in compliance with paragraph 1 shall be stored with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results.
3. Rightsholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.
4. Member States shall encourage rightsholders, research organisations and cultural heritage institutions to define commonly agreed best practices concerning the application of the obligation and of the measures referred to in paragraphs 2 and 3 respectively.

Article 4 – Exception or limitation for text and data mining

1. Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining.
2. Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining.
3. 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 rightsholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.

4. This Article shall not affect the application of Article 3 of this Directive.

Article 18 – Principle of appropriate and proportionate remuneration

1. Member States shall ensure that where authors and performers license or transfer their exclusive rights for the exploitation of their works or other subject matter, they are entitled to receive appropriate and proportionate remuneration.
2. In the implementation in national law of the principle set out in paragraph 1, Member States shall be free to use different mechanisms and take into account the principle of contractual freedom and a fair balance of rights and interests.

Article 19 – Transparency obligation

1. Member States shall ensure that authors and performers receive on a regular basis, at least once a year, and taking into account the specificities of each sector, up to date, relevant and comprehensive information on the exploitation of their works and performances from the parties to whom they have licensed or transferred their rights, or their successors in title, in particular as regards modes of exploitation, all revenues generated and remuneration due.
2. Member States shall ensure that, where the rights referred to in paragraph 1 have subsequently been licensed, authors and performers or their representatives shall, at their request, receive from sub-licensees additional information, in the event that their first contractual counterpart does not hold all the information that would be necessary for the purposes of paragraph.
   Where that additional information is requested, the first contractual counterpart of authors and performers shall provide information on the identity of those sub-licensees.
   Member States may provide that any request to sub-licensees pursuant to the first subparagraph is made directly or indirectly through the contractual counterpart of the author or the performer.
3. The obligation set out in paragraph 1 shall be proportionate and effective in ensuring a high level of transparency in every sector.
   Member States may provide that in duly justified cases where the administrative burden resulting from the obligation set out in paragraph 1 would become disproportionate in the light of the revenues generated by the exploitation of the work or performance,

the obligation is limited to the types and level of information that can reasonably be expected in such cases.

4. Member States may decide that the obligation set out in paragraph 1 of this Article does not apply when the contribution of the author or performer is not significant having regard to the overall work or performance, unless the author or performer demonstrates that he or she requires the information for the exercise of his or her rights under Article 20(1) and requests the information for that purpose.
5. Member States may provide that, for agreements subject to or based on collective bargaining agreements, the transparency rules of the relevant collective bargaining agreement are applicable, on condition that those rules meet the criteria provided for in paragraphs 1 to 4.
6. Where Article 18 of Directive 2014/26/EU is applicable, the obligation laid down in paragraph 1 of this Article shall not apply in respect of agreements concluded by entities defined in Article 3(a) and (b) of that Directive or by other entities subject to the national rules implementing that Directive.

Article 20 – Contract adjustment mechanism

1. Member States shall ensure that, in the absence of an applicable collective bargaining agreement providing for a mechanism comparable to that set out in this Article, authors and performers or their representatives are entitled to claim additional, appropriate and fair remuneration from the party with whom they entered into a contract for the exploitation of their rights, or from the successors in title of such party, when the remuneration originally agreed turns out to be disproportionately low compared to all the subsequent relevant revenues derived from the exploitation of the works or performances.
2. Paragraph 1 of this Article shall not apply to agreements concluded by entities defined in Article 3(a) and (b) of Directive 2014/26/EU or by other entities that are already subject to the national rules implementing that Directive.

5.3. Harmonised rules on artificial intelligence

Source: REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013,

(EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

Excerpts from Article 27

The seven principles include human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being and accountability.

Transparency means that AI systems are developed and used in a way that allows appropriate traceability and explainability, while making humans aware that they communicate or interact with an AI system, as well as duly informing deployers of the capabilities and limitations of that AI system and affected persons about their rights.

Article 50 – Transparency obligations for providers and deployers of certain AI systems

2. Providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated. Providers shall ensure their technical solutions are effective, interoperable, robust and reliable as far as this is technically feasible, taking into account the specificities and limitations of various types of content, the costs of implementation and the generally acknowledged state of the art, as may be reflected in relevant technical standards. This obligation shall not apply to the extent the AI systems perform an assistive function for standard editing or do not substantially alter the input data provided by the deployer or the semantics thereof, or where authorised by law to detect, prevent, investigate or prosecute criminal offences.
4. Deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated. This obligation shall not apply where the use is authorised by law to detect, prevent, investigate or prosecute criminal offence. Where the content forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, the transparency obligations set out in this paragraph are

limited to disclosure of the existence of such generated or manipulated content in an appropriate manner that does not hamper the display or enjoyment of the work. Deployers of an AI system that generates or manipulates text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated. This obligation shall not apply where the use is authorised by law to detect, prevent, investigate or prosecute criminal offences or where the AI-generated content has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication of the content.

5. The information referred to in paragraphs 1 to 4 shall be provided to the natural persons concerned in a clear and distinguishable manner at the latest at the time of the first interaction or exposure. The information shall conform to the applicable accessibility requirements.
6. Paragraphs 1 to 4 shall not affect the requirements and obligations set out in Chapter III, and shall be without prejudice to other transparency obligations laid down in Union or national law for deployers of AI systems.
7. The AI Office shall encourage and facilitate the drawing up of codes of practice at Union level to facilitate the effective implementation of the obligations regarding the detection and labelling of artificially generated or manipulated content. The Commission may adopt implementing acts to approve those codes of practice in accordance with the procedure laid down in Article 56 (6). If it deems the code is not adequate, the Commission may adopt an implementing act specifying common rules for the implementation of those obligations in accordance with the examination procedure laid down in Article 98(2).

Article 53 – Obligations for providers of general-purpose AI models

1. Providers of general-purpose AI models shall:
   a. draw up and keep up-to-date the technical documentation of the model, including its training and testing process and the results of its evaluation, which shall contain, at a minimum, the information set out in Annex XI for the purpose of providing it, upon request, to the AI Office and the national competent authorities;
   b. draw up, keep up-to-date and make available information and documentation to providers of AI systems who intend to integrate the general-purpose AI model into their AI systems. Without

prejudice to the need to observe and protect intellectual property rights and confidential business information or trade secrets in accordance with Union and national law, the information and documentation shall:

   i. enable providers of AI systems to have a good understanding of the capabilities and limitations of the general-purpose AI model and to comply with their obligations pursuant to this Regulation; and
   ii. contain, at a minimum, the elements set out in Annex XII;

c. put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790;
d. draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office.

MINISTRY OF EDUCATION AND CULTURE
FINLAND

P.O. Box 29 FI-00023 Government, Finland
Tel. + 358 295 16001 (switchboard)
okm.fi/en

ISSN 1799-0351 PDF
ISBN 978-952-415-143-6 PDF