Council of the European Union

**Brussels, 24 November 2025**
**(OR. en)**

**15422/25**

**PI 194**
**INTER-REP 112**

**NOTE**

| | |
|---|---|
| From: | General Secretariat of the Council |
| To: | Delegations |
| Subject: | Presentation by a representative of the Copyright Infrastructure Task Force (agenda item 5) at the Working Party on Intellectual Property (Copyright) on 13 November 2025 |

Delegations will find attached a presentation (Annex I) and the accompanying speaking points (Annex II), as delivered by a representative of the Copyright Infrastructure Task Force to the Working Party on Intellectual Property (Copyright) at its meeting on 13 November 2025.

This document contains a presentation by an external stakeholder and the views expressed therein are solely those of the third party it originates from. This document cannot be regarded as stating an official position of the Council. It does not reflect the views of the Council or of its members.


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

## Report on the CITF First Project

Philippe Rixhon, Brussels, 13 November 2025

## Origin and mission of the Copyright Infrastructure Task Force

The creative industries will benefit from –

- a well-functioning copyright infrastructure
  *Stocktaking document of the Working Party on Intellectual Property, Council of the European Union, 2019*
- an open rights data framework
  *Study on Copyright and New Technologies, European Commission, 2022*

The Copyright Infrastructure Task Force is a forum identifying and promoting standards and technologies needed to raise interoperability, trustworthiness, and machine-readability of copyright data.

2/19

# The three layers of the copyright infrastructure

The space where the EU creative industries meet

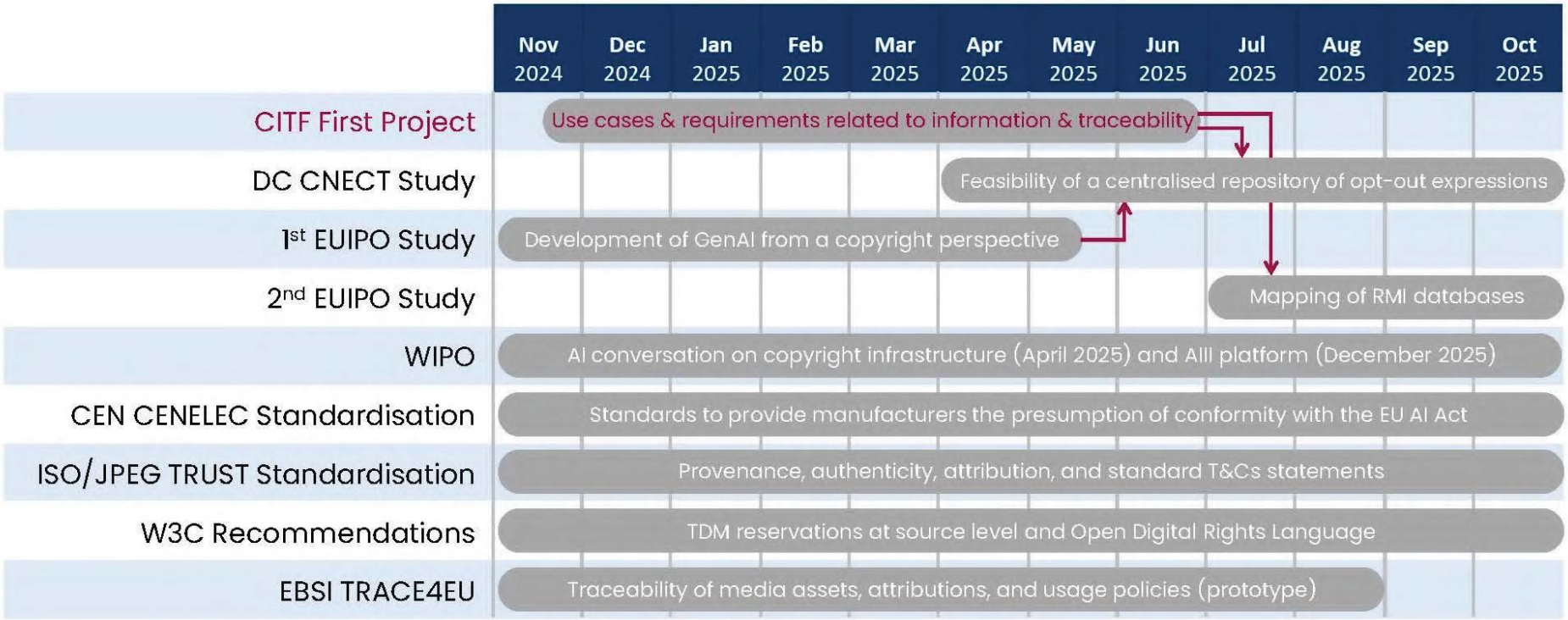
Google Meta OpenAI Spotify TikTok, etc.

COPYRIGHT INFRASTRUCTURE TASK FORCE

CENELEC ISO W3C ETSI ITU, etc.

Harmonisation of copyright in info. society EC/2001/29
Copyright in the Digital Single Market EU/2019/790
AI Act EU/2024/1689, Berne Convention, UDHR, etc.

# The context of the CITF First Project

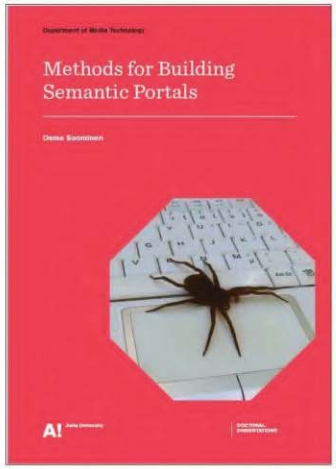



4/19

# Five partners and five use cases

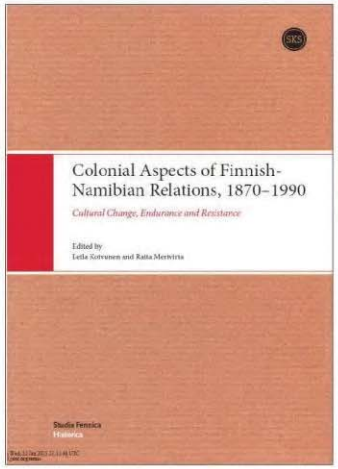
THE NATIONAL LIBRARY OF FINLAND

L N B NATIONAL LIBRARY OF LATVIA




PhD Thesis


Scientific article




eBook

Image


News article

Culture Information Systems Centre Republic of Latvia

TAL TECH valunode

5/19

# From use cases to maximum hypotheses and requirements

## BY 2030 –

- All content will be digitized and available online.
- AI applications will grow and multiply.
- Creators will be fairly, appropriately, proportionally and transparently remunerated.

## RIGHTS MANAGEMENT INFORMATION

- Standardise interoperable asset and actor identifiers
- Assert and enforce AI-related reservations

## TRACEABILITY & TRANSPARENCY

- Identify content that has been generated by AI
- Document training and generation algorithms use to produce AI-generated content

## Considerations and knowledge acquisition

**Works and media assets**

what is what, and who can tell, and the need for asset identifiers (*identification metadata*)

**Authors and other actors**

who is who, and who is accredited with work authorship or right ownership, and the need for agent identifiers (*identification metadata*)

**Attribution, terms and conditions, and remuneration**

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*) and counters (*usage metadata*)

**Provenance and authenticity**

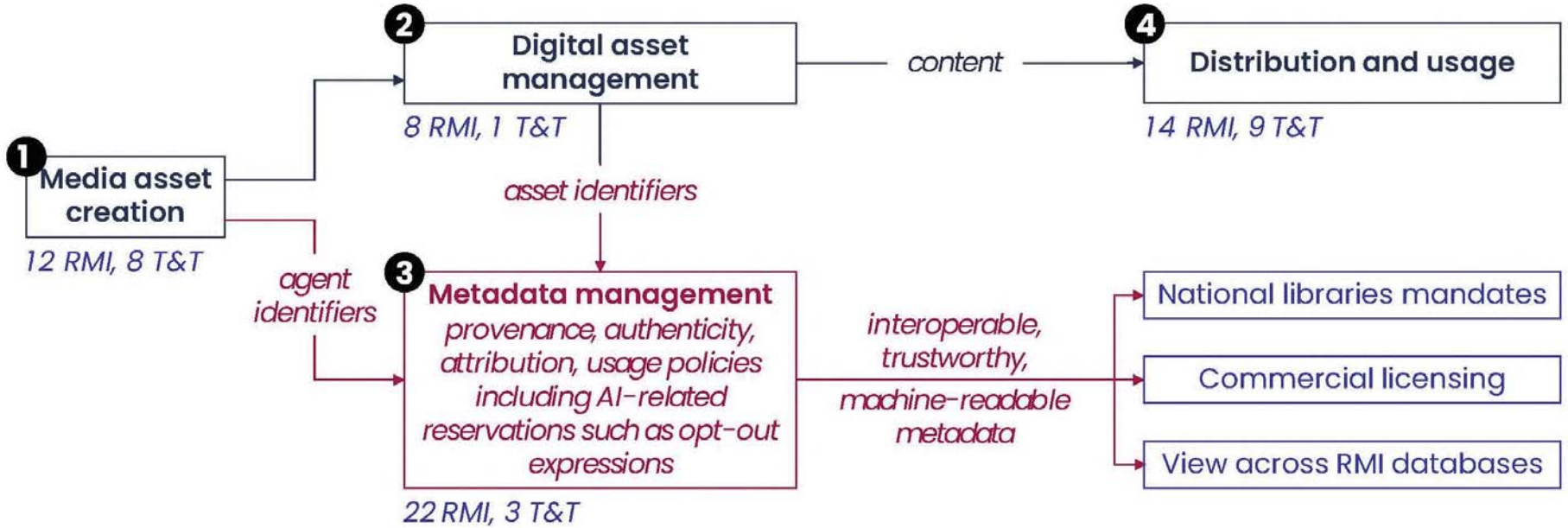
where does that media asset come from, and the need for provenance and authenticity assertions (*descriptive metadata*)

**Digital Single Market**

how do we assure the free movement of copyright data and the need for interconnection (*administrative metadata*)

