CONTENT RECOGNITION TOOLS AND THE NEW EU COPYRIGHT DIRECTIVE (2019/790)
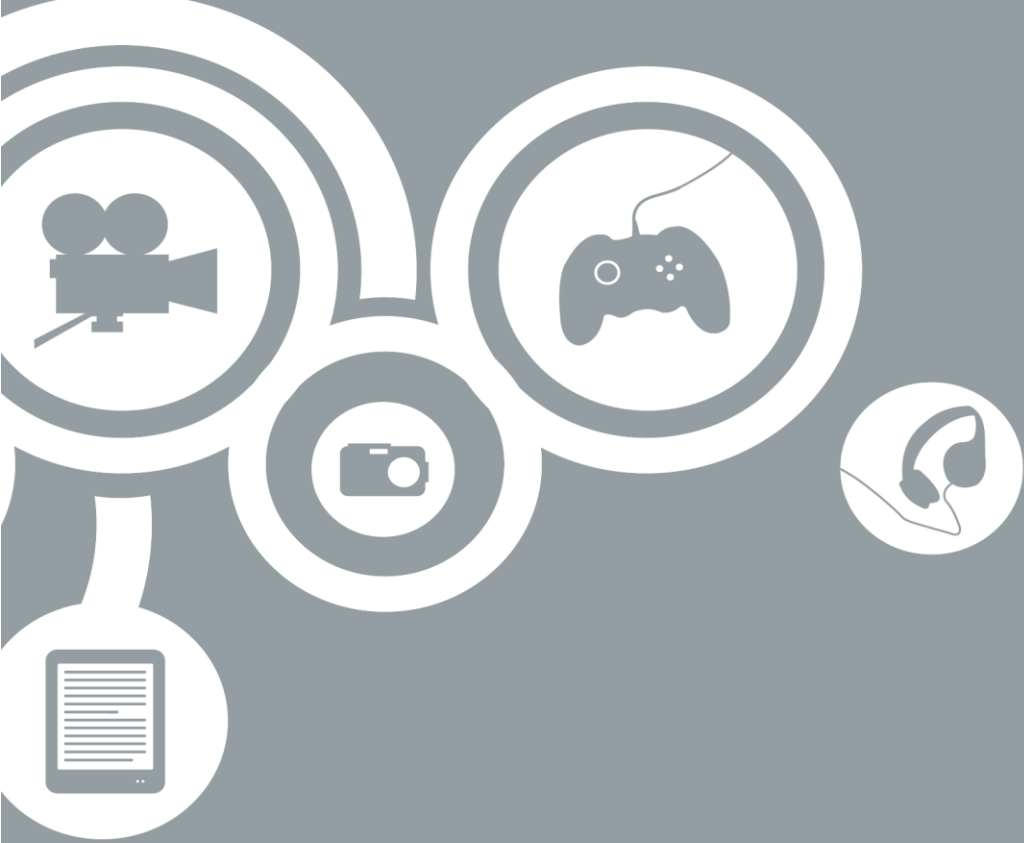
State for the art, deployment, expectations and main learnings.

CSPLA – Hadopi – CNC Mission

CONTEXT

Hadopi

A NEW EUROPEAN COPYRIGHT DIRECTIVE

- Before the vote of Article 17 of the European Directive (EU) 2019/790 of 17 April 2019, online sharing platforms were considered as hosting platforms, not responsible by default for the contents posted by their users.

- Now, some online sharing platforms are considered as realising acts of exploitation when they offer copyrighted contents that have been uploaded by their users. They must therefore seek authorisation from rightholders to do so. If no autorisation is granted, platforms must demonstrate that :

  - They have made their "best efforts" to seek appropriate authorisations.
  - They make their "best efforts" to disable access to unauthorised protected works, based on the "relevant and necessary information" provided to them by rightholders.
  - Upon notification, they remove and prevent the re-upload of unauthorised contents.

Hadopi

THE CSPLA – HADOPI – CNC MISSION

- The new legal framework does not mention any particular technology but, by referring to the "high standards of the sector", gives in fact a central role to content recognition tools.

- On April 1, 2019, a joint mission on content recognition tools has been initiated in France, with a triple objective:

  - Make an up-to-date inventory of the existing tools (especially the ones that are used by platforms).
  - Evaluate the efficiency of such tools.
  - Make recommendations in relation to the transposition of the directive.

- The mission gathered for the first time representatives and experts from the Higher Council on Literary and Artistic Property (CSPLA), the Authority for the Dissemination of Works and the Protection of Rights on the Internet (Hadopi) and the National Centre for Cinema and the Moving Image (CNC).
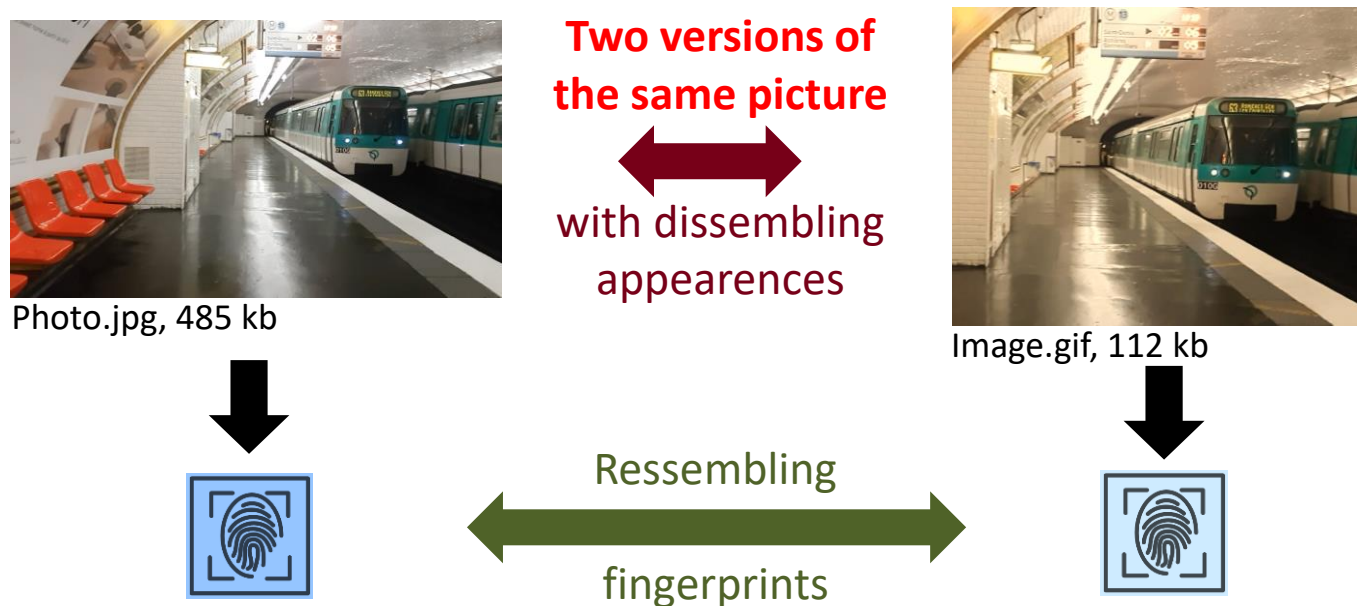
THE MAIN CONTENT RECOGNITION SOLUTION AS OF TODAY: FINGERPRINTING

CONTENT RECOGNITION BASED ON DIGITAL FINGERPRINTS

- The main technology used for automated content recognition on online sharing platforms as of today is called "fingerprinting".

- Content recognition is made based on the comparison of digital fringerprints.

Photo.jpg, 485 kb

Two versions of the same picture

with dissembling appearences

Image.gif, 112 kb

Ressembling

fingerprints

- A digital fingerprint is a simplified representation of a content.

