Interoperable and trustworthy copyright data

Brussels, 16 June 2025

Karina Bandere
National Library of Latvia

Niko Partanen
National Library of Finland

Philippe Rixhon
Valunode, Estonia

Jānis Ziediņš
Culture Information Systems Centre, Latvia

Origin and mantra 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

Rights management information must be –

- trustworthy
- interoperable, and
- machine-readable.

The three layers of the copyright infrastructure

TECHNICAL LAYER
providing mechanisms for trust and interoperation of identification and rights management information

SEMANTIC LAYER
addressing the challenges of declaration and intermediation where the meaning of identification and rights management information is made clear and consistent

FOUNDATIONAL LEGAL LAYER
defining authors' rights by conventions, treaties, directives, acts, and regulations

The Copyright Infrastructure Task Force is a forum…

… to identify and promote the standards and technologies needed to raise trustworthiness, interoperability, and machine-readability of rights data.

1st use case: AI & copyright

- Information: standardise interoperable identifiers and AI-related reservations
- Traceability: identify content that has been generated by AI
- Transparency: document the training and generation algorithms used to produce AI-generated content

Methodology

Market assessment

Prioritised use cases

Engagement plan

Liaison agreements

Terms & definitions

Requirements & criteria

Deliverables

Innovative architecture of identifiers and metadata schemas

Copyright-specific, current, interactive technology guidelines

Objectives of the first project

Objectives:

- Investigate opt-out and the use of generative AI through use cases and scenarios
- Assume future developments and formulate requirements for the copyright infrastructure
- Analyse technology-neutral requirements for their *benefits* and *challenges*
- Adapt the point of view from the library sector

Partners:

- National Library of Finland
- National Library of Latvia
- Culture Information Systems Centre (Latvia)
- TalTech (Estonia)
- Valunode (Estonia)

Perspective of the National Libraries

Typical roles fulfilled by a National Library:

- Responsibility for the description, preservation, and availability of the national published heritage
- Supporting entire library sector, working closely with museums, galleries and archives
- National publication metadata is openly available, forming the National Bibliography
- Example: National Library of Finland (NLF) provides various identifiers: ISBN, ISSN – now ISNI

Legal deposit collections as a source of materials – examples from the Finnish context:

- Printed material 1707–
- Nearly comprehensive audio record collection 1901–
- Online material 2006–
- The NLF offers legal deposit collections to researchers and other users, in line of the provisions of the Legal deposit act, within the scope of the exception rules defined in the Copyright Act

Perspective of the National Libraries, cont'd

National Libraries and copyright infrastructure

- The National Bibliography documents the published cultural heritage of a country.
- Our data is openly available and interoperable: mutual benefits, clear and distinct roles.
- In the Finnish context: existing ISNI collaboration with the Finnish CMOs

Opt-out, or other AI-related reservations

- Many possible methods of implementation
- May have a large impact to common infrastructures at different sectors

Use of generative AI

- AI-generated content will need standardized and machine-readable indications of AI use.
- Developments in these regulations at EU level demand new technical and practical solutions.

Actions of the first CITF project

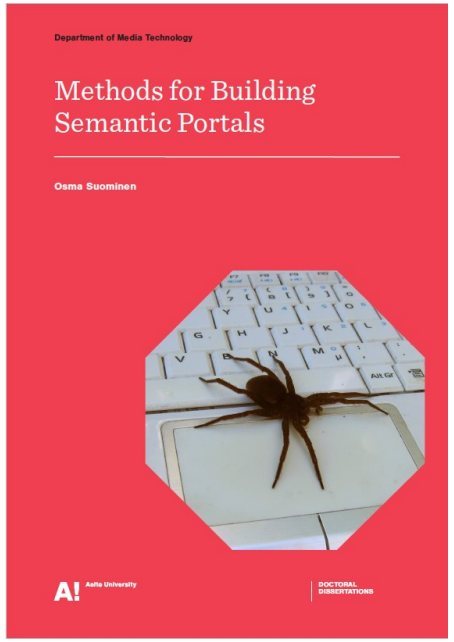
Completed work:

- Select 5 real media assets, describe 2025 life cycles, and analyse current IT infrastructure
- Consider 2030 life cycle scenarios – assuming that –
  - all content will be digitised and available online
  - AI applications will grow and multiply
- Conduct workshops on asset IDs, assertions, interconnection, and actor IDs
- Identify identifiers, metadata, AI-related reservations, and technologies required to support 2030 lifecycle scenarios
  - Resulting in report, with appendices for *Use cases*, *Requirements* and *Terms & Definitions*

Upcoming work:

- Preparing the report for publication by the Ministry of Education and Culture in Finland
- Estimated schedule: September 2025

Example of a media asset, metadata, 2025 context, and current IT support



Methods for Building Semantic Portals

PhD thesis

Writing the thesis:

- Copyright is in place immediately

Preparing for publication:

- Publication is agreed between the authors and the publisher, agreements are made with other authors and publishers.
- Potentially there could already be opt-out expressions and information about the use of generative AI.

