

**AI Conversations 11 – Copyright Infrastructure – the backbone to making copyright work  
Senior Ministerial Adviser Anna Vuopala, 23 April 2025**

Dear audience, across the globe,

I'm honored to present some key features of the copyright infrastructure -**the backbone to making copyright work**. A two-day conference on this critical topic is what many "infrastructure enthusiasts" have hoped for a long time. Alongside some of you, **there is a rather limited group of experts dealing with this critical topic**. Among them Philippe Rixhon, whom I have had the pleasure to work with on copyright infrastructure over many many years. **Let's make this crowd bigger!**

The copyright infrastructure is about providing technological solutions to implement in practice provisions in the copyright laws. It is for **rightholders and it's for innovation**.

In order to talk about **key elements of the Copyright infrastructure**, it is important to clarify the term itself. Vocabulary is important.

The concept "copyright infrastructure" is **not coming up on search engines** often and **it does not have one single definition**.

Work in this field **requires co-operation** of experts from different backgrounds: **lawyers, technologists, economists, and practitioners like librarians who all tend to speak different languages**.

However, the copyright infrastructure, as generally understood in the EU context, is a **"a set of rules, technologies and institutions that frame the management of data in the creative sectors"** which

aims to improve **authoritative and updated information on rightsholders, terms and conditions, and licensing opportunities.**

It is not easy to grasp **all the various dimensions of it.** Think of for instance **the non-commercial uses of copyright protected works** libraries, archives and museums. The research on AI!

### **1. The problem everyone (working in copyright) can identify with**

But one data-related **problem that everyone can relate to is that creators are not being correctly and transparently paid in the digital environment.** One of the most notable issues is the so-called **black box problem** relating to undistributed mechanical rights remunerations in the music industry.

**In short.** If a work is not identified, that is, does not have an ISWC or the creator is not a member of CISAC and recognized in CIS-NET, the remuneration paid by the user, like Spotify, cannot be properly distributed.

**The fundamental purpose** of the copyright system is to ensure that authors' economic and moral rights are respected. This means to ensure that authors are **credited and paid** correctly and transparently for the use of their works.

Under EU law, remuneration for authors and performers must be appropriate, proportionate, and transparent. EU law also provides the creators with the **right to receive regular reports** on the use of their works. **Numerous sources show that creators are often neither accurately nor transparently remunerated, if at all.** Reporting is intrinsically linked to the existence of an identifier to identify the creator entitled to remuneration. **Yet, this identifier is often missing, rendering remuneration tedious or even impossible.**

**Proper infrastructure** is needed to ensure that the law works as intended. This can be achieved step by step, by coordinated action, in particular, by **identifying and supporting different use cases underpinned by legislation. We need more work on this – globally!**

At the **Copyright Infrastructure Task Force, the CITF**, legal, tech and metadata experts from **17 European countries** are specifying **an open rights data framework** that will provide rightsholders the possibility to express their terms and conditions in a machine-readable format. **This in turn will support a sustainable licensing market for works in the AI era. At the same time, it will support innovation, scientific research in artificial intelligence.**

**Traditionally metadata schemes** are developed for sector-specific purposes, so there are only a **few metadata standards** for content and rights which **are designed to be sector-neutral.**

Opening and integrating the rights data framework means **supporting and enhancing the existing framework, it does not mean replacing it.**

Opening and integrating the rights data framework **in close collaboration with 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.**

Regarding new technologies, we need to be openminded and positive. **Technologies like the “verified credentials”** can provide the required **trust framework for copyright data**. But the quality of copyright data needs to be ensured before taking up any distributed technology.

There are also a lot of identified **untapped opportunities for micro-licensing**, for example the opportunity to properly remunerate **press publishers’** neighboring rights or commercial text and data mining.

**What needs to be built is the connections** between the data repositories in a suitable manner, **most likely in a federated manner**. We also need to improve data quality and flow while **safeguarding data privacy and trade secrets**.

Finally, the **exponential increase in new content being uploaded onto platforms** makes using human resources e for the too expensive for the copyright system to cope with in an efficient manner.

**Machine readability of copyright data is not a choice, it is a must!**

## **2. How is remuneration of creators ensured now?**

**The remuneration of creators rely on “Rights Management Information (RMI) that is:**

“information which

identifies the work,

the author of the work,

the owner of any right in the work, or

information about the terms and conditions of use of the work,

**and any numbers or codes that represent such information,**

when any of these **items of information is attached to a copy of a work**

**or appears in connection with the communication of a work to the public”** (Art. 12 (2) WIPO

Copyright Treaty; comp with Art. 19 (2) WIPO Performances and Phonograms Treaty).

