TDM Reservation Protocol

Quick history

- 2019: EU Copyright Directive for a Single Market (CDSM)
- 2021: Creation of the TDMRep CG at W3C
- 2022: first version of the specification (HTTP, file on a website, and HTML)
- 2024: second and third versions (plus EPUB and PDF)
- 2024: EU AI Act comes into force
- 2025: EU Code of Practice for General-Purpose AI - EU AI Transparency Template

TDM, some definitions

Text and data mining *means any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations.*

Recital 18: New technologies enable the ***automated computational analysis of information in digital form****, such as text, sounds, images or data, generally known as text and data mining. Text and data mining makes the processing of large amounts of information with a view to gaining new knowledge and discovering new trends possible.*

EU Copyright Directive

Text and data mining *is the use of automated analytical techniques to analyse text and data for patterns, trends and other useful information.*

Intellectual Property Office, UK

Text and data mining *is the process of deriving information from machine-read material. It works by copying large quantities of material, extracting the data, and recombining it to identify patterns.*

UK Government

TDMRep in brief

One boolean property

- 1 = TDM rights reserved.
- 0 (or nothing) = TDM rights not reserved.

An optional Web link

- Pointing at a TDM Licensing Policy setup by the rightsholder.

W3C ODRL

TDM Reservation Protocol

Links

- TDMRep Specification:
  https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240510/

- Project pages:
  https://w3c.github.io/tdm-reservation-protocol/

- Community Group page (with posts):
  https://www.w3.org/community/tdmrep/

TDMRep adoption

- See https://w3c.github.io/tdm-reservation-protocol/docs/adopters.html
- **In and outside the EU**
- Major trade publishers
- Major academic and STM publishers
- Major digital platforms (ebook distributors, aggregators)
- Major newspapers

-> TDMRep is one of the opt-out mechanisms studied by EUIPO in is study "The development of Generative AI from a Copyright Perspective" (2025) url

-> TDMRep cannot be ignored by AI providers (see Code of Practice)

How TDMRep is linked to the EU CDSM

Article 4

"1. Member States shall provide for **an exception or limitation to the rights** provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive **for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining**. 17.5.2019 L 130/113 Official Journal of the European Union EN

3. **The exception or limitation** provided for in paragraph 1 **shall apply on condition that the use of works** and other subject matter referred to in that paragraph **has not been expressly reserved by their rightsholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online**."

How the EU AI Act is linked to the CDSM

Article 4

Providers of general-purpose AI models shall: ...

(c) put in place a policy to comply with Union law on copyright and related rights, and in particular **to identify and comply with**, including through state-of-the-art technologies, **a reservation of rights** expressed pursuant to **Article 4(3) of Directive (EU) 2019/790**;

url

How the work at IETF s linked to the EU AI Code of Practice i

robots.txt .... AI-Pref

Copyright Chapter contains a crucial measure (Measure 1.3), which commits signatories (AI providers) to:

"employ web-crawlers that read and follow instructions expressed in accordance with the **Robot Exclusion Protocol (robots.txt)... and any subsequent version** ..."

Crucially, the CoP also says AI providers must:

"identify and comply with **other appropriate machine-readable protocols** to express rights reservation..."

These sections indirectly point to the W3C TDMRep and the IETF AI-Pref.