Publication

- When the thesis is published, metadata is sent to the National Library, which provides ISBN, ISSN and ISNI.
- If generative AI would be used, the indications of this would need to be declared at this point.
- Similarly, if the author wants to declare AI-related reservations

Example of current metadata


```
{
  "@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."
}
```

First CITF Project – Selected media assets
+ metadata, 2025 life cycles, and current IT support

XX Latvian Song and X Dance Celebration
Participants at the Mežaparks Great Bandstand

Photograph, 1990

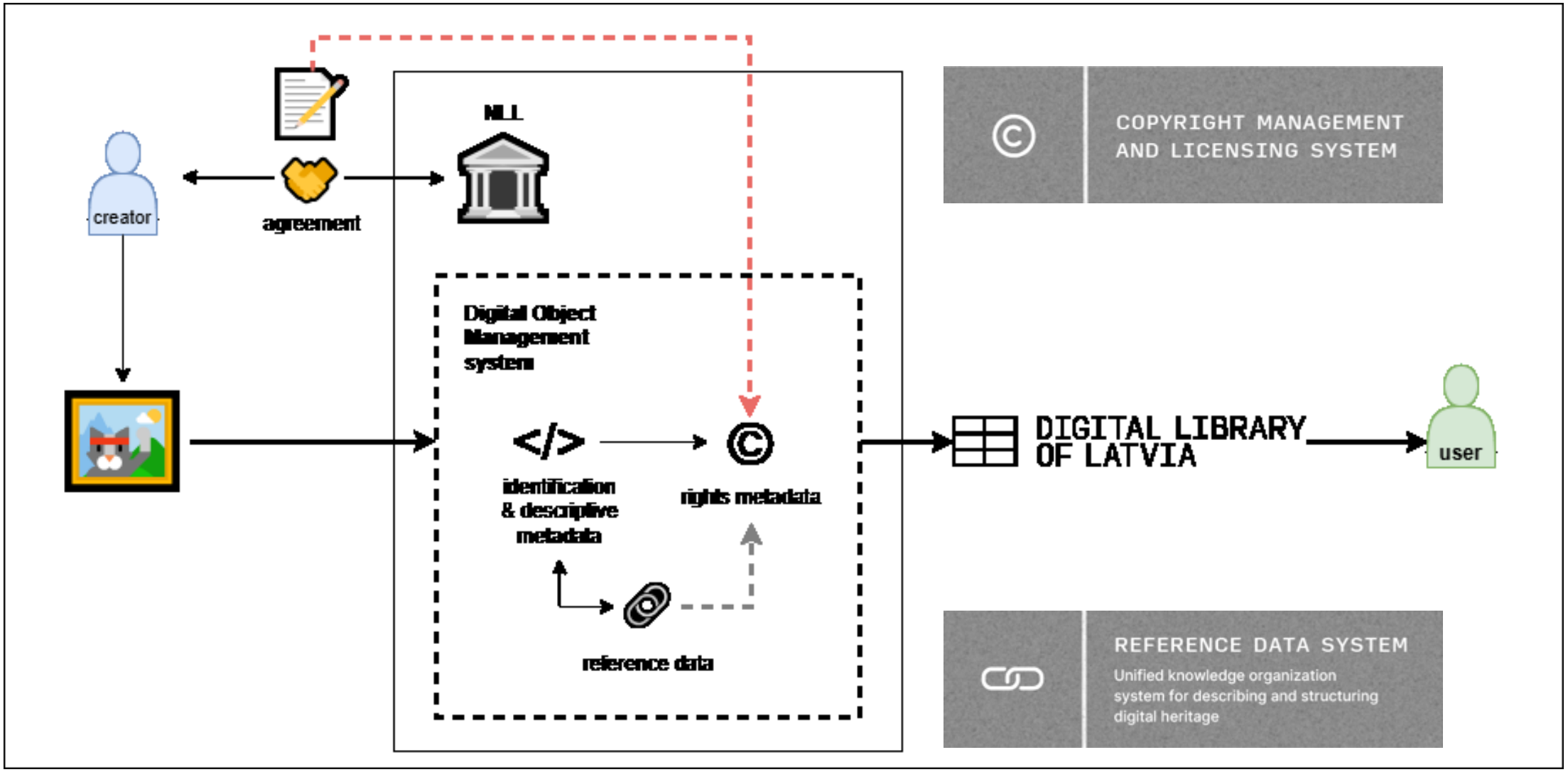
Sanita Burķīte Ilga Rampāne Aldona Volkova