7/19

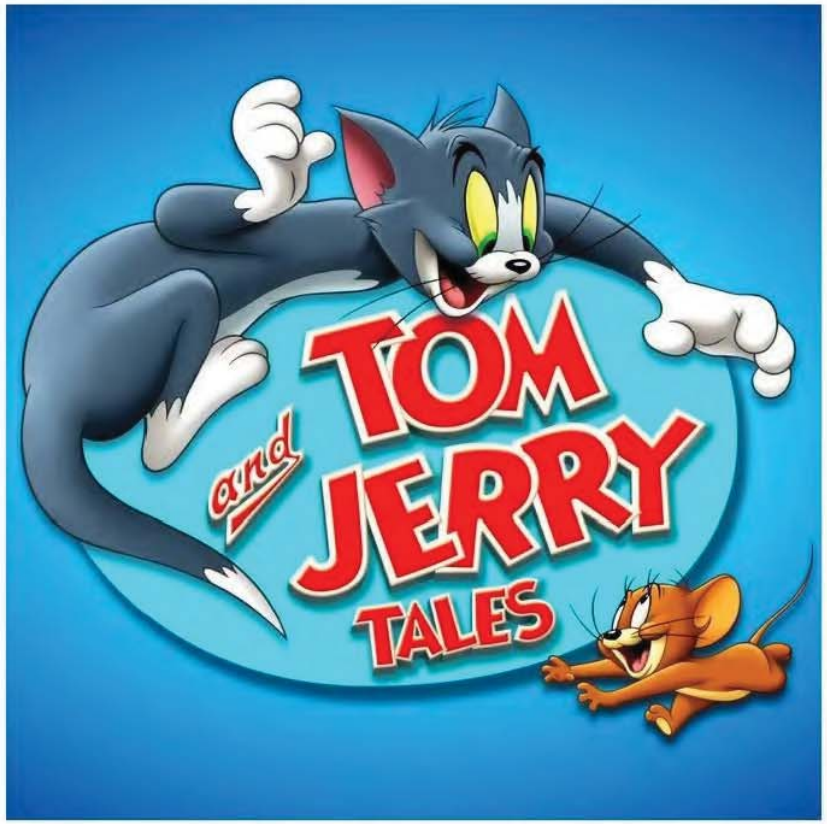
# Requirements




56 requirements related to rights management information (RMI)
21 requirements related to traceability and transparency (T&T)