- Fingerprinting technique applies to audio, video and still images. It can also work for text and possibly for software and applications (video games).
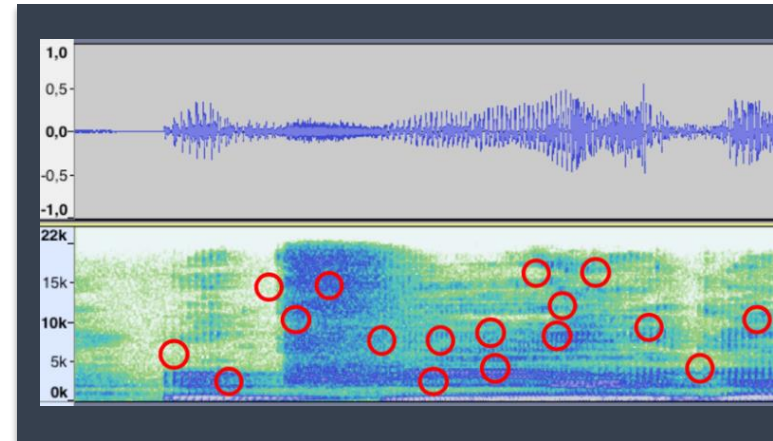
Hadopi

CONTENT RECOGNITION BASED ON DIGITAL FINGERPRINTS

Conceptual illustration of image fingerprinting

Source : CNRS-IRISA (L. Amsaleg)

Conceptual illustration of audio fingerprinting

Source : CNRS-IRISA (L. Amsaleg)
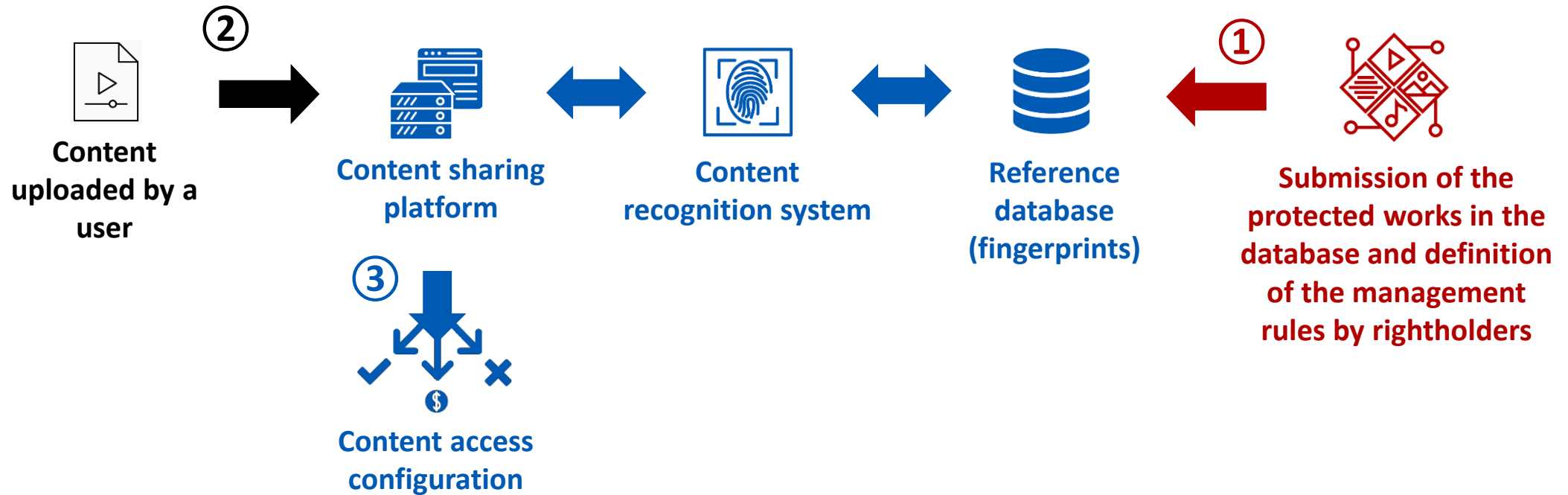
Conceptual illustration of video fingerprinting

Source : Ina – Institut National de l'Audiovisuel

Conceptual illustration of fingerprinting as possibly applied to text

Hadopi

SIMPLIFIED FUNCTIONING OF FINGERPRINTING SYSTEMS

Content uploaded by a user

②

Content sharing platform

③

Content access configuration

Content recognition system

Reference database (fingerprints)

①

Submission of the protected works in the database and definition of the management rules by rightholders

1. Generation of the reference fingerprints

Point of attention : possible conflicts between fingerprints

2. Definition of management rules

Point of attention : possible contradictory rules

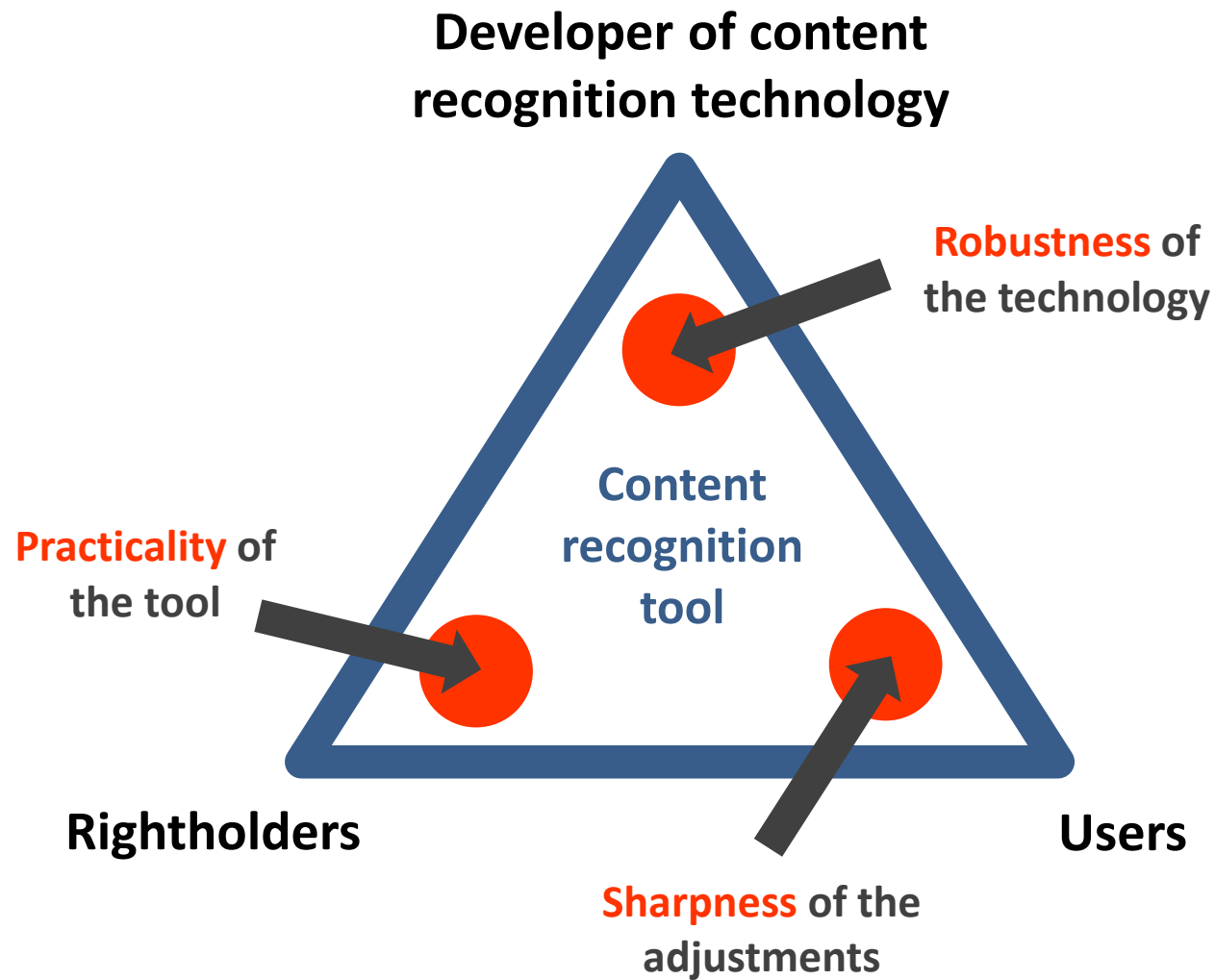## 3. Disputes management or claims resolution

It is sometimes possible for users to truncate or remove the copyrighted material from the uploaded content in order to resolve the claim

CONTENT RECOGNITION TOOLS' ASSESSMENT

Hadopi

ASSESSMENT OF CONTENT RECOGNITION TOOLS

Developer of content recognition technology

Robustness of the technology

Content recognition tool

Practicality of the tool

Sharpness of the adjustments

Rightholders

Users

- The **capability** and the robustness of the technology are just one facet of the content recognition tools' assessment.

