

EXECUTIVE SUMMARY

Interoperable, trustworthy, and machine-readable copyright data in the AI era CITF First Project Report (2025:23)

The copyright infrastructure is generally understood as a set of rules, technologies and institutions that frame the management of copyright data in the creative sectors. In the AI era, defining the semantic layer of the copyright infrastructure – i.e., building a framework where identification and rights management information is made clear and consistent – is a crucial step to make copyright, data, and AI laws work in practise.

In November 2023, building on the work of the previous five years, the Copyright Infrastructure Task Force (CITF) was formed by four EU Member States – Estonia, Finland, Latvia, and Lithuania. Today, the CITF counts participants from more than 20 European, North-American, and other partner countries and liaises with European and international organisations. The CITF acts as a forum to promote an open rights data framework (ORDF) based on the interoperability, trustworthiness, and machine-readability of copyright data.

The first project of the CITF defined requirements in 2025 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 in 2026.

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

First, the project discussed the life cycles of 5 use cases – a doctoral thesis, a scientific article, an eBook, an image, and an article in a periodical – and analysed how they are impacted by new technical, commercial and regulatory developments, particularly around generative AI.

Then, the project identified the metadata to be collected and processed in 4 phases of the life cycles (creation, asset management, metadata management, and distribution & usage). This led to more than 75 requirements for 3 main purposes (National Libraries mandates, commercial licensing, and search across existing repositories) considering 4 clusters of metadata (agent identifiers, asset identifiers, rights management information, and interconnection metadata).

Finally, the project presented its results at an ad hoc CITF seminar held in Brussels in June 2025, published its report in September and organised a series of productive stakeholders' feedback sessions in October-December 2025. More use cases will be considered in 2026.