

WIPO Conversation on IP and Frontier Technologies, Eleventh Session

Sharing session and Open floor discussion, 24 April 15.20 – 16.50 Geneva time.

Statement by Giulia Marangoni, AIE International Relations department and Co-Chair of the W3C CG Text and Data Mining Reservation Protocol.

The TDMRep Protocol for rights reservation in the context of AI

Good afternoon,

My name is Giulia Marangoni, I'm responsible for International Relations at the Public Affairs department of the **Italian Publishers Association** and I'm the **Co-Chair of the W3C Community Group Text and Data Mining Reservation Protocol**, which developed an opt-out solution known as TDMRep.

- The **TDMRep** was designed to provide rightsholders with a **machine-readable solution** for rights reservation for text and data mining in compliance with the **EU Copyright Directive in the Digital Single Market**
- The Directive provides for **an exception for text and data mining for purposes other than research**, unless rightsholders reserved their rights in an appropriate manner, including machine readable means for content publicly available online
- Afterwards, the **EU AI Act recalled that if an AI provider relies on a TDM exception, it must respect the TDM opt-out** under the Copyright directive

Since the first release of the protocol in 2022 and subsequent updates in 2024, the TDMRep has been **widely adopted by the European publishing industry**, including major European books, press and STM publishers, digital platforms and distributors, and has become a **consolidated industry practice**. It is also used beyond the publishing industry and outside of Europe.

TDMRep is based on open standards and protocols widely used in the web and in the content industry. It features two key elements:

- A **tdm-reservation element**, to specify if TDM rights are reserved, in compliance with the DSM directive
- A **tdm-policy element**, providing information on the licenses available, either in human readable format (such as the publisher's license information web page) or in machine-readable files, using ODRL format (Open Digital Rights Language 2.2)

Different implementation methods are possible. Currently, TDM Rep currently applies to:

- **Web content**, at any level of granularity (audio, text, images, video)
- to **EPUB files**, embedded in EPUB metadata, in compliance with the Epubcheck

- To **PDF files**, embedded in XMP metadata
- In principle, it is scalable to other content types and formats

This makes TDMRep the only opt-out solution to be usable both in a location and asset manner, making it extremely flexible. In addition, the TDM opt-out and policy can be expressed within **ONIX for Books format**, the main standard to communicate book product information along the book supply chain.

The **web implementation** of the TDMRep was specifically developed as **horizontal opt-out**, meaning that rights reservation apply to any AI operator without needing to name crawlers or class of crawlers individually. This simplifies greatly administration by rightholders, without making it more complicated for AI actors.

At technical level, the protocol was developed as a **separate and compatible solution** from tools such as **robot.txt**, specifically designed **to control content findability**. Indeed, a prerequisite for the TDMRep protocol was to allow rightholders to reserve TDM rights without affecting access to other services such as content indexing and ranking in search pages results.

The **W3C Community Group** that developed the TDMRep was co-chaired by **EDRLab and AIE** and promoted by the **Federation of European Publishers (FEP)**, with the involvement of national publishers' associations and the STM association and in contact with the European Commission. Now the group features more than 50 members from the publishing sector, and **representatives of other opt-out initiatives that collaborate with TDMRep**.

Currently, the **W3C Community group TDMRep** is in close contact with other **opt-out initiatives to foster interoperability** between different machine-readable solutions and to discuss specific **use cases for more granular vocabulary** for rights reservations in the context of AI.