- For complete evaluation, the following aspects must also be observed:

  ➢ The **functionalities** offered to rightholders and the practicality of their implementation.

  ➢ The sharpness that rightholders demonstrate in the way they **use** tools, taking into account copyright exceptions.

Hadopi

GLOBAL METHODOLOGY

- Public and private evaluation protocols exist in order to assess the efficiency of content recognition tools – but the methodologies (and the results) are not always published.

- The goal is to test tools in an exhaustive or targeted way and compare the observed results with the expected ones.

- Methodology chosen for the study: « stress tests » rather than a global evaluation protocol (due to limited time and resources), with 4 sets of tests with increasing complexity, inspired by practical observations and particular cases.

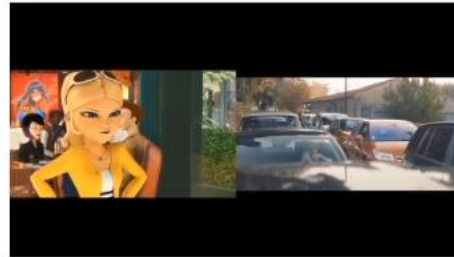
Hadopi

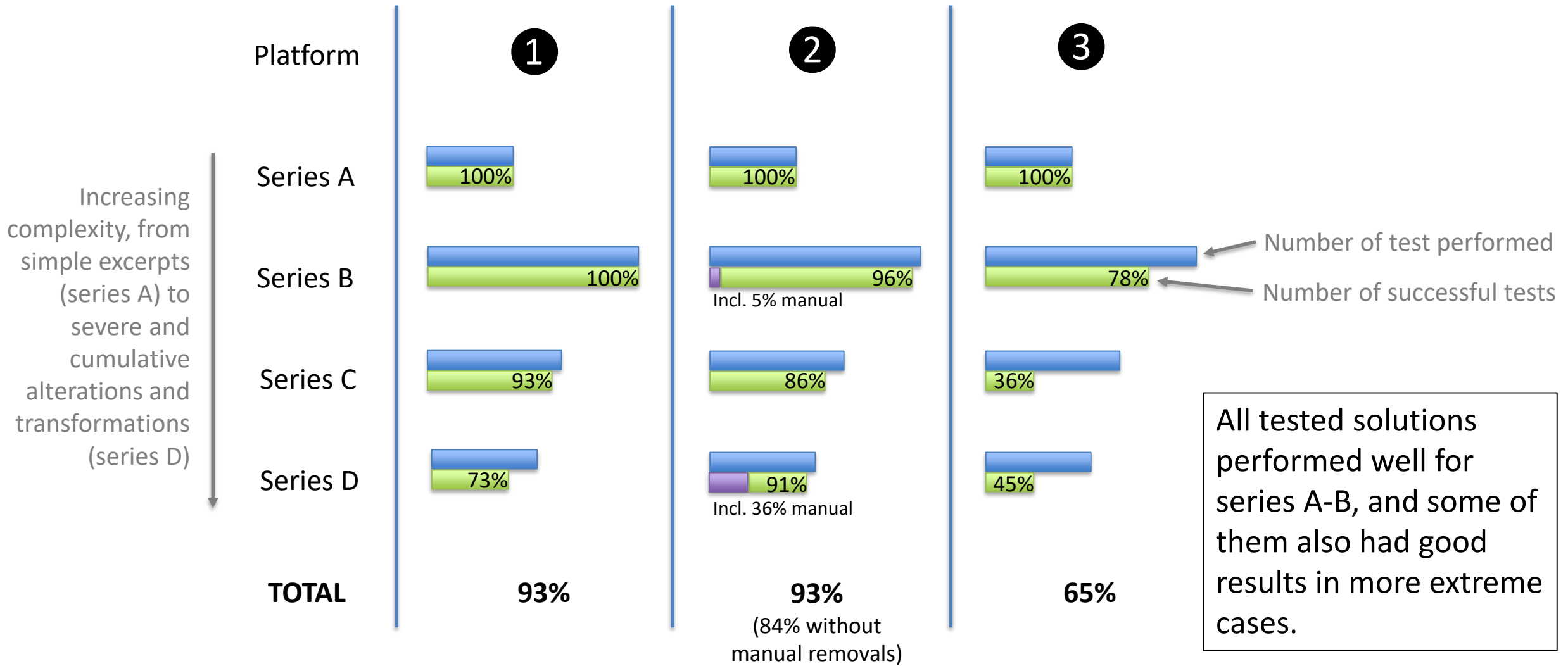# SETS OF TESTS (AUDIOVISUAL CONTENTS)

Series A	Series B	Series C	Series D

Sources : Gaumont et TF1

EXAMPLES OF RESULTS (AUDIOVISUAL CONTENTS)

Platform: ❶ ❷ ❸

Increasing complexity, from simple excerpts (series A) to severe and cumulative alterations and transformations (series D)

	❶	❷	❸
Series A	100%	100%	100%
Series B	100%	96% Incl. 5% manual	78%
Series C	93%	86%	36%
Series D	73%	91% Incl. 36% manual	45%
TOTAL	**93%**	**93%** (84% without manual removals)	**65%**

Number of test performed
Number of successful tests

All tested solutions performed well for series A-B, and some of them also had good results in more extreme cases.

MUSICAL CONTENTS

- YouTube, Facebook and Shazam have been tested as a user of the platform and Audible Magic has been tested directly without the intervention of a platform (Type-3 i20 solution).

Normal excerpt   Speed variation   Tone variation   Multiple alterations

- Varying results... but for a reason:

  ➤ Strong alteration tolerance with tools that are intended to be flexible in the way they work (but occasional false positives).
  ➤ Good alteration tolerance on mainstream platforms, where uploaded contents are often of average quality.
  ➤ Lower alteration tolerance with tools that are intended to be more precise (but no false positives).