8/19

# Tom and Jerry Tales



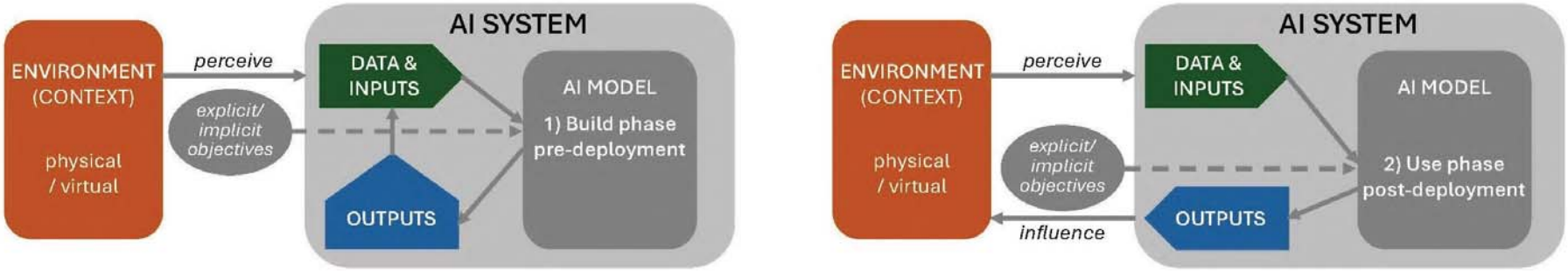

© Warner Bros. Discovery

Midjourney output

9/19

# Text and data mining

## Distinguishing the AI build phase from the AI use phase

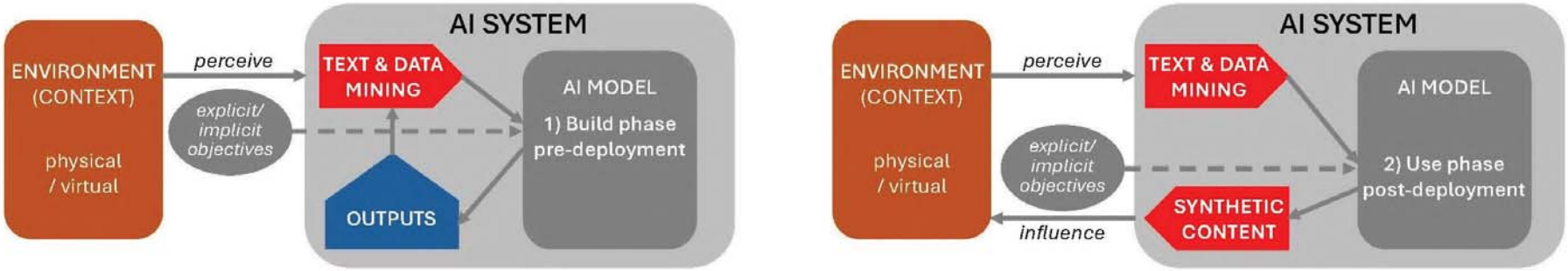



Source: OECD

10/19

# Text and data mining

## Distinguishing the AI build phase from the AI use phase

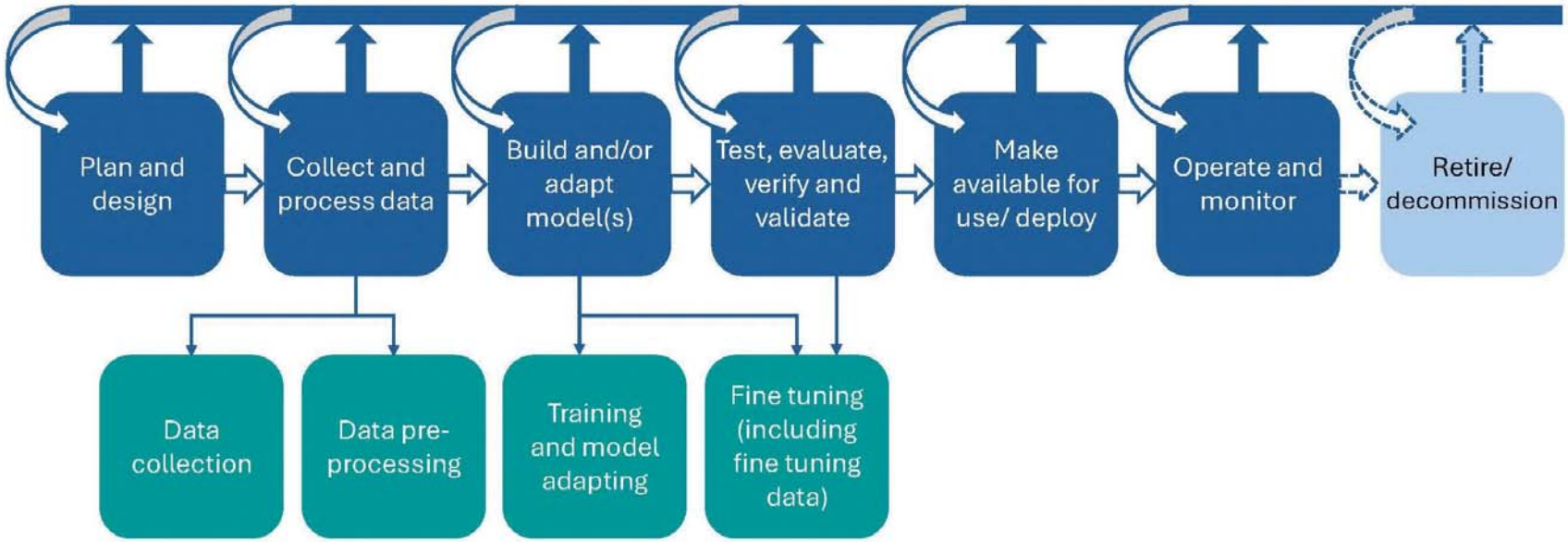



Source: OECD

10/19

# Text and data mining

## Identifying potential touch points between AI and copyright

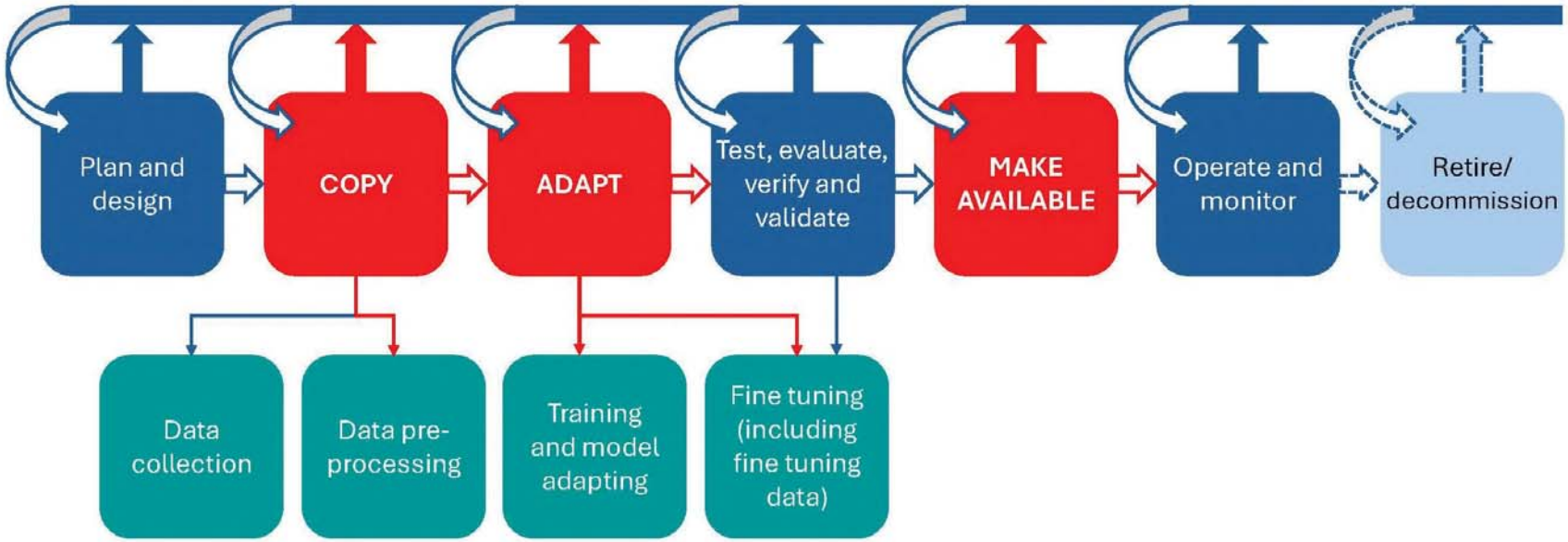



Source: OECD

11/19

# Text and data mining

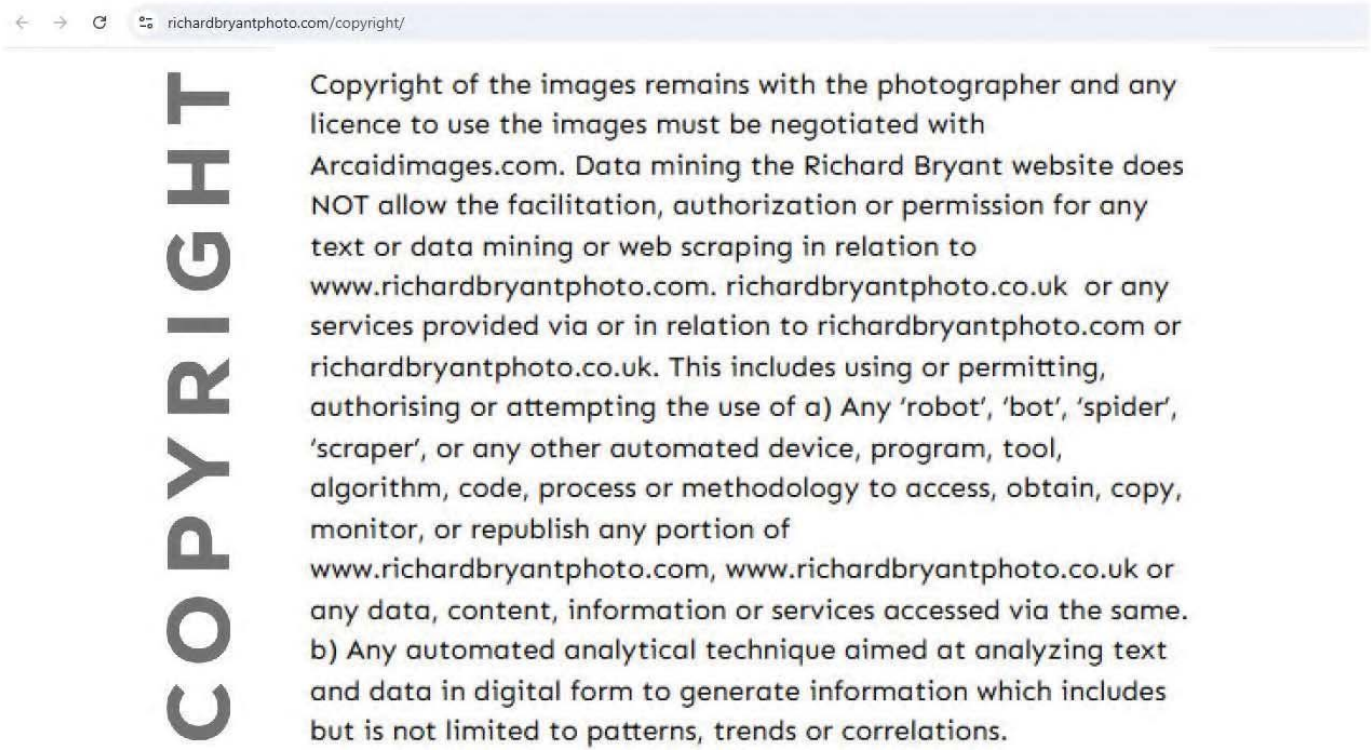
## Identifying potential touch points between AI and copyright



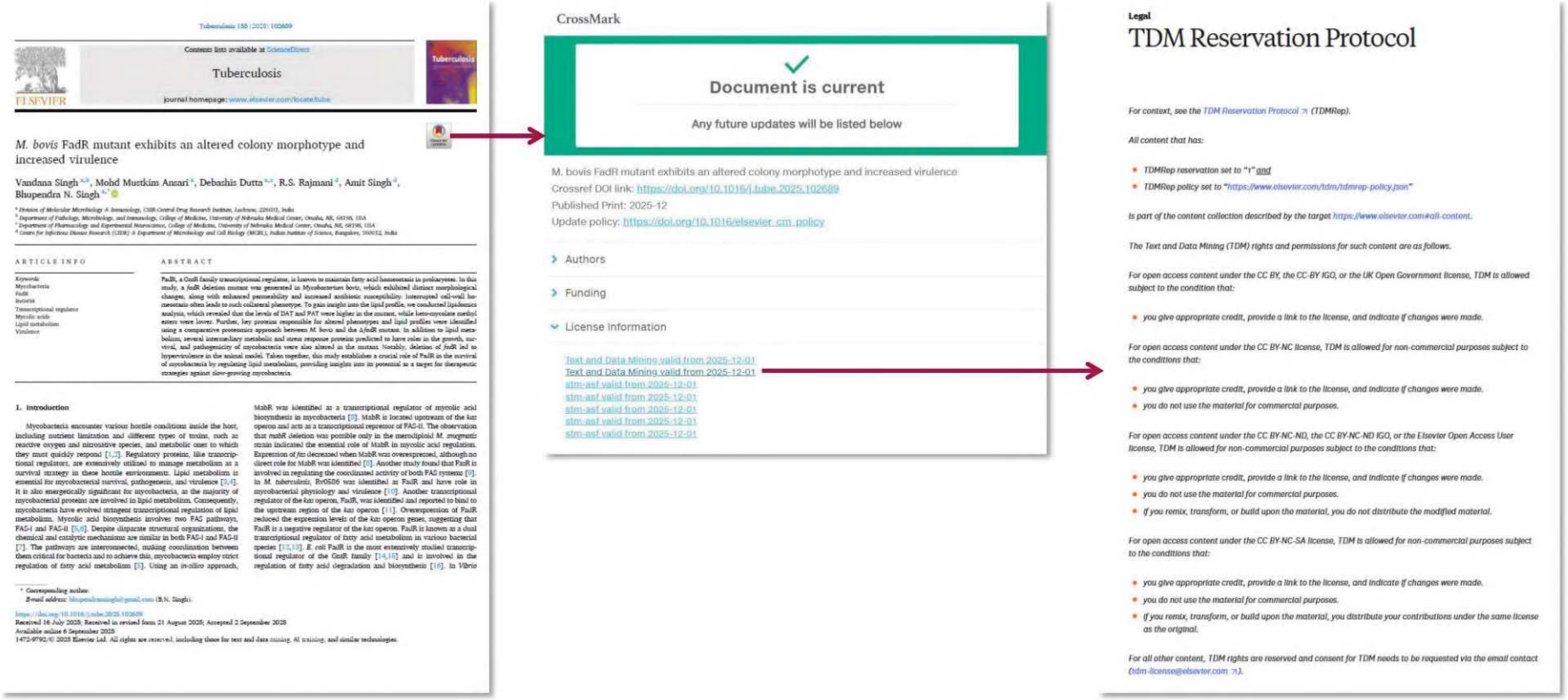

Source: OECD

11/19

# Machine-readable reservations (natural language)



# Machine-readable reservations (rights language)



Tuberculosis 158 (2025) 102689

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: www.elsevier.com/locate/tube

*M. bovis* FadR mutant exhibits an altered colony morphotype and increased virulence

Vandana Singh [a,b], Mohd Mustkim Ansari [a], Debashis Dutta [a,c], R.S. Rajmani [d], Amit Singh [d], Bhupendra N. Singh [a,*]

ARTICLE INFO

ABSTRACT

1. Introduction

CrossMark

Document is current

Any future updates will be listed below

M. bovis FadR mutant exhibits an altered colony morphotype and increased virulence

Crossref DOI link: https://doi.org/10.1016/j.tube.2025.102689

Published Print: 2025-12

Update policy: https://doi.org/10.1016/elsevier_cm_policy

Authors

Funding

License Information

Text and Data Mining valid from 2025-12-01

Text and Data Mining valid from 2025-12-01

stm-asf valid from 2025-12-01

stm-asf valid from 2025-12-01

stm-asf valid from 2025-12-01

stm-asf valid from 2025-12-01

stm-asf valid from 2025-12-01

Legal

TDM Reservation Protocol

For context, see the TDM Reservation Protocol (TDMRep).

All content that has:

- TDMRep reservation set to "1" and
- TDMRep policy set to "https://www.elsevier.com/tdm/tdmrep-policy.json"

Is part of the content collection described by the target https://www.elsevier.com#all-content.

The Text and Data Mining (TDM) rights and permissions for such content are as follows.

For open access content under the CC BY, the CC-BY IGO, or the UK Open Government license, TDM is allowed subject to the condition that:

- you give appropriate credit, provide a link to the license, and indicate if changes were made.

For open access content under the CC BY-NC license, TDM is allowed for non-commercial purposes subject to the conditions that:

- you give appropriate credit, provide a link to the license, and indicate if changes were made.
- you do not use the material for commercial purposes.

For open access content under the CC BY-NC-ND, the CC BY-NC-ND IGO, or the Elsevier Open Access User license, TDM is allowed for non-commercial purposes subject to the conditions that:

- you give appropriate credit, provide a link to the license, and indicate if changes were made.
- you do not use the material for commercial purposes.
- if you remix, transform, or build upon the material, you do not distribute the modified material.

For open access content under the CC BY-NC-SA license, TDM is allowed for non-commercial purposes subject to the conditions that:

- you give appropriate credit, provide a link to the license, and indicate if changes were made.
- you do not use the material for commercial purposes.
- if you remix, transform, or build upon the material, you distribute your contributions under the same license as the original.

For all other content, TDM rights are reserved and consent for TDM needs to be requested via the email contact (tdm-license@elsevier.com).

With the kind permission of Elsevier

Under the hood



```
<?xpacket begin="" id="W5M0MpCehiHzreSzNTczkc9d"?>
<x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmptk="Adobe XMP Core 9.1-c001 79.675d0f7,
2023/06/11-19:21:16        ">
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
        <rdf:Description rdf:about="" xmlns:crossmark="http://crossref.org/crossmark/1.0/"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:jav="http://www.niso.org/schemas/jav/1.0/"
xmlns:pdf="http://ns.adobe.com/pdf/1.3/" xmlns:pdfx="http://ns.adobe.com/pdfx/1.3/"
xmlns:prism="http://prismstandard.org/namespaces/basic/3.0/"
xmlns:tdm="http://www.w3.org/ns/tdmrep/" xmlns:xmp="http://ns.adobe.com/xap/1.0/"
xmlns:xmpMM="http://ns.adobe.com/xap/1.0/mm/"
xmlns:xmpRights="http://ns.adobe.com/xap/1.0/rights/">
            <crossmark:CrossmarkDomainExclusive>true</crossmark:CrossmarkDomainExclusive>
            <crossmark:DOI>10.1016/j.tube.2025.102689</crossmark:DOI>
            <crossmark:MajorVersionDate>2010-04-23</crossmark:MajorVersionDate>
            <crossmark:CrossMarkDomains>
                <rdf:Seq>
                    <rdf:li>elsevier.com</rdf:li>
                    <rdf:li>sciencedirect.com</rdf:li>
                </rdf:Seq>
            </crossmark:CrossMarkDomains>
            <dc:format>application/pdf</dc:format>
            <dc:identifier>10.1016/j.tube.2025.102689</dc:identifier>
            <dc:publisher>
                <rdf:Bag>
                    <rdf:li>Elsevier Ltd</rdf:li>
                </rdf:Bag>
            </dc:publisher>
            <dc:description>
                <rdf:Alt>
                    <rdf:li xml:lang="x-default">Tuberculosis, 155 (2025) 102689.
doi:10.1016/j.tube.2025.102689</rdf:li>
                </rdf:Alt>
            </dc:description>
            <dc:title>
                <rdf:Alt>
                    <rdf:li xml:lang="x-default">M. bovis FadR mutant exhibits an altered colony
morphotype and increased virulence</rdf:li>
                </rdf:Alt>
```