**In practice this consists of identifiers, metadata schemes, data-repositories and IT systems**

-**International standard identifiers** such as ISWC for musical works, ISRC for recordings, ISBN for books, and ISAN for audiovisual works. These are developed separately in each industry.

- There are also some **sector-neutral** identification systems, like the ISNI and the ISCC, that are now being increasingly taken up. These provide for new mechanisms for the capture and curation of metadata. AND

-**Databases and other data repositories that contain copyright data** such as databases of CMOs, national copyright registries, and data repositories of cultural heritage institutions.

Generally speaking, the types of metadata that are mostly collected in these databases are

- **descriptive metadata** for instance type of content, genre and so on.

**Identification metadata on works and authors is a key element** from the copyright point of view and

for this there is a multitude of identifiers as I just mentioned.

Another key element is

- **“rights metadata” which means elements related to rights, terms and conditions**, or IP rights restrictions (for instance “opt outs” from text and data mining). For this type of metadata, **open and sector-neutral standard should be used.**
- Why does this matter? It means that unlike content or parties, licenses **do not have international standard identifiers** – and computers cannot unambiguously handle things unless they have unique identifiers. The availability and use of these **identifiers** will enable the automation of licensing, distribution, and remuneration and make these processes simpler, faster, more accurate and less expensive.

Furthermore, there are **metadata on use of works** (usage data) alongside the **administrative metadata that provides trust in the aforementioned data.**

Collective management organisations play **a key role in a functioning copyright infrastructure.**

The EU directive on collective management from 2014 lists in several articles information standards that CMO’s must respect. The WIPO Collective Management Toolkit from **this year includes the most recent source of information relating to international identifiers (NB! the only single place they are all listed in so far).** It also lists the **Exchange formats and protocols** (such as CIS-NET and IDA) and the **IT schemes used by the creative industries** (such as DDEX and ONIX).

We hear about problems relating to the respect **of the law, not about the lack of respect of the “infrastructure”**. **A big issue not addressed in any way is the unauthorized removal of rights management information, data stripping**. Still this data is protected by the 1996 WIPO Internet Treaties which have been in force more than 20 years.

### **Issues with copyright registers**

Copyright is granted automatically upon the creation and fixation of an original work without the need for any formalities. Consequently, there is **no single source of information** on copyright protected works. Data repositories holding various kinds of copyright data at CMOs, national libraries, and tech companies are fragmented into several hundred around the globe.

**None of these databases are complete or up to date**. All of them would benefit from data enrichment, which can only be **achieved through large scale global co-operation**.

**However, a mapping study is being commissioned by the EUIPO at this very moment**. It is a detailed sector-by-sector mapping of all public and private initiatives to record, manage or facilitate access to information on copyright protected works. The study's findings should serve as foundational research for the feasibility of a service allowing users of the copyright infrastructure to **access and view data included in these databases**. This will be **an important source of information to paint the broader picture of the dispersed and fragmented data environment**.

### **Industry initiatives**

Even today, after years of facilitated discussions only a limited number of **cross-sectoral initiatives have emerged**.

However, the audiovisual and news sector has organized themselves in a EU consortium to blueprint a **Trusted Media Data Space**. This data space will create common rules for data sharing and act as a hub for different kinds of services, like fact checking, licensing and access to copyright data.

**The Mechanical Licensing Collective in the US** partially succeeded to tackle the “black box” problem. However, significant sums still 1 billion dollars of remuneration **remain undistributed to the rightful owners**. Collaboration with rightsholders to match metadata is ongoing but slow.

**Some stakeholders, like actors and scriptwriters or choreographers, have been kept outside copyright infrastructure developments**. To get the work started, funding has been granted for **crediting data in a single portal**, following a discussion on copyright infrastructure development at national level (in Finland).

Interesting initiatives were also shared in the **AI information session** at SCCR 46 that referred to promoting **metadata relating to local enterprises, traditional knowledge and traditional cultural expressions**.

### **3. What can be done about it?**

In 2022, the EU Commission assessed **the problems related to copyright data** management in creative industries in the **study Copyright and New Technologies in 2022**. The study covered music, publishing, audiovisual and images sectors. It collected data from European Union Member States and also the United States, United Kingdom and Canada. **It is a valuable source of information for the global level as well.**

After a thorough survey, and although the situation had improved a lot in music and publishing for instance, the **study assessed that the copyright infrastructure was deemed to be not yet fully functional.**

The study recommended among other things:

- **raising copyright awareness** on the side of creators and users,
- **mastering metadata** by developing education, in particular for professionals in the creative industries,
- **defining data governance** *i.e.* raising trust in rights management information – who can issue and verify credentials on authors and works

and

- **opening the rights data framework:** making it trustworthy and sector-neutral by defining a limited sets of standards with which the industries could **update existing or build new metadata solutions.**

**At the CITF we have started to work together on a common copyright infrastructure,** while building a **level playing field for standards and technology innovation** that everyone can use and benefit from.