STILL IMAGES

Basic tests done with the IMATAG and Videntifier technologies with regard to the recognition of still images. The different technologies may be used depending on the platforms' needs (search for exact images and/or similar images).
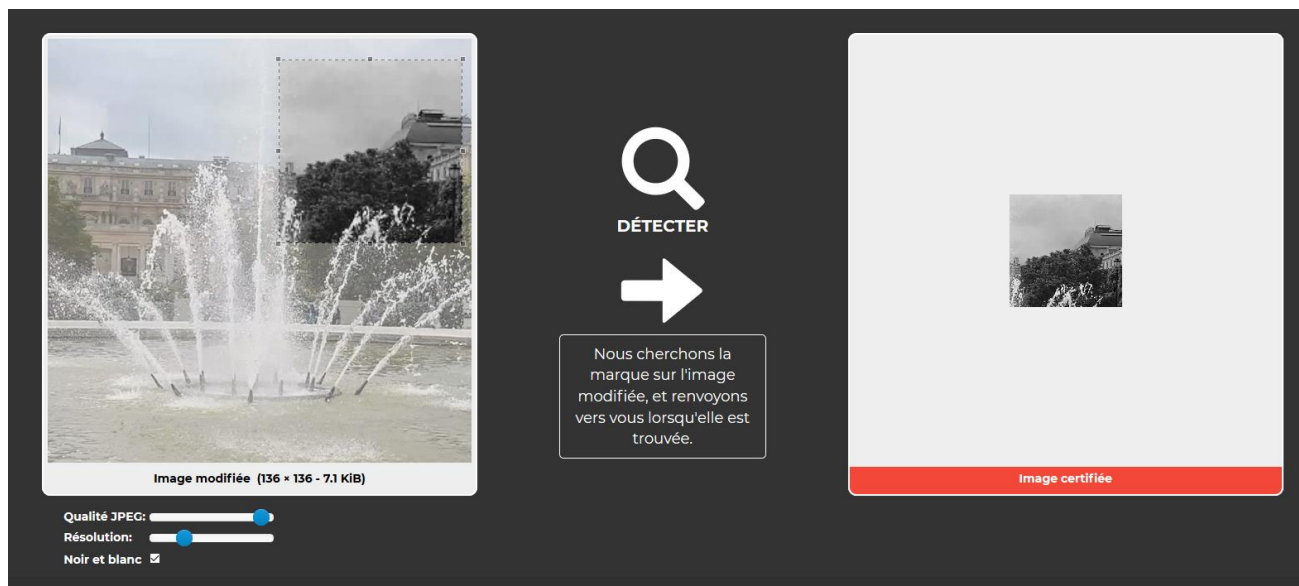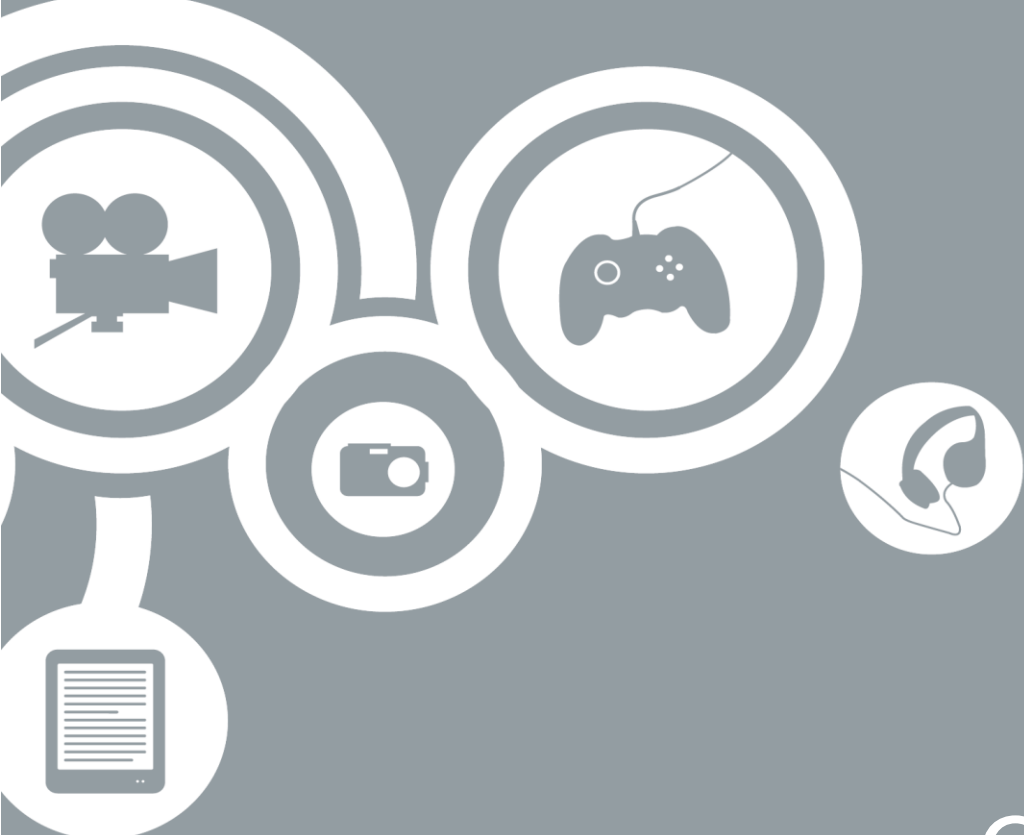

Source : IMATAG

Example of <u>fidelity check</u> (identical content)

Source : Videntifier

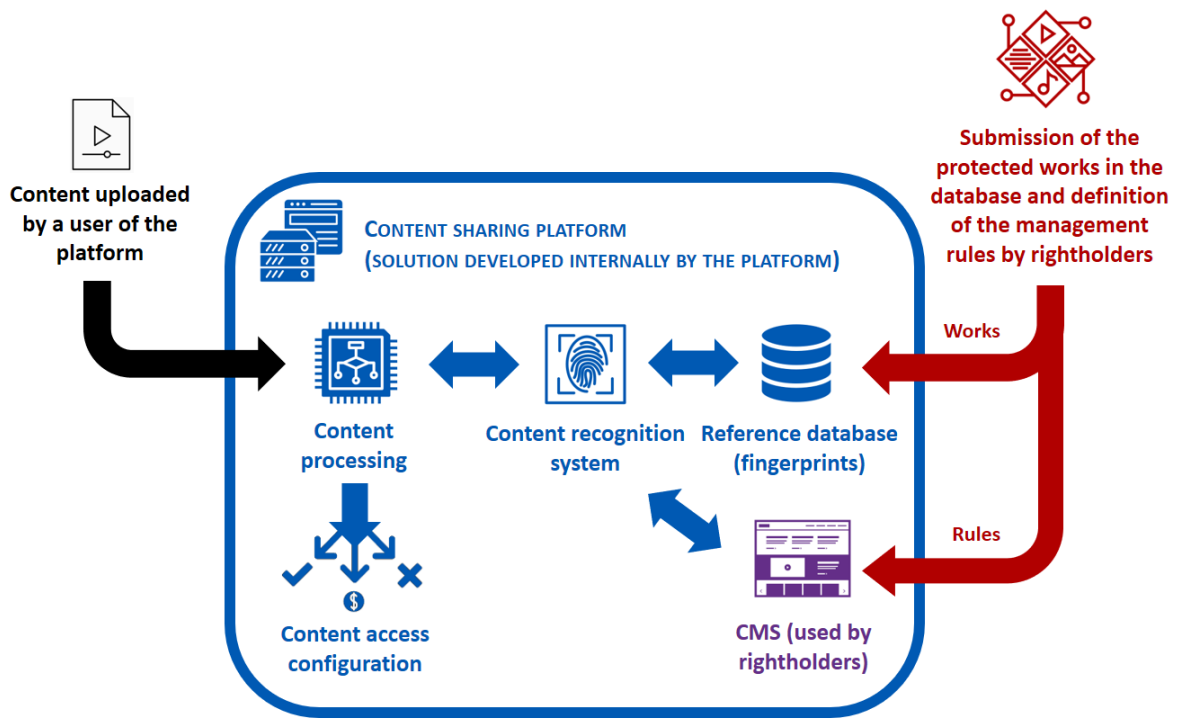Example of <u>similarity check</u> (similar content)