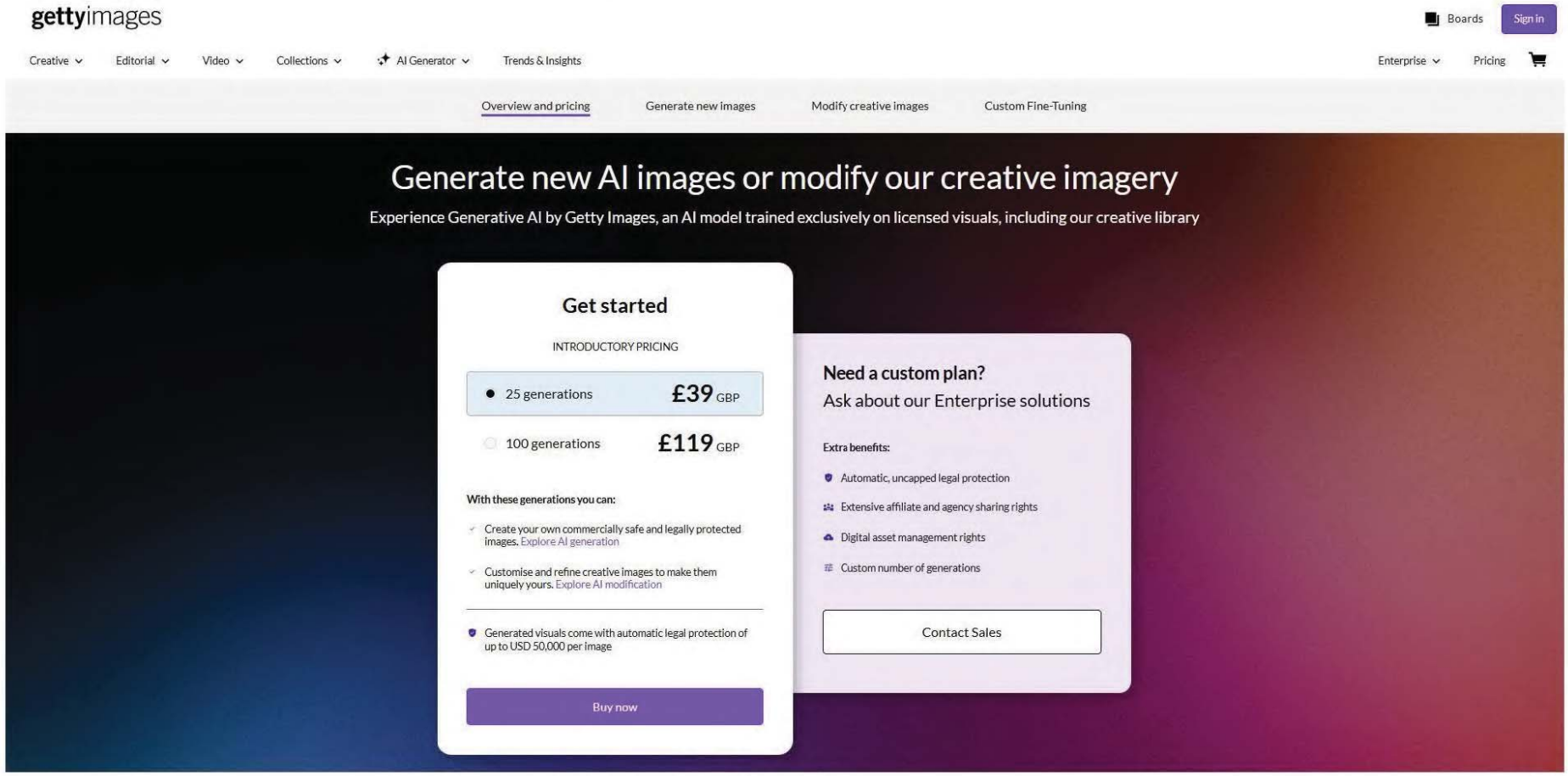
With the kind permission of Elsevier

```
                </rdf:Alt>
            </dc:title>
            <dc:creator>
                <rdf:Seq>
                    <rdf:li>Vandana Singh</rdf:li>
                    <rdf:li>Mohd Mustkim Ansari</rdf:li>
                    <rdf:li>Debashis Dutta</rdf:li>
                    <rdf:li>R.S. Rajmani</rdf:li>
                    <rdf:li>Amit Singh</rdf:li>
                    <rdf:li>Bhupendra N. Singh</rdf:li>
                </rdf:Seq>
            </dc:creator>
            <dc:subject>
                <rdf:Bag>
                    <rdf:li>Mycobacteria</rdf:li>
                    <rdf:li>FadR</rdf:li>
                    <rdf:li>Rv0494</rdf:li>
                    <rdf:li>Transcriptional regulator</rdf:li>
                    <rdf:li>Mycolic acids</rdf:li>
                    <rdf:li>Lipid metabolism</rdf:li>
                    <rdf:li>Virulence</rdf:li>
                </rdf:Bag>
            </dc:subject>
            <jav:journal_article_version>VoR</jav:journal_article_version>
            <pdf:CreationDate--Text>11th September 2025</pdf:CreationDate--Text>
            <pdf:Producer>Acrobat Distiller 8.1.0 (Windows)</pdf:Producer>
            <pdf:Keywords>Mycobacteria,FadR,Rv0494,Transcriptional regulator,Mycolic acids,Lipid
metabolism,Virulence</pdf:Keywords>
            <pdfx:CreationDate--Text>11th September 2025</pdfx:CreationDate--Text>
            <pdfx:CrossmarkDomainExclusive>true</pdfx:CrossmarkDomainExclusive>
            <pdfx:CrossmarkMajorVersionDate>2010-04-23</pdfx:CrossmarkMajorVersionDate>
            <pdfx:doi>10.1016/j.tube.2025.102689</pdfx:doi>
            <pdfx:robots>noindex</pdfx:robots>
```

Under the hood

```
            <pdfx:CrossMarkDomains>
                <rdf:Seq>
                    <rdf:li>sciencedirect.com</rdf:li>
                    <rdf:li>elsevier.com</rdf:li>
                </rdf:Seq>
            </pdfx:CrossMarkDomains>
            <prism:aggregationType>journal</prism:aggregationType>
            <prism:copyright>© 2025 Elsevier Ltd. All rights are reserved, including those for text
and data mining, AI training, and similar technologies.</prism:copyright>
            <prism:coverDate>2025-12-01</prism:coverDate>
            <prism:coverDisplayDate>1 December 2025</prism:coverDisplayDate>
            <prism:doi>10.1016/j.tube.2025.102689</prism:doi>
            <prism:issn>1472-9792</prism:issn>
            <prism:pageRange>102689</prism:pageRange>
            <prism:publicationName>Tuberculosis</prism:publicationName>
            <prism:startingPage>102689</prism:startingPage>
            <prism:url>https://doi.org/10.1016/j.tube.2025.102689</prism:url>
            <prism:volume>155</prism:volume>
            <tdm:policy>https://www.elsevier.com/tdm/tdmrep-policy.json</tdm:policy>
            <tdm:reservation>1</tdm:reservation>
            <xmp:CreateDate>2025-09-11T12:11:12</xmp:CreateDate>
            <xmp:CreatorTool>Elsevier</xmp:CreatorTool>
            <xmp:MetadataDate>2025-09-11T12:20:33</xmp:MetadataDate>
            <xmp:ModifyDate>2025-09-11T12:20:33</xmp:ModifyDate>
            <xmpMM:DocumentID>uuid:1b499bed-4ce8-4c0c-b682-0d58baae1cbe</xmpMM:DocumentID>
            <xmpMM:InstanceID>uuid:b6ed5266-03cb-4855-90c0-636283357f42</xmpMM:InstanceID>
            <xmpRights:Marked>True</xmpRights:Marked>
        </rdf:Description>
    </rdf:RDF>
</x:xmpmeta>
<?xpacket end="w"?>
```