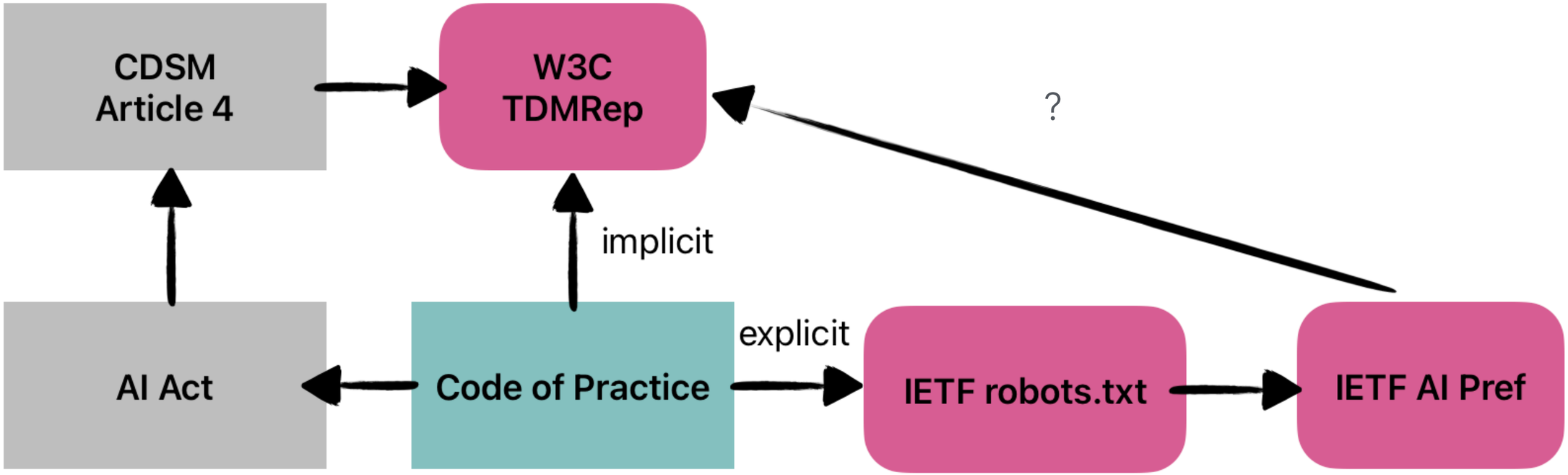
Measure 1.3 Identify and comply with rights reservations when crawling the World Wide Web

(1) In order to help ensure that Signatories will identify and comply with, including through state-of-the-art technologies, machine-readable reservations of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790 if they use web-crawlers or have such web-crawlers used on their behalf to scrape or otherwise compile data for the purpose of text and data mining as defined in Article 2(2) of Directive (EU) 2019/790 and the training of their general-purpose AI models, Signatories commit:

a) to employ web-crawlers that read and follow instructions expressed in accordance with the Robot Exclusion Protocol (robots.txt), as specified in the Internet Engineering Task Force (IETF) Request for Comments No. 9309, and any subsequent version of this Protocol for which the IETF demonstrates that it is technically feasible and implementable by AI providers and content providers, including rightsholders, and

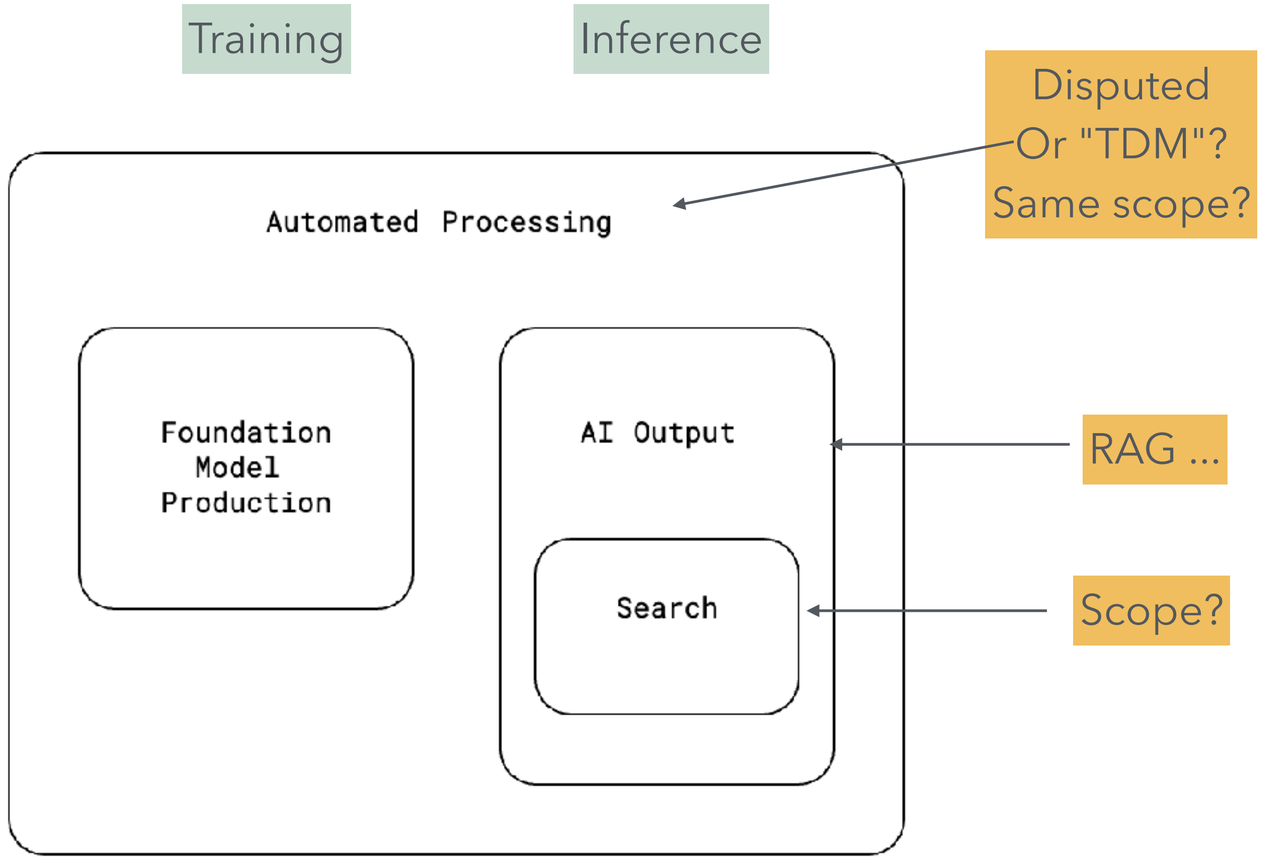
b) to identify and comply with other appropriate machine-readable protocols to express rights reservations pursuant to Article 4(3) of Directive (EU) 2019/790, for example through asset-based or location-based metadata, that have either have been adopted by international or European standardisation organisations, or are state-of-the-art, including technically implementable, and widely adopted by rightsholders, considering different cultural sectors, and generally agreed through an inclusive process based on bona fide discussions to be facilitated at EU level with the involvement of rightsholders, AI providers and other relevant stakeholders as a more immediate solution, while anticipating the development of standards.