This work is based on synergies and cooperation to support **the IP potential of creatives,** SMEs, and society at large.

## **Now a few words about the use cases to develop machine readable terms and conditions**

**The Copyright Infrastructure Task Force, the CITF**, is a network or a coalition "of the willing" to promote "trustworthy, interoperable and machine-readable copyright data".

Our objective is to **support the entire copyright ecosystem**, that is all the different entities in the cultural and creative sectors, to build common approaches to identify works, parties, and their interrelations in a trustworthy manner.

The CITF is **meeting monthly online** and reports to the Council working party on Copyright. It is **working in close cooperation** with EU institutions, EUIPO, **its planned Copyright Knowledge Center**, the WIPO and its initiatives relating to infrastructure and AI. The **CITF also liaises with industry projects to seek synergies and alignments** and promotes and coordinates the access **to EU funding for the needed** investments.

**The EU AI Act coming into force in August this year contains obligations for General Purpose AI providers to respect copyright law** and *in particular the reservations related to text and data mining "through state-of-the-art technologies"*.

In practice, this **requires the deployment of new solutions that leverage identifiers and metadata schemes that are currently standardized**. Therefore, during the past 1,5 years, the CITF has worked on an AI and copyright use case that specifically addresses that. We are looking at **machine-readable terms and conditions, that are ways for the rightholder to express preferences relating to the use of their works**.

The first project of the CITF is run by **practitioners from the National Libraries of Finland and Latvia supported by IP experts from the Tallinn University of Technology.**

The project **builds on 2030 life cycle scenarios assuming that:**

- **all content will be digitised and available online,**
- **only machines will manage rights and royalty,**
- AI applications will grow and multiply, and
- creators of original material will be fairly, appropriately, proportionally, and transparently remunerated.

The project, that will present its outcomes on the 16<sup>th</sup> of June, aims to **define requirements to express in machine readable language the replies to the following questions:**

1) **who did what**, as creator or contributor

2) **what is it?** A work or other protected subject matter

**3) what can be done with it?** The terms and conditions for the use, licenses, including exceptions and payment of remuneration.

**Questions that sound simple but** require a whole lot of common effort to be expressed in a uniform way.

This would be the beginning of an **open rights data framework**. This framework would grow with each new use case that would **build upon and be informed** by the previous use cases!

**Coming back to what would make copyright work,**

we need to **focus on our own and our partners' practices relating to rights management information**. Copyright Infrastructure **is often considered invisible or "unseen" to most outside each industry**. In my mind, this is due to **too few opportunities to talk about common denominators in needs**, and whether there could **be a common sector-neutral set of metadata standards to use as a basis to build IT systems and services in accordance with specified industry needs**.

- **The elephant must be cut into edible pieces**. This exercise begins with drawing attention to, and taking stock of **developments** relating to metadata on works, authors, attribution, terms and conditions – at national and regional level – The easiest way is to participate in **coordinated use case activity** – one model is the First CITF Project, maybe by national IP Offices, to define technical criteria to express trustworthy, interoperable, and machine-readable copyright data.

**Then we could gather on WIPO level to discuss and take action!**

**We also need to raise awareness**. We need to promote **the visibility** of the copyright infrastructure at the crossing of **law, business practice, and technology**. A global coordinated campaign for awareness and education would be excellent!

We can build on CLIP. We can use Global ID's. **The EUIPO is likely to host the CITF in the future as part of the Copyright Knowledge center opening in November 2025. This center**

**will be providing creators, rightsholders and users** the best possible support to manage copyright in the AI era.

**To conclude,**

#### **4. How can you contribute to the solution?**

We all know it is necessary to address the needs and challenges of fast-evolving technologies and we all know that now is the time – we are already late.

**Therefore, I urge each and every one listening today to commit to a proper copyright infrastructure.**

**The question is,** how can we involve all aspects of the infrastructure and make copyright work in the AI era? **We should welcome building a platform for copyright infrastructure here at WIPO!**

I hope for an enlightening exchange today **that focuses on technical solutions** and not on policy or legal amendments that belong elsewhere, such as the standing committees.

I thank you!