Sadarbība — Turcijas bibliotēku attīstības virzītājspēks

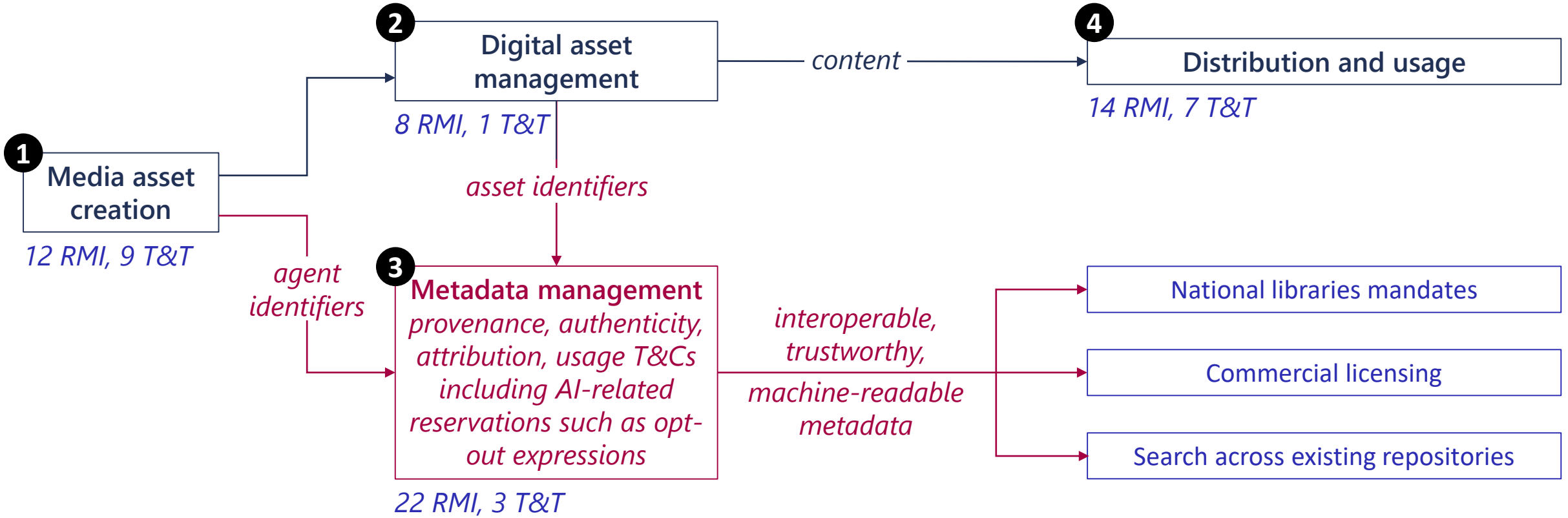
Cooperation - Aspects of the development of libraries in Turkey

Article in the periodical “Word of Libraries”, 2011

Lifecycle of a Heritage Item in a Typical Digital Workflow



Context of 75+ requirements for 3 main purposes across 4 life cycle steps in 2030



Requirements:

- *RMI: for rights management information*
- *T&T: for traceability and transparency*

Author identity, requirements (excerpts)

1) Various AGENT IDENTIFIERS shall have clearly defined purposes and explicit mappings between them that define when they can be used interchangeably and when they differ in purpose.
2) AUTHORS and CONTRIBUTORS shall be identified with standard and persistent identifiers.
3) AUTHORS and CONTRIBUTORS shall be distinguished; CONTRIBUTOR ROLES should be standardised per creative sector and declared.