Relationship between TDMRep and AI-Pref



IETF AI-Pref, draft 4

- Vocabulary: https://www.ietf.org/archive/id/draft-ietf-aipref-vocab-04.html
- in robots.txt: https://www.ietf.org/archive/id/draft-ietf-aipref-attach-04.html



ex: *bots=y, train-ai=n, ai-output=n, search=y*

Main differences

- TDMRep directly addresses the EU CDSM Article 4 (but can be used elsewhere)
  - Does not try to re-define "Text and Data Mining"
  - Does not state NO, but RIGHTS RESERVED
  - Is perceived as an all or nothing approach, which does not take into account subtleties of modern Search pipeline.
- AI Pref aims at offering a fined-grained vocabulary for the usage of content by AI systems, including Search pipelines.
  - Every bit is heavily discussed; the consensus will take time.
  - Its top level term ("Automated Processing") is particularly controversial.
- **Can we blend both approaches?**

Possible evolution of TDMRep

Keep tdm-reservation & tdm-policy as-is, extend the ODRL policy

```
{
  "@context": [
    "http://www.w3.org/ns/odrl.jsonld",
    "http://www.w3.org/ns/tdmrep.jsonld"
  ],
  "@type": "Offer",
  "profile": "http://www.w3.org/ns/tdmrep",
  "uid": "https://provider.com/policies/1",
  ...
  "permission": [{
      "action": "tdm:mine",
      "constraint": [{
        "leftOperand": "purpose",
        "operator": "neq",
        "rightOperand": "ietf:train-ai"
        }
      ]
    }
  ]
}
```

ODRL is rich

```
{
  "@context": [
    "http://www.w3.org/ns/odrl.jsonld",
    "http://www.w3.org/ns/tdmrep.jsonld"
  ],
  "@type": "Offer",
  "profile": "http://www.w3.org/ns/tdmrep",
  "uid": "https://provider.com/policies/1",
  ...
  "permission": [{
      "action": "tdm:mine",
      "constraint": [{
        "leftOperand": "purpose",
        "operator": "eq",
        "rightOperand": "ietf:train-ai"
        },
      "duty": [{
        "action": "obtainContent"
        }]

      ]
    }
  ]
}
```

Operator	Description	Example Usage
`eq`	Equal to	`age eq 18`
`neq`	Not equal to	`country neq "US"`
`lt`	Less than	`price lt 10.00`
`lteq`	Less than or equal to	`count lteq 5`
`gt`	Greater than	`dateTime gt "2025-12-31"`
`gteq`	Greater than or equal to	`dateTime gteq "2025-01-01"`
`isA`	Is a member of (e.g., belongs to a set or category)	`userType isA "premium"`
`hasPart`	Contains or includes	`document hasPart "chapter3"`
`isPartOf`	Is contained within	`user isPartOf "organizationX"`
`isAllOf`	Must include all of the specified values	`tags isAllOf ["educational", "free"]`
`isAnyOf`	Must include at least one of the specified values	`format isAnyOf ["PDF", "EPUB"]`
`isNoneOf`	Must not include any of the specified values	`region isNoneOf ["US", "CA"]`
`in`	The left operand is in the set defined by the right operand (spatial/temporal)	`location in "FR"`