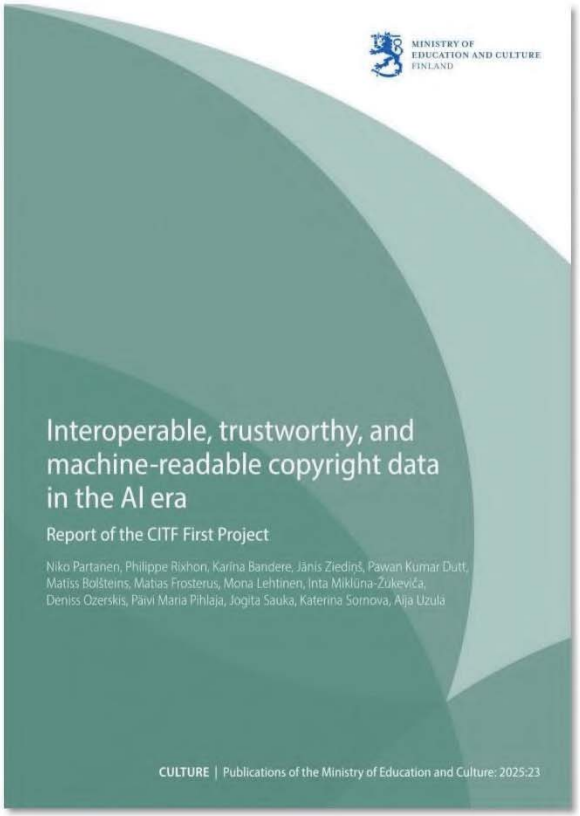
# Opt-in opportunities




© Getty Images

15/19

# Stakeholders' outreach and feedback




## Dissemination

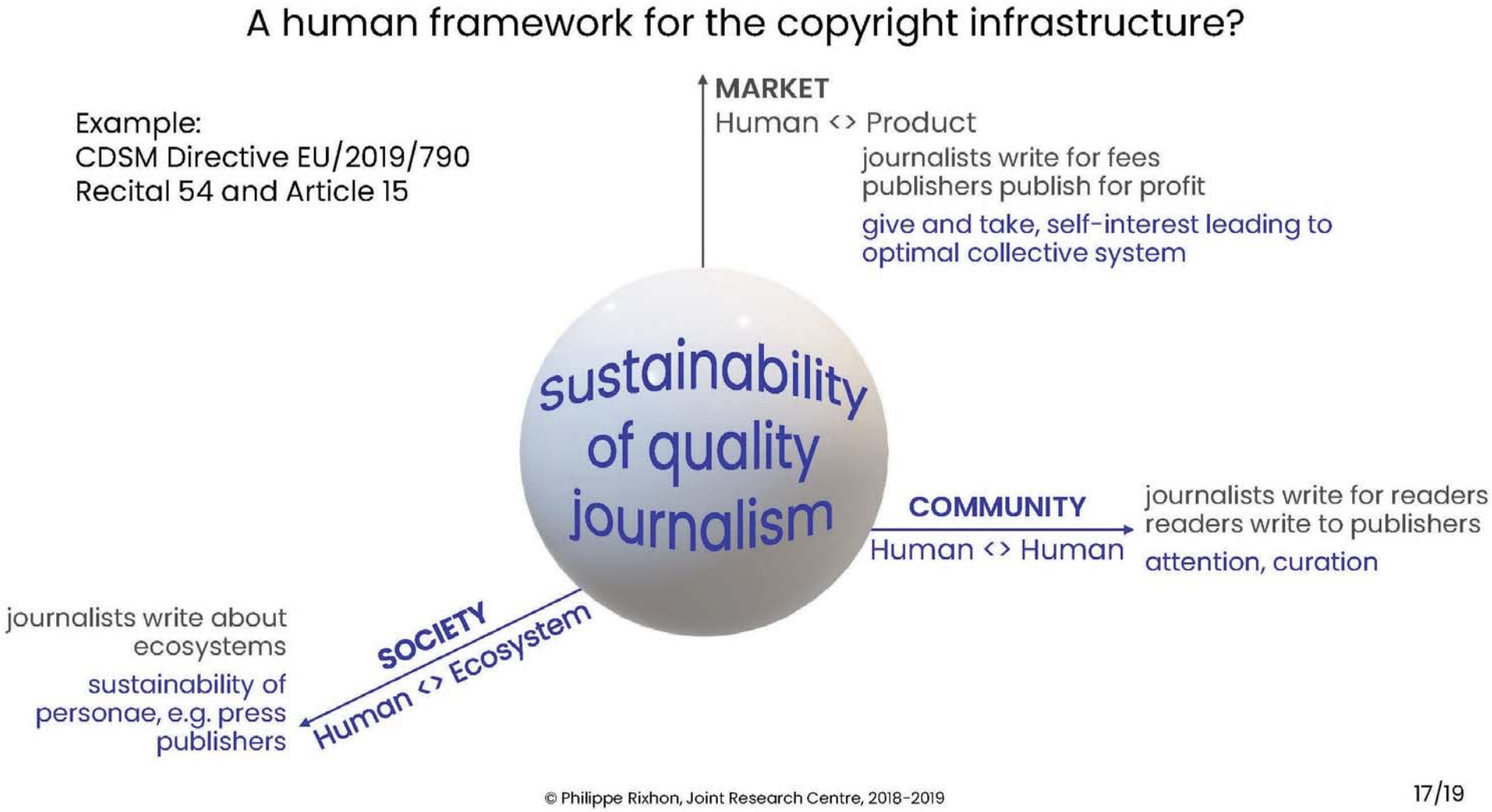
- Monthly meetings and website
- Seminar on 16 June 2025, report, and video

## Feedback on 28 October 2025

- Creators' control on their works
- What about DIY authors and performers?
- CITF Knowledge Centre?
- Author's label "Made by a human"
- CITF next projects?
  - Deduplication
  - RMI robustness
  - eIDAS, EUDI, and copyright
  - Personality rights
  - Contextual minimum sets of metadata

Next feedback sessions: 19 November in Alicante and 2 December 2025 in Geneva

16/19



# Call to action

The CITF transitioned from a Finno-Estonian exploration to a European function while becoming –

- the forum of Member States focused on the semantic layer of copyright and related rights infrastructure
- a creator and disseminator of knowledge

The CITF needs a growth platform to pursue its mission –

- in an open, iterative, and interdisciplinary way
- as a steward of the human condition in the AI era

18/19


## Resources

- Press release: here
- Press: here
- Report: here
- Video: here
- Website: here

## Contacts

- Anna Vuopala, Finnish Ministry of Education & Culture
  Email: anna.vuopala@gov.fi
- Philippe Rixhon, Valunode OÜ
  Email: prixhon@valunode.com

19/19


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

Report on the CITF First Project to the Working Party on Intellectual Property (Copyright)
at the Council of the European Union

### Origin and mission of the Copyright Infrastructure Task Force

Six years ago, under the Finnish presidency of the European Union, the Working Party on Intellectual Property (Copyright) issued a stocktaking document on a well-functioning copyright infrastructure. The document is ageing well – it is keeping its relevance.

Following the stocktaking document and the Directive on Copyright in the Digital Single Market, the Commission launched a study on Copyright and New Technologies with two objectives: checking how new technologies could be used to support copyright management, and exploring the relationship between copyright and artificial intelligence.

In that study, I argued that if one really wants a Digital Single Market, one must open the rights data framework. Rights data shall travel across creative sectors, Member States, and distribution channels. But, free movement of data has a price. It supposes that this data is interoperable, trustworthy, and machine-readable.

Three years ago, the governments of Estonia and Finland established a task force to look into that matter. The Copyright Infrastructure Task Force (CITF) is a forum identifying and promoting standards and technologies needed to raise interoperability, trustworthiness, and machine-readability of copyright data. My esteemed colleague, Anna Vuopala, kept you informed on our progress.

### The three layers of the copyright infrastructure

The definitions of "infrastructure" abound and differ. This is why it is helpful to distinguish the three layers of a copyright infrastructure –

- The foundational legal layer that defines author's rights by conventions, treaties, directives, acts, and regulations,
- The technical layer that provides mechanisms for trust and interoperation of identification and rights management information, and
- The middle layer – the semantic layer – that addresses the challenges of declaration and intermediation where the meaning of identification and rights management information is made clear and consistent; the layer that enables engineers to instruct machines in a consistent way on the technical layer.

Report on the CITF First Project – Philippe Rixhon – 13 November 2025 1

Practically, the legal layer includes the dynamic acquis Communautaire and its references – the relevant international treaties. The technical layer is the space where millions of European creators and one and a half million of European SMEs meet very large platforms.

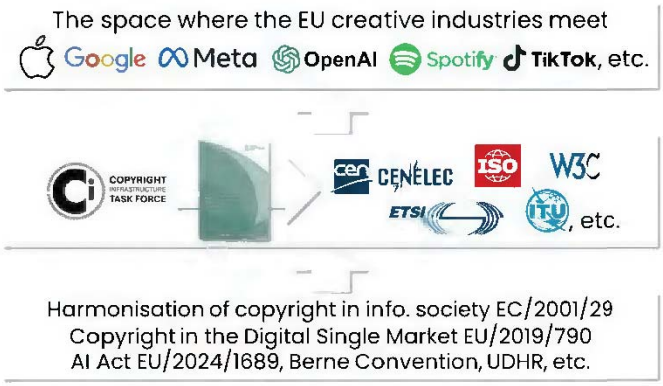



Figure 1: The three layers of the copyright infrastructure

In the middle, we primarily have standardisation bodies. They follow a straightforward methodology. Step 1: define use cases; if there is no practical use case, there is no need to standardise. Step 2: develop standardisation requirements to address the use cases. Step 3: conduct the standardisation process based on consensus. And step 4: provide reference software to help the technical layer implement the new standards. The CITF – a standardisation forum – focuses on steps 1 and 2 – use cases and requirements – in liaison with standards developing organisations, and in liaison with the actors on the legal and technical layers. The CITF plays – at least graphically – a central role in the development of the copyright infrastructure – *conditio sine qua non* for the emergence of a Digital Single Market.

## The context of the CITF first project

Two years ago, a discussion with the Commission led to the formulation of the CITF First Project: addressing the AI and Copyright conundrum. Accepting that challenge led us to look at the copyright infrastructure in general. The First Project ran from November 2024 to June 2025. Not in a void, but in connection with a series of parallel initiatives:

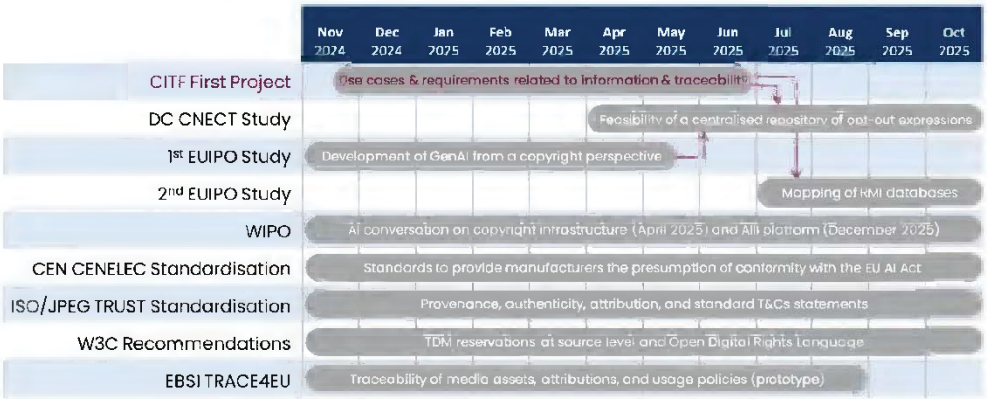



Figure 2: The context of the CITF First Project

The DG CNECT study on the feasibility of a repository of opt-out expressions, whereby the terms of reference of that study asked the researchers to look at the CITF work. The EUIPO study on GenAI from a copyright perspective that the CITF used for its First Project. The EUIPO mapping of copyright databases and metadata that is informed by the CITF First Project. The WIPO activities which are taken into account by the CITF and in turn offer the CITF a platform for dissemination. The standards developing organisations, CEN CENELEC, ISO, and W3C with which the CITF collaborates. And the European Blockchain Service Infrastructure (EBSI), which participates in the CITF meetings and developed a first prototype to track and trace media assets, attributions, and usage policies. Now, there is a robust, coordinated ecosystem to realise the stocktaking document of the Working Party.

## Five partners and five use cases

The CITF First Project was managed by the National Library of Finland. They led a consortium including the National Library of Latvia, the Culture Information Systems Centre of Latvia, which built a first copyright infrastructure, the Department of Law at the Tallinn University of Technology, and Valunode also based in Tallinn. Librarians have the longest and most sophisticated practice of taxonomy – how one names things – and ontology – how one relates things. Their knowledge is essential to build an effective semantic layer.