Applicability:

- *NL: for National Libraries* ✓
- *CL: for commercial licensing* ✓
- *IR: for interconnected repositories* ✓

Author identity, benefits (excerpts)

- Minimum metadata requirements about an AGENT leads to increased interoperability and streamlined online transactions
- Possibility to distinguish AGENTS without using sensitive personal data
- Facilitated attribution of a WORK to an AUTHOR
- Persistence of the AGENT IDENTIFIERS given at creation or publication throughout the lifecycle of a media asset

For commercial licensing: accurate usage remunerations based on industry practices, e.g., standardised royalty splits among AUTHORS and CONTRIBUTORS

Author identity, challenges (excerpts)

- Anonymity, aliases, and data privacy
- Distinction between natural and legal persons
- Some AGENT IDENTIFIERS require a previous publication or WORK, which limits their applicability at the moment of creation.
- ROLES should be standardised and translated across countries.

Author identity, emerging technologies to watch (excerpts)

The EU Digital Identity Wallet – due by the end of 2026 – will enable AGENTS to access online and offline services, store and share digital documents, and create binding signatures.

Benefits –

- protection of personal data; AGENTS share only what is necessary,
- improved cybersecurity thanks to strong security standards,
- less administration and lower cost of AGENT authentication.

WIPO Global ID – a single and unified ID per natural person or legal entity used in global IP community, granted following a defined identity proofing system, with a common set of relevant data, data quality and access control managed directly by end-user or its delegate.

Globally accepted by participating authorities, with changes in user data being shared across participating authorities.

Reservation solutions, requirements (excerpts)

The EUIPO study *"The development of generative AI from a copyright perspective"* categorizes the following types of reservation:

- Location-based: website terms and conditions
- Work-based or repertoire-based: unilateral declarations
- Work-based: **licensing constraints**, e.g., "opt-outs" and "opt-ins"

1) Most attributions, terms and conditions are asserted by proxies. It should be possible to verify that proxies are mandated to make such assertions.

2) T&Cs assertions should be valid as long as 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.

Applicability:

- *NL: for National Libraries*
- *CL: for commercial licensing* ✓
- *IR: for interconnected repositories*

Reservation solutions, benefits (excerpts)

- Positively expressed moral and commercial rights of the original authors impacted by text and data mining activities can technically be protected.
- Authors' opportunity to opt-out does not concern use of materials in research organisations and cultural heritage organisations for purposes of scientific research: Text and data mining exception for scientific research can technically be implemented.
- The respect of reservation solutions by AI service providers – training or deploying large language models – could positively impact the quality of generated content.

Reservation solutions, challenges (excerpts)

- Notwithstanding out-of-scope commercial and legal aspects, the respect of GDPR and trade secrets could add a data processing burden to rights management processes.
- Logical considerations of opt-out mechanisms seem to conclude that it could be very complex or even impossible to scale them in the case of human-machine hybrid content – typically, during the AI training phase.
- However, opt-in mechanisms could theoretically scale, because they only require positive expressions.

GenAI traceability and transparency, requirements (excerpts)

From Art. 50 §2 of the AI Act EU/2024/1689:

1) GenAI Output shall be tagged, and the tag shall be detectable and machine-readable.
2) The tagging process should be effective – tagging what shall be tagged, interoperable, robust, and reliable.

From Art. 50 §4 of the AI Act EU/2024/1689 (re. deep fakes):

3) GenAI Output shall be disclosed as such, and the disclosure shall be detectable and human-readable.
4) Human-made and AI-generated content shall be distinguished and identified.

Applicability:

- *NL: for National Libraries* ✓
- *CL: for commercial licensing* ✓
- *IR: for interconnected repositories* ✓

GenAI traceability and transparency, benefits (excerpts)

- Consumers can use metadata to filter materials depending on whether they are generated by AI or not.
- Responsible AI can be fostered by requiring disclosure.

GenAI traceability, challenges (excerpts)

- Identification of hybrid assets, e.g., books translated by AI but edited by humans
- Definition of minor and major uses of AI
- Definition of where human creativity begins and ends

Outcomes and perspectives

Public report

1) Objectives
2) Context
3) CITF
   - National Libraries
   - Methodology
4) Terms and definitions
5) Use cases
6) Requirements
7) Annexes
   - Terms and definitions
   - Use cases
   - Requirements
   - Presentation

CITF working papers

1) Charter
2) Engagement plan
3) Actions and lessons learned
4) Forward risk analysis
5) Recommendations

Perspectives

1) Input for standardisation
2) Establishment
   - Outreach
   - Further research
   - Dissemination