ORGANISATIONAL MODELS OF FINGERPRINTING TOOLS
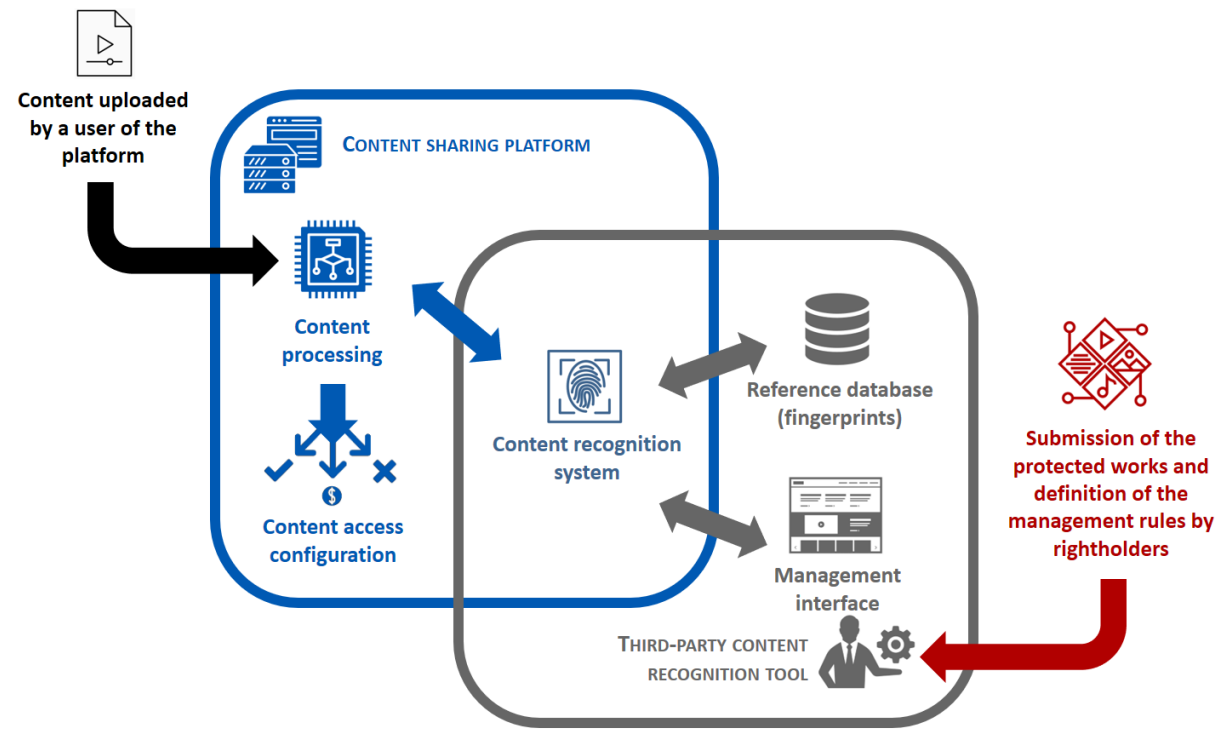
TYPICAL ORGANISATIONAL MODELS

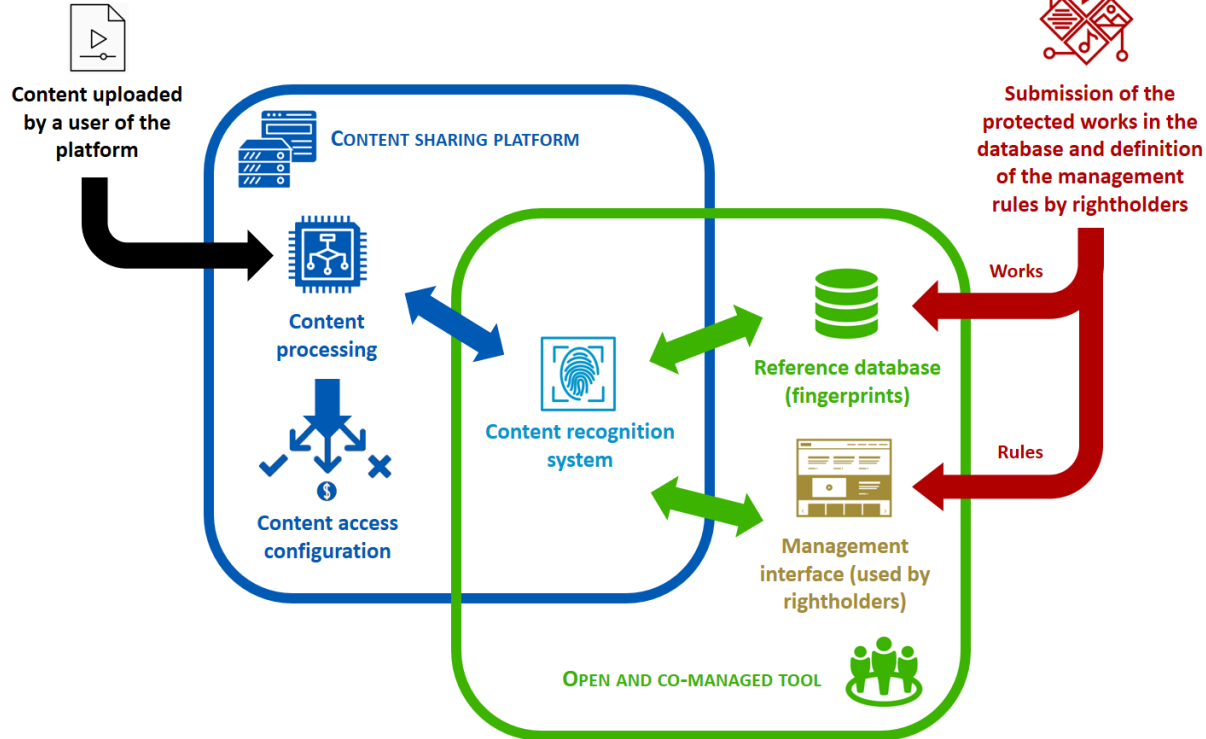Tools developed by the platforms

Tools developed by third parties

Examples :

YouTube Content ID

Rights Manager

Examples (used by dailymotion, twitch, SOUNDCLOUD, etc.) :

Audible Magic

ina-signature TAKE THE FINGERPRINT OF YOUR CONTENT

Hadopi

ALTERNATIVE ORGANISATIONAL MODELS

Open and co-managed tools

Content uploaded by a user of the platform

CONTENT SHARING PLATFORM

Content processing

Content access configuration

Content recognition system

Reference database (fingerprints)

Management interface (used by rightholders)

OPEN AND CO-MANAGED TOOL

Submission of the protected works in the database and definition of the management rules by rightholders

Works

Rules

Example :

Qwant (project)

Centralised service provision

Submission of the protected works and definition of the management rules by rightholders

CENTRALISED SERVICE (ONE-STOP SHOP)

CONTENT SHARING PLATFORMS (SOLUTIONS DEVELOPED INTERNALLY BY THE PLATFORMS)

THIRD-PARTY CONTENT RECOGNITION TOOLS

OPEN AND CO-MANAGED TOOLS

Examples :

ALPA
ASSOCIATION DE LUTTE CONTRE LA PIRATERIE AUDIOVISUELLE
GUICHET UNIQUE

Blue Efficience
Le service d'empreinte universelle
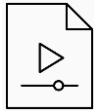
Hadopi

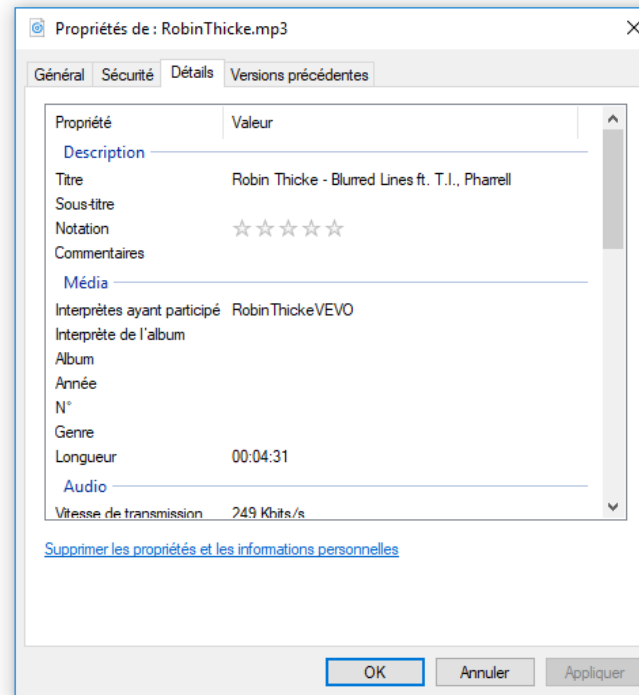# OTHER CONTENT RECOGNITION SOLUTIONS

Hadopi

OTHER EXISTING SOLUTIONS

Hashing : to recognise easily identical files.

d1921aa0ca3c1146a01520c04e6caa9e

Metadata analysis : a basic but fragile method.

Digital watermarking: an interesting but still underused alternative.

User-ID   Network-ID

Content-ID

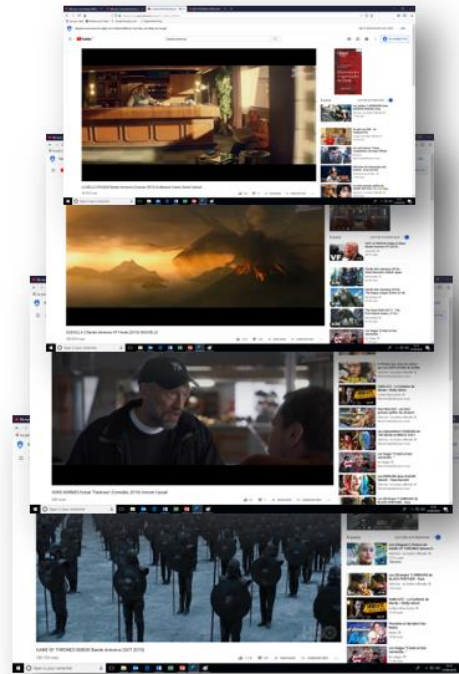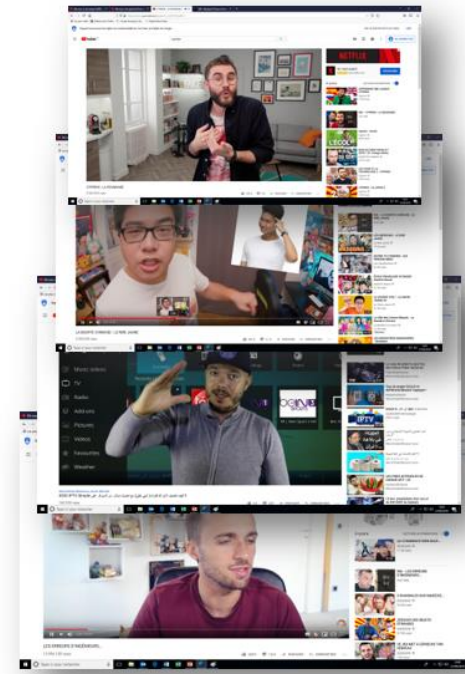# PROSPECTS AND COMPLEMENTARY OR ALTERNATIVE SOLUTIONS

Hadopi

- Machine learning can help to improve current technologies. It can also help to recognise content types without requiring fingerprints but this technique has inherent limits and constraints.