PhD Thesis | Scientific article | eBook | Image | News article

Figure 3: The 5 use cases of the CITF First Project

The National Library of Finland choose three items, a PhD thesis, a scientific article, and an eBook. The National Library of Latvia picked two items, an image and a news article. Then, they collected a maximum of metadata around the items, including all rights data. Finally, they described the full life cycle of each of the items and how these life cycles are currently supported by information and communication technologies.

## From use cases to maximum hypotheses and requirements

The project team formulated maximum hypotheses to be on the safe side of the AI and copyright conundrum. We said that by 2030 all content will be digitised and available online, and that AI applications will grow and multiply. Even if both statements will not be confirmed, it is prudent to consider the most challenging theoretical situation. And, we said that in that situation, creators will be paid fairly, appropriately, proportionally, and transparently. Fairly and appropriately are socio-political commercial considerations. Proportionally and transparently are first and foremost technical requirements. We are not equipped to meet them. The necessary data is barely available.

Two clusters of requirements have been developed from the acquis Communautaire – from the well-established directives on copyright and from the emerging legislation and guidelines around artificial intelligence –

- We need to know who did what – that is the core question for the management of moral and commercial authors' rights. It means that we need to know who is who and what is what. Therefore, we need to standardise interoperable asset and actor identifiers. It may sound trivial. But two years ago, for example, the photography sector was not considering the question. Then, we need to assert and enforce AI-related reservations. Indeed, we generally need to know what we may do with what. That is the domain of usage policies. An opt-out expression is nothing else than a usage policy. And, in the 21st century, policies better be machine-readable.
- The second cluster covers the hotly debated traceability and transparency requirements. We need to identify content that has been generated by AI. We also need to document training and generation algorithms used to produce AI-generated content. On the semantic layer, we do not to enter that debate but specify systems which will support the implementation and enforcement of the law and judgements, whatever they are.

## Considerations and knowledge acquisition

The CITF First Project considered five topics –

- Works and media assets – what is what, and who can tell, and the need for asset identifiers (*identification metadata*)
- Authors and other actors – who is who, and who is accredited with work authorship or right ownership, and the need for agent identifiers (*identification metadata*)
- Attribution, terms and conditions, and remuneration – 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*) and meters (*usage metadata*)
- Provenance and authenticity – where does that media asset come from, and the need for provenance and authenticity assertions (*descriptive metadata*)
- The Digital Single Market – how do we assure the free movement of copyright data and the need for interconnection (*administrative metadata*)

The project team considered the five types of metadata defined in the study Copyright and New Technologies. *Identification metadata* to identify things and people. *Rights metadata* to assert attributions, terms, and conditions. *Usage metadata* for a transparent trade of content. *Descriptive metadata* for search and enjoyment of content. And *administrative metadata* to trust and exchange all other metadata.

To equip the team, we organised four knowledge transfer workshops led by experts from organisations liaising with the CITF. The team members became acquainted with the latest developments around asset and agent identifiers, the expression of assertions, and technical interconnection.

So, this was a team of practitioners, sharing their daily professional lives and own properties, and gaining expertise. Motivated experts, who all gave their full attention to the project from its inception to its final report. They deserve praise and gratitude.

## Requirements

Together, we formulated 77 requirements –

56 requirements related to rights management information (RMI), and
21 requirements related to traceability and transparency (T&T).

We could of course have merged or split some of them. We looked at the four generic phases of the life cycles of the five items selected by the national libraries: (1) media asset creation, (2) digital asset management – the management of the digital media files, (3) metadata management, and (4) distribution and usage.

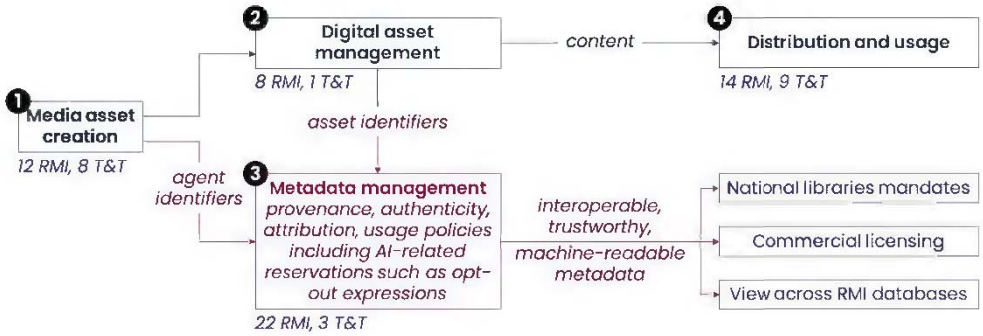
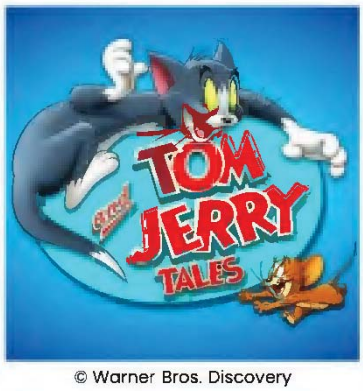



Figure 4: Requirements

We considered three purposes: the mandates of national libraries, commercial licensing, and the possibility to view rights management information across existing databases. We considered the two clusters of requirements for the four phases of life cycles for the three purposes, looking at asset and agent identifiers, provenance, authenticity, attribution and usage policies, and guided by our mantra that the Digital Single Market requires free movements of interoperable, trustworthy, and machine-readable metadata.

## Tom and Jerry Tales

We are in the middle of the codes, guidelines, and judgements season. Some serious stuff, and some tales. You remember Tom and Jerry. Please keep in mind the picture on the left.




© Warner Bros. Discovery

Midjourney output

Figure 5: Tom and Jerry

Now, please look to the right. You do not know who these cat and mouse are. Or you do. Or you recognise them if you put on special glasses. Or the right is only an adaptation of the left. Or it is not. Or you do not know. And if you are my public for an hour, I communicate these pictures to you, or I don't. And if I do, we can consider this as scientific research, or we can't. I have just established that this is much more complex that it seems... and confirmed that the truth is in the eyes of the believers. At the semantic layer, we do not believe, we think clinically. And we don't deal with truth, but with trust and security coefficients. Recently, several teams started to develop trust frameworks – frameworks relying on factual trust indicators, not on faith.

We have observed in the last weeks that creative industries are resilient, inventive, and able to compromise. Maybe, you cannot protect the cartoon on the left, but you can protect the characters of Tom and Jerry, and you could protect the style of Warner Bros. More work will be needed on the middle layer.

## Distinguishing the AI build phase from the AI use phase

Seeing some oversimplistic taxonomies of AI actions, the team of the CITF First Project looked at the Commission Guidelines on the definition of an artificial system established by Regulation EU 2024/1689 (AI Act). The definition refers to OECD studies and comprises seven main elements: 1 a machine-based system, 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, 5 infers from the input it receives, how to generate outputs, 6 such as predictions, content, recommendations, or decisions, 7 that can influence physical or virtual environments.

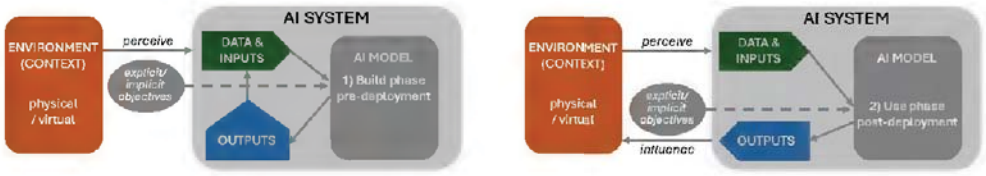



Figure 6a: The two phases of an AI system (Source: OECD)

The definition of an AI system adopts a lifecycle-based perspective encompassing two main phases: on the left, the pre-deployment or "build" phase of the system, and, on the right, the post-deployment or "use" phase of the system.

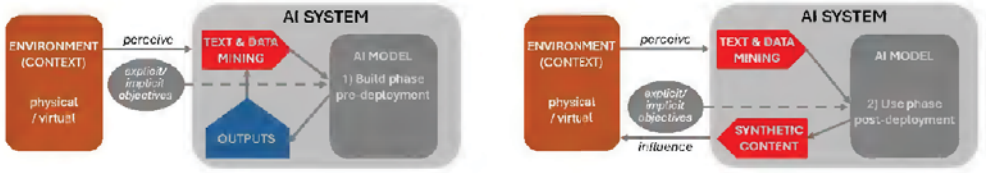



Figure 6b: The two phases of an AI system (Source: OECD)

The point for the copyright infrastructure is that the text and data mining on the left is not the same as the text and data mining on the right – the one for RAG – Retrieval Augmented Generation. If a taxonomy includes only one text and data mining, it is probably not very helpful.

## Identifying potential touch points between AI and copyright

Now, let us have a closer look at the text and data mining during the build phase. The OECD describes it as an iterative multi-step process.

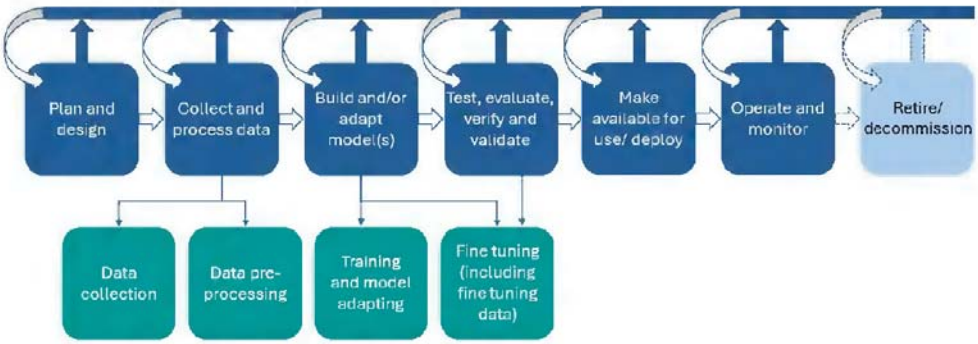



Figure 7a: Text and data mining during the build phase (Source: OECD)

It has to collect and process data. So, it must access and/or store a copy of this data and process it somewhere. So, there is a usage and this usage produces a derivative, however one calls that derivative. And if this derivative is not made available, then it is probably useless. So, the derivative of a copy is probably made available.