Signal in the http header, along with IA-Pref signal

Possible evolution of TDMRep - along with AI Pref

```
HTTP/1.1 200 OK
Date: Wed, 14 Jul 2021 12:07:48 GMT
Content-type: text/html
tdm-reservation: 1
tdm-policy: https://provider.com/policies/policy.json
Content-Usage: train-ai=n
```

Issue: what if there is discrepancy between HTTP prefs and ODRL prefs?

The TDMRep spec could indicates that HTTP takes precedence over ODRL.

Signal embedded in publications

Possible evolution of TDMRep - no AI Pref equivalent

Publication

tdm-reservation = 1

tdm-policy ---------->

Any Website

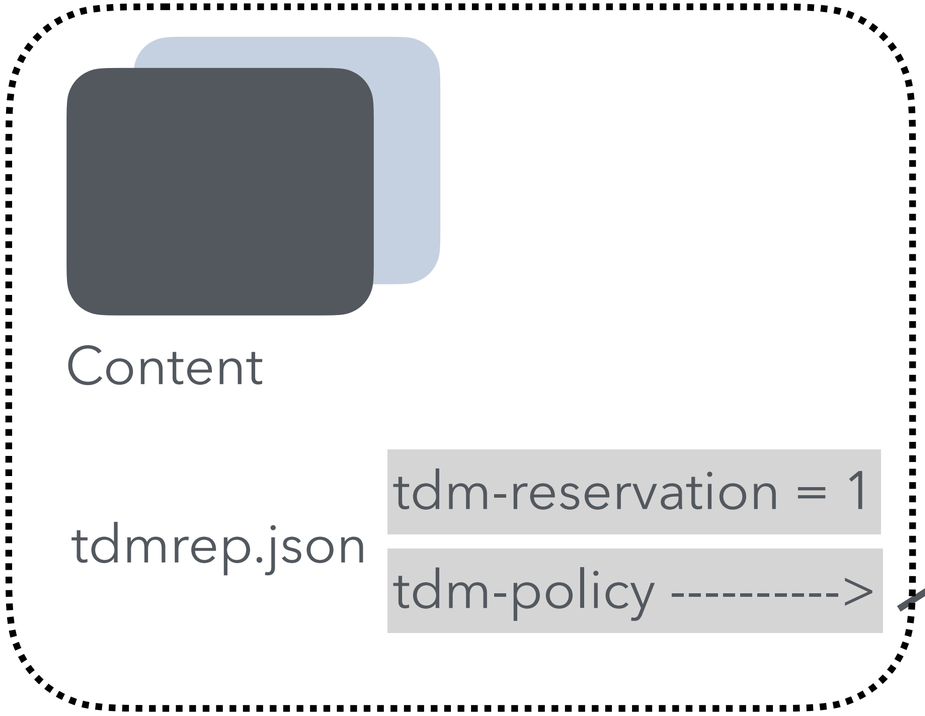
```
{
  "@context": [
    "http://www.w3.org/ns/odrl.jsonld",
    "http://www.w3.org/ns/tdmrep.jsonld"
  ],
  "@type": "Offer",
  "profile": "http://www.w3.org/ns/tdmrep",
  "uid": "https://provider.com/policies/1",
  ...
  "permission": [{
      "action": "tdm:mine",
      "constraint": [{
        "leftOperand": "purpose",
        "operator": "neq",
        "rightOperand": "tdm:train-genai"
        }
      ]
    }
  ]
}
```

No modification of existing publications :-)

Signal in tdmrep.json

Possible evolution of TDMRep - Alternative = adopting robots.txt

Web Server



Standarization of TDMRep

- Currently a CG
- Is there a real requirement to make in an international standard before AI Pref has its vocabulary ready? (see the AI Code of Practice)
  - If yes, most likely ISO (using fast-track).
- If we're not in a hurry, if AI Pref succeeds and we agree with its vocabulary:
  - We'll extend TDMRep with the AI-Pref vocabulary.
  - Then go for a W3C recommendation or ISO standard.