Sport

Drama

"YouTubers"

OTHER COMPLEMENTARY OR ALTERNATIVE SOLUTIONS

Automated speech recognition (ex. : YouTube) and comparison with reference databases.
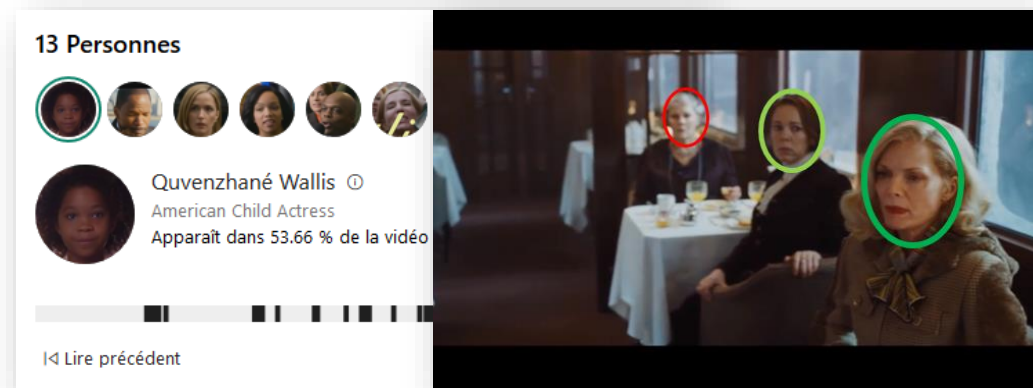

Recognition of logos or distinctive marks.

Optical character recognition (ex. : lyrics appearing on some videos).

Actors or character recognition and comparison with databases containing photographs and lists of casts.

Hadopi

- Description of image content, to be compared with reference databases (synopsis, summaries, etc.) – this technique is still experimental, but already used by Facebook.

- In the future: description and analysis of actions and dialogues (called « story analysis » or « action analysis »), plagiarism or reappropriation detection, multiformat content recognition (example: text v. video, etc.)

INTERNET USERS SURVEY

Hadopi

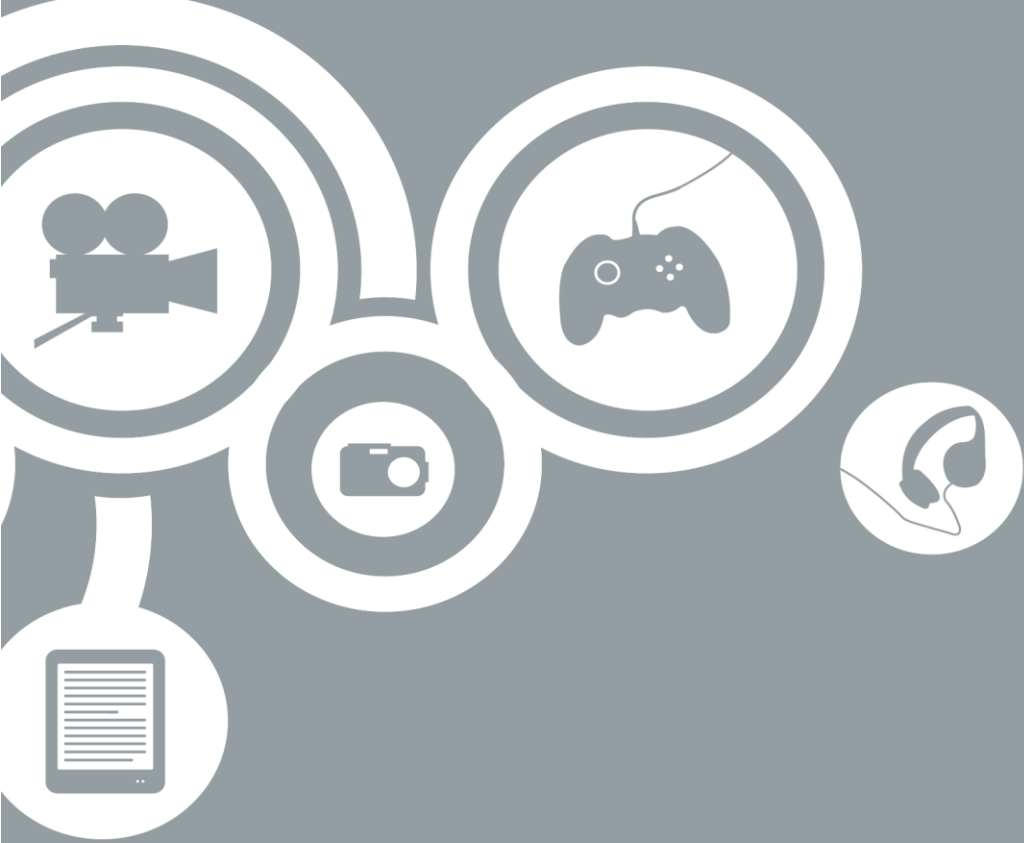# METHODOLOGY

Survey conducted by the OpinionWay institute.
User samples have been surveyed through an auto-administrated online questionnaire on the CAWI (Computer Assisted Web Interview) system.

Sampling :

PHASE 1 : FRAMEWORK	PHASE 2 : FOCUS ON PEOPLE SHARING AUDIO AND VIDEO CONTENTS
3040 Internet users **Aged 15 ans and above, representing the Internet users population** *according to gender, age, socio-professional categories, region, urban typology, consumption rate of cultural goods online and piracy rate** → **Determination of the penetration rates of people sharing contents online and people who had some shared contents blocked**	**1000 Internet users who shared video and audio contents online, based on the profiling made during phase 1** **+** **« Boost » sample for a total of 300 sharers who had some contents blocked,** *a minima* → **Determination of the different types of blockings, dispute rate and follow-up**

+

Phase 1 was conducted from August 19 until September 4, 2019

Phase 2 was conducted from October 23 until November 15, 2019

*Quotas taken from the Hadopi barometer standard, April 2019

Hadopi

RELATIVELY GOOD UNDERSTANDING OF PLATFORMS' RULES

In bold and colour: correct answer	TRUE	FALSE
Posting excerpts from music or films by another author does not require permission	31%	**69%**
Posting musical scores online does not require authorisation	32%	**68%**
Posting a live video of one's television as it shows the broadcast of a competition does not require authorisation	35%	**65%**

Base : all Internet users

Hadopi

GOOD UNDERSTANDING OF BLOCKING RULES…

Platforms and social media may remove content when the actual author has prohibited it from being shared
- True: 87%
- False: 13%
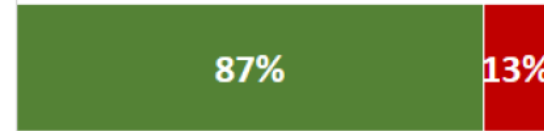
Platforms and social media may prevent cultural content from being posted on-line
- True: 75%
- False: 25%

Creators can receive compensation if they so desire, every time a user posts their work on content-sharing
- True: 71%
- False: 29%