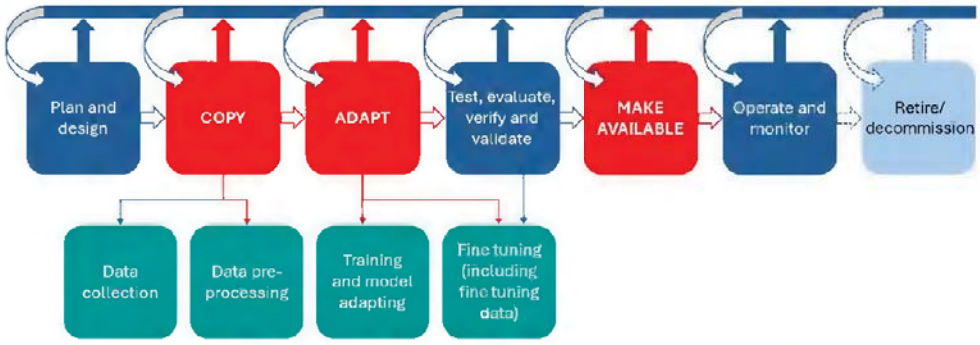



Figure 7b: Text and data mining during the build phase (Source: OECD)

This is where the team of the CITF First Project stopped. We identified potential touch points between AI and copyright. It was not up to us to decide if permissions are needed, or if remunerations are due to proceed with the build phase of an AI system. It was up to us to identify the potential anchor points for traceability, transparency, and reservations.

You may be left wanting more. In this case, have a look at the guide of the stages of AI model development that could involve copyright-relevant acts published last month by Eleonora Rosati. Her simplified guide confirmed that the CITF team was on the right track. Professor Rosati is a jurist, expert in both copyright management and enforcement. She knows more.

## Machine-readable reservations (natural language)

A lot of things have been said and written about the machine-readability of AI-related reservations, commonly reduced to their negative formulation "opt-out".

> Copyright of the images remains with the photographer and any licence to use the images must be negotiated with Arcaidimages.com. Data mining the Richard Bryant website does NOT allow the facilitation, authorization or permission for any text or data mining or web scraping in relation to www.richardbryantphoto.com. richardbryantphoto.co.uk or any services provided via or in relation to richardbryantphoto.com or richardbryantphoto.co.uk. This includes using or permitting, authorising or attempting the use of a) Any 'robot', 'bot', 'spider', 'scraper', or any other automated device, program, tool, algorithm, code, process or methodology to access, obtain, copy, monitor, or republish any portion of www.richardbryantphoto.com, www.richardbryantphoto.co.uk or any data, content, information or services accessed via the same. b) Any automated analytical technique aimed at analyzing text and data in digital form to generate information which includes but is not limited to patterns, trends or correlations.

Figure 8: An opt-out reservation (Source: www.richardbryantphoto.com)

The opt-out expression of Figure 8 is expressed in natural language, and it is machine-readable. It is possible to express this very reservation in sophisticated rights languages such as W3C's Open Digital Rights Language (ODRL). It would require natural language processing and/or IT wizards to keep the richness of what this photographer wants to convey, but it is possible.

## Machine-readable reservations (rights language)

Great progress is made. Have a look at Figure 9a. This is a scientific article published by Elsevier.

On the top right, you see the CrossMark button.

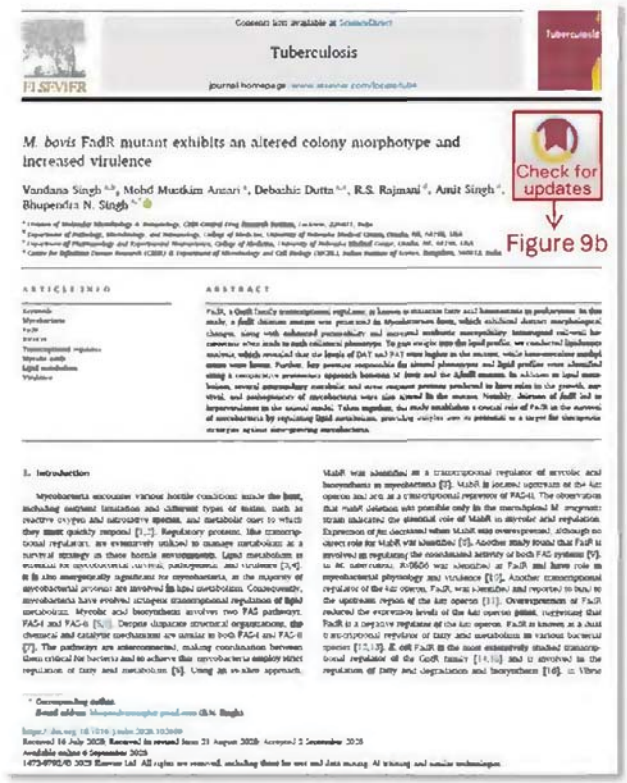



Figure 9a: An opt-out declaration, the CrossMark button (Source: Science Direct, Elsevier)

When you click on that button, a window appears with a series of pop-up menus. One of them concerns license information. The license information includes several links:

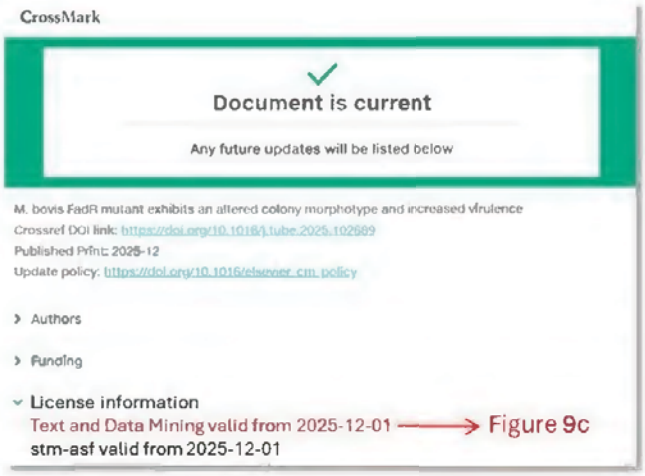
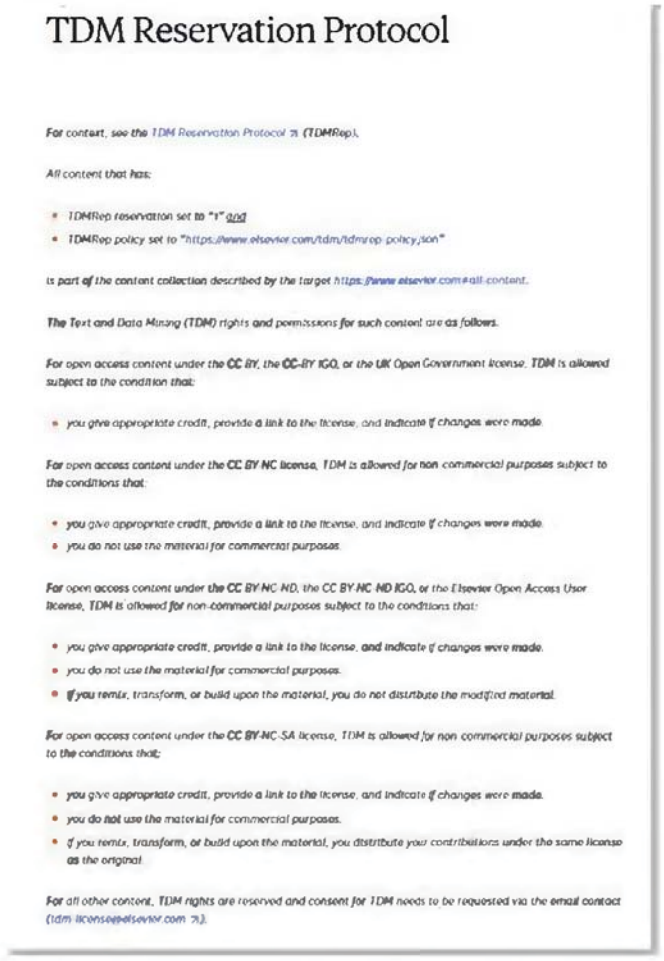



Figure 9b: The CrossMark window (Source: Science Direct, Elsevier)

When you click on the Text and Data Mining link, you access the Elsevier's TDM Reservation Protocol in natural language.



TDM Reservation Protocol

For context, see the TDM Reservation Protocol ↗ (TDMRep).

All content that has:

- TDMRep reservation set to "1" and
- TDMRep policy set to "https://www.elsevier.com/tdm/tdmrep-policy.json"

is part of the content collection described by the target https://www.elsevier.com/#all-content.

The Text and Data Mining (TDM) rights and permissions for such content are as follows.

For open access content under the CC BY, the CC-BY IGO, or the UK Open Government license, TDM is allowed subject to the condition that:

- you give appropriate credit, provide a link to the license, and indicate if changes were made.

For open access content under the CC BY-NC license, TDM is allowed for non-commercial purposes subject to the conditions that:

- you give appropriate credit, provide a link to the license, and indicate if changes were made.
- you do not use the material for commercial purposes.

For open access content under the CC BY-NC-ND, the CC BY-NC-ND IGO, or the Elsevier Open Access User license, TDM is allowed for non-commercial purposes subject to the conditions that:

- you give appropriate credit, provide a link to the license, and indicate if changes were made.
- you do not use the material for commercial purposes.
- if you remix, transform, or build upon the material, you do not distribute the modified material.

For open access content under the CC BY-NC-SA license, TDM is allowed for non-commercial purposes subject to the conditions that:

- you give appropriate credit, provide a link to the license, and indicate if changes were made.
- you do not use the material for commercial purposes.
- if you remix, transform, or build upon the material, you distribute your contributions under the same license as the original.

For all other content, TDM rights are reserved and consent for TDM needs to be requested via the email contact (tdm.licensing@elsevier.com ↗).

Figure 9c: TDM Reservation Protocol (Source: Science Direct, Elsevier)

This is precisely what the CITF is looking at. Identifying the article, identifying the authors, attributing the article to the authors, enabling the expression of usage policies, requiring the robustness of the link between the article and the CrossMark, and the robustness of the link between the CrossMark and Elsevier's TDM Reservation Protocol. Also requiring that this protocol shall be human-readable and machine-readable.

And making a distinction between the mandatory requirements "shall", the recommendations "should", and the options "may". Yes, shall, should and may have standardised meanings.

Please fasten your seat belts. We now go under the hood, as we did during the CITF First Project. When we did that, we discovered for example that sometimes a work appears in two databases for operational reasons, and that the rights assertions for the same work might differ from one database to the next due to some technical reasons. You saw the link between the foundational legal layer and the CITF. Now, I show you the link between the CITF and the technical layer.