Legend: ■ True ■ False

Base : all Internet users

Hadopi

...BUT A LOWER UNDERSTANDING OF RULES RELATED TO COPYRIGHT EXCEPTIONS

In bold and colour: correct answer	TRUE	FALSE
Posting a parody of music, a film or other cultural content does not require permission	**55%**	45%
Individuals may post content of which they are not the author if they clearly indicate who is the actual author of the content	67%	**33%**

Base : all Internet users

Hadopi

GOOD SOCIAL ACCEPTANCE OF RULES IMPLEMENTED BY THE PLATFORMS (1/2)

Platform rules that appear justified or not by Internet users

	Completely warranted	Somewhat warranted	TOTAL
To enable authors to be compensated	44%	43%	87%
To enable platforms to comply with copyright	47%	45%	92%
To battle the piracy of cultural content (music, films, series, etc.)	49%	41%	90%

Base : all Internet users

Fully agree

Social media and video platforms should remove content if an author objects to its being shared

| 48% | 41% | 7% | 3% | **89 %** |

It is normal that authors exercise their rights if they wish to block their content or receive remuneration

| 40% | 48% | 8% | 3% | **88 %** |

The author's refusal to let their work be shared constitutes a form of censorship

| 12% | 35% | 35% | 17% | **47 %** |

The author's refusal to let a work be shared harms the creation

| 12% | 34% | 39% | 14% | **46 %** |

Content should not be blocked by authors, because sharing helps advertise their work

| 11% | 37% | 38% | 13% | **48 %** |

■ Strongly agree   ■ Somewhat agree   ■ Somewhat disagree   ■ Strongly disagree

Base : all Internet users

Hadopi

A THIRD OF INTERNET USERS SHARE AUDIO AND VIDEO CONTENTS ON SOCIAL PLATFORMS

80% of Internet users have at least one social media account*

58% of Internet users share content on the social media

Original works only — 24%

Combination of content — 19%

Personal content only — 42%

Videos — 31%
Audio recordings — 6%

33% of Internet users share audio or video content

Images — 48%

Texts — 30%

Sports — 8%

Video games — 4%
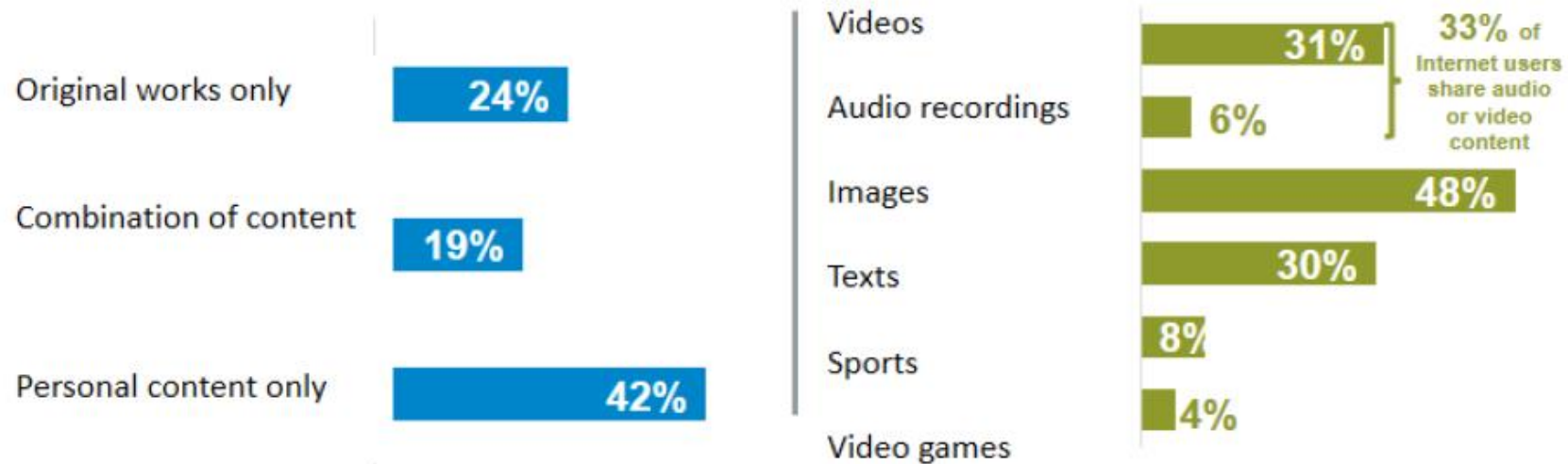
*YouTube, Facebook, Instagram, Twitter, Dailymotion or Reddit

Base : all Internet users

15% OF INTERNET USERS SHARE ORIGINAL WORKS

24% of Internet users share only original works		19% of Internet users share mixed works		42% of Internet users share only personal works	
Videos	15%	Videos	10%	Videos	15%
Audio recordings	3%	Audio recordings	2%	Audio recordings	3%
Images	18%	Images	13%	Images	35%
Texts	14%	Texts	9%	Texts	17%
Sports	4%	Sports	3%	Sports	4%
Video games	2%	Video games	1%	Video games	2%

Base : Internet users ages 15 and above

Hadopi
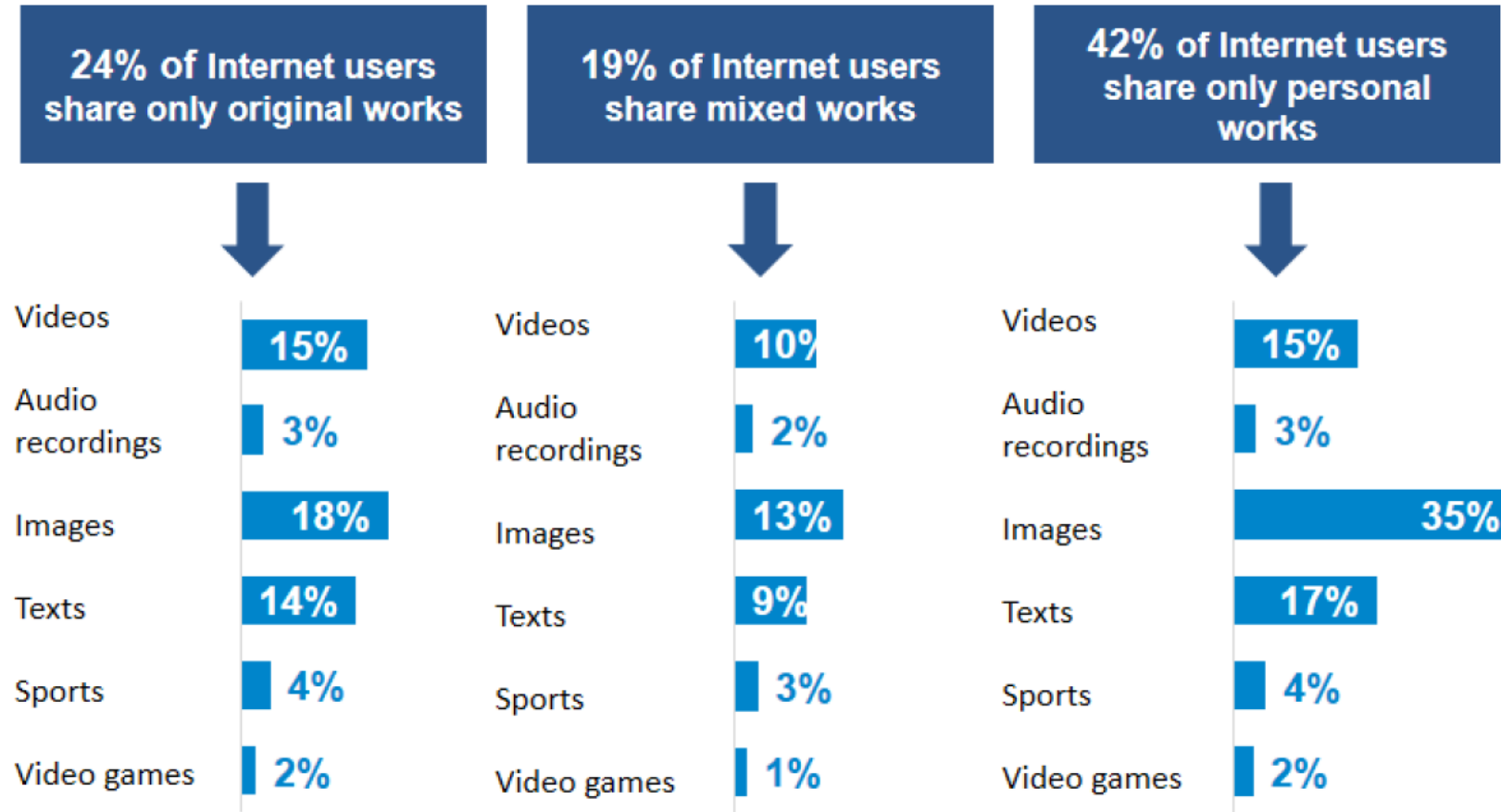
ALMOST HALF OF THE BLOCKINGS ARE DISPUTED