Figure 10 supports the mention of a few characteristics.

```
1   <?xpacket begin="" id="W5M0MpCehiHzreSzNTczkc9d"?>
2   <x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmptk="Adobe XMP Core 9.1-c001 79.675d0f7,
    2023/06/11-19:21:16        ">
3       <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
4           <rdf:Description rdf:about="" xmlns:crossmark="http://crossref.org/crossmark/1.0/"
    xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:jav="http://www.niso.org/schemas/jav/1.0/"
    xmlns:pdf="http://ns.adobe.com/pdf/1.3/" xmlns:pdfx="http://ns.adobe.com/pdfx/1.3/"
    xmlns:prism="http://prismstandard.org/namespaces/basic/3.0/"
    xmlns:tdm="http://www.w3.org/ns/tdmrep/" xmlns:xmp="http://ns.adobe.com/xap/1.0/"
```

Figure 10a: Machine-readable syntax (Source: Science Direct, Elsevier)

For Machine A to be understood by Machine B, Machine A must tell Machine B in which language Machine A speaks. This is readable on figure 10a at the lines 3 and 4. This is where standardisation development bodies set the tone. They do so when standardisation forums such as the CITF commissions the music – so to speak. At line 3, we see the mention of the RDF syntax of W3C. RDF stands for Resource Description Framework, a standard model for data interchange on the Web. At line 4, Machine B will see that Elsevier's TDM reservation is expressed as recommended by W3C, and Machine B will be able to understand that particular opt-out expression.

```
26              <dc:title>
27                  <rdf:Alt>
28                      <rdf:li xml:lang="x-default">M  bov  FadR mutant exhibits an altered colony
    morphotype and increased virulence</rdf:li>
29                  </rdf:Alt>
30              </dc:title>
31              <dc:creator>
32                  <rdf:Seq>
33                      <rdf:li>Vandana Singh</rdf:li>
34                      <rdf:li>Mohd Mustkim Ansari</rdf:li>
35                      <rdf:li>Debashis Dutta</rdf:li>
36                      <rdf:li>R.S. Rajmani</rdf:li>
37                      <rdf:li>Amit Singh</rdf:li>
38                      <rdf:li>Bhupendra N. Singh</rdf:li>
39                  </rdf:Seq>
40              </dc:creator>
```

Figure 10b: Machine-readable title and authors (Source: Science Direct, Elsevier)

At line 26, "dc" is the mention of the Dublin Core Metadata Element Set standardised at ISO. When Machine B reads dc:title, then Machine B knows what comes next. And Machine B also understands the list of creators (lines 31 to 40).

```
54              <pdf:Producer>Acrobat Distiller 8.1.0 (Windows)</pdf:Producer>
55              <pdf:Keywords>Mycobacteria,FadR,Rv0494,Transcriptional regulator,Mycolic acids,Lipid
    metabolism,Virulence</pdf:Keywords>
56              <pdfx:CreationDate--Text>11th September 2025</pdfx:CreationDate--Text>
57              <pdfx:CrossmarkDomainExclusive>true</pdfx:CrossmarkDomainExclusive>
58              <pdfx:CrossmarkMajorVersionDate>2010-04-23</pdfx:CrossmarkMajorVersionDate>
59              <pdfx:doi>10.1016/j.tube.2025.102689</pdfx:doi>
```

Figure 10c: Machine-readable provenance assertions (Source: Science Direct, Elsevier)

At line 54, you see a first provenance assertion: the tool which has been used to produce the article, in this case: Acrobat Distiller 8.1.0 for Windows.

On figure 10d, you recognise the short form of the rights expression (line 68), the link to the long form of the TDM reservation (line 78), the article identified by DOI (line 71), and the journal "Tuberculosis" identified by ISSN (line 72) and title (line 74).

```
68              <prism:copyright>© 2025 Elsevier Ltd. All rights are reserved, including those for text
    and data mining, AI training, and similar technologies.</prism:copyright>
69              <prism:coverDate>2025-12-01</prism:coverDate>
70              <prism:coverDisplayDate>1 December 2025</prism:coverDisplayDate>
71              <prism:doi>10.1016/j.tube.2025.102689</prism:doi>
72              <prism:issn>1472-9792</prism:issn>
73              <prism:pageRange>102689</prism:pageRange>
74              <prism:publicationName>Tuberculosis</prism:publicationName>
75              <prism:startingPage>102689</prism:startingPage>
76              <prism:url>https://doi.org/10.1016/j.tube.2025.102689</prism:url>
77              <prism:volume>155</prism:volume>
78              <tdm:policy>https://www.elsevier.com/tdm/tdmrep-policy.json</tdm:policy>
79              <tdm:reservation>1</tdm:reservation>
```

Figure 10d: Machine-readable rights assertions (Source: Science Direct, Elsevier)

In addition to the technical framework, the TDM Reservation Protocol enables transparency for rights marked as reserved. Users who are interested in obtaining these rights can reach out directly for instance via e-mail, to the publisher through the contact information provided. This approach ensures clarity and facilitates lawful access for text and data mining activities.

## Opt-out or opt-in?

Can we stack opt-out expressions? There is a natural and useful theory of positively dependent events. There is, as yet, no corresponding theory of negatively dependent events. Take one work. Assume that the work is the result of multiple generations of hybrid manipulations. A next generation occurs when a creator – human or machine – generates a new creation out of a previous one. An hybrid manipulation is performed by human and machine. And both human and machine can be plural. And humans may disagree on opt-out. This is important because text and data mining is a series of multiple generations of hybrid manipulations as we have seen on figure 7a. This argument is purely mathematical. It evokes neither the Berne Convention nor any other legal consideration. Anyway, the semantic layer is business model agnostic. It facilitates both usage policies – opt-out and opt-in. With a subtle distinction: if you can opt-in, you can also opt-out. But if you can only opt-out, for example with a simplistic expression, you can not necessarily opt-in.

## Opt-in opportunities

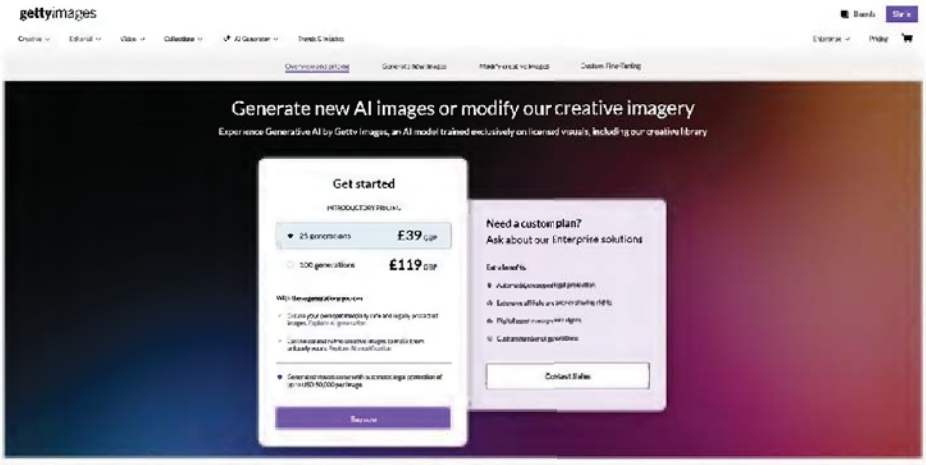



Figure 11: Image generator trained on licensed content (Source: Getty Images)

You can of course opt-in and stack opt-in expressions. This idea is getting traction. Getty Images provides an AI generator trained on licensed images. This is a commercial endeavour. Users of the AI generator register and pay a fee or a subscription. They own the IP – if any – on their productions. The rightsholders whose content has been used to train the AI generator each gets a share of the revenue generated. Customer's generated outputs are not added to the online Getty Images catalogue and therefore only accessible to the customer who has opted in.

A European collective management organisation is developing something technically and commercially similar for music.

## Stakeholders' outreach and feedback

The CITF is disseminating its work through open monthly meetings and a website.

On the 16th of June 2025, we organised a well-attended seminar focusing on the outcomes of the First Project "Interoperable, trustworthy and machine-readable copyright data: requirements in a 2030 scenario". Followed by a session on "Trust in Media" providing a glimpse into trust frameworks and verifiable credentials used in the media sector to fight deepfakes and debunk the AI slop. Followed by a panel providing space for exchanges across sectors on "Options for Creators in the AI Era".

We published the report of the First Project and the video of the seminar.

We organized feedback sessions, online, in Alicante, and in Geneva. The participants insisted on the importance of creators' control on their works and the need of a label "Made by a human" or "Made by humans". They raised questions such as what about DIY authors and performers? Who will care about them? Or should the CITF run a digital knowledge centre? They also suggested topics for the next CITF projects: deduplication of copyright records, robustness of rights management information, the impact of eIDAS and the EU Digital Identity wallets on the management of copyright, the protection and monetisation of personality rights, or the definition of contextual minimum sets of metadata.

## For a human-centred copyright infrastructure

At the three level of the copyright infrastructure, we consider the relationship between humans and their products, especially the relationships between authors and works. When we study recital 54 and article 15 of the CDSM directive, we contemplate journalists writing for fees and news publishers publishing for profit. Good so. But is it enough? Is it enough in the era of artificial intelligence? Is it enough when we don't clearly distinguish between human and synthetic creation anymore? No, it is not enough – at whatever level of the copyright infrastructure.

Figure 12 is a result of a project conducted at the Joint Research Centre in 2018-2019. The AI tsunami makes it more relevant in 2025. It shows the full space of the human condition, defined by Hannah Arendt in 1958. Her *homo faber* acts on the market axis. Her *homo civis* acts on the community axis. This is the axis where the creative industries don't function along value chains, but on a value network... and one doesn't standardise data flows on a network as one does it on a chain. If the role of a particular human is changing from context to context and from time to time, one cannot signal that role as a static attribute. Considering Arendt's *homo civis* has a direct, pragmatic impact on the semantic layer of the copyright infrastructure.


Figure 12: Considering the CDSM Art. 15 through the three angles of the human condition

Her *homo laborans* acts on the society axis. This is the axis of sustainability. This is the axis where the technical layer of the copyright infrastructure should care about the amount of electricity it consumes. This is the axis where the technical layer of the copyright infrastructure should care about the mental health of the people it employs to filter fake news and infringing contents. Considering Arendt's *homo laborans* has a direct, pragmatic impact on the technical layer of the copyright infrastructure. For example, inhumane tasks should only be performed by machines.

In the age of media machines, the humans dealing with human creation and with any layer of an infrastructure supporting human creation shall consider all aspects of the human condition to protect and foster what we cherish most – our humanity.

## Call to action

It took 6 years between the publication of the stocktaking document and today – exactly 6 years. The creative industries have 8 months left to comply with the new regulations. There is an emergency.

The CITF transitioned from a Finno-Estonian exploration to a European function while becoming the forum of Member States focused on the semantic layer of copyright and related rights infrastructure, as well as a creator and disseminator of knowledge.

The CITF now needs a growth platform to pursue its mission in an open, iterative, and interdisciplinary way, as a steward of the human condition in the AI era

**Resources:** Press release, Press, Report, Video, Website

## Contacts

Anna Vuopala, Finnish Ministry of Education & Culture
Email: anna.vuopala@gov.fi

Philippe Rixhon, Valunode OÜ
Email: prixhon@valunode.com