58% of Internet users share content on the social media

11% are blocked

6% of Internet users have received messages announcing that their content has been blocked from platforms for copyright reasons

3% after sharing exclusively original works

51% for original works

29% for mixed content

2% after sharing mixed content

43% contested the decision

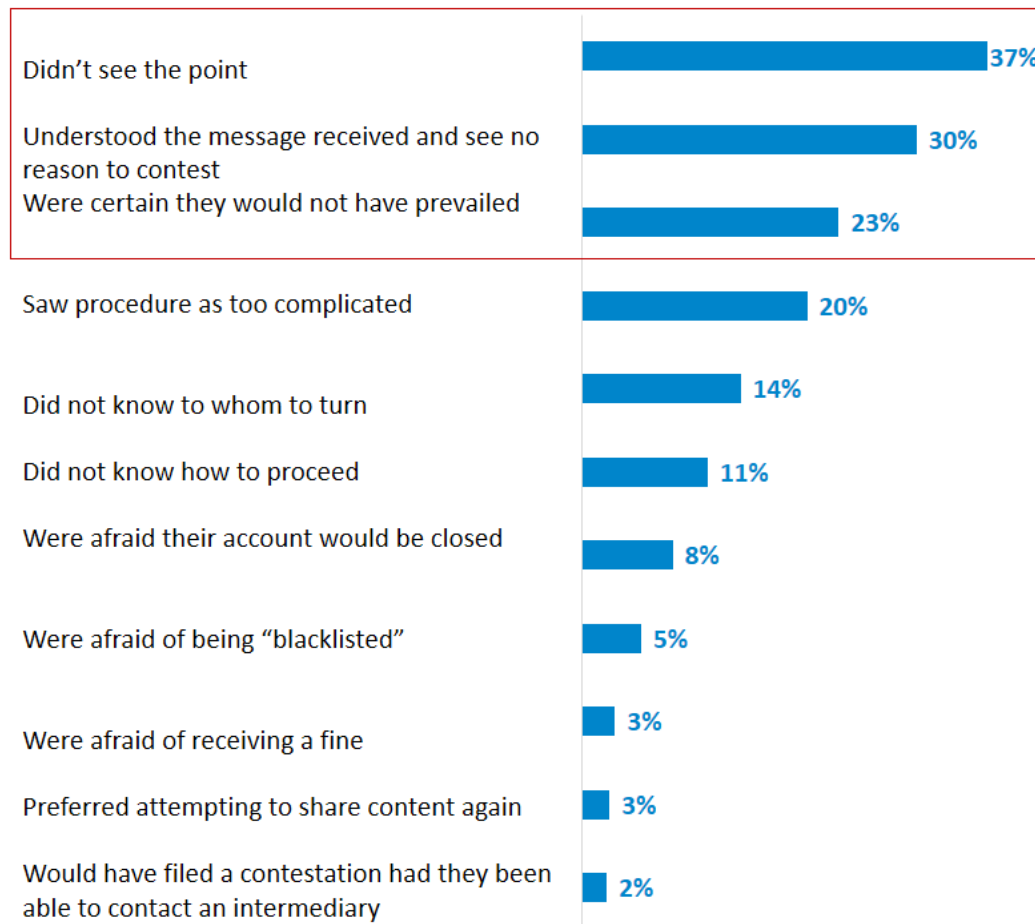**3% of Internet users have challenged the blocking of their content at least once**

Base : Internet users ages 15 and above

Hadopi

UNDISPUTED BLOCKINGS ARE DUE TO A LACK OF INTEREST OR DUE TO THE UNDERSTANDING OF THE REASON WHY THE BLOCKING OCCURRED

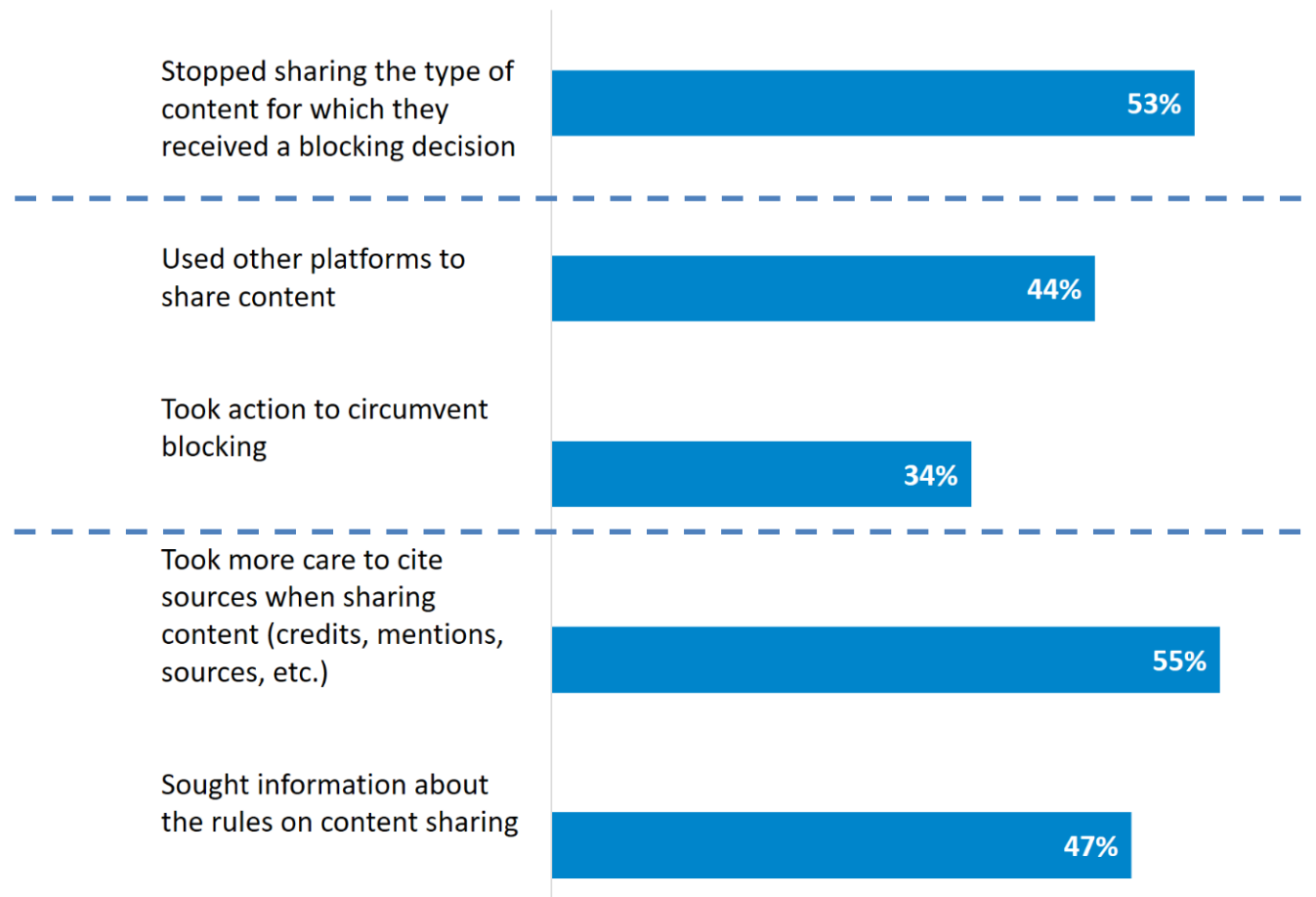*Reasons for not disputing the blocking of an audio or video content*

Reason	%
Didn't see the point	37%
Understood the message received and see no reason to contest	30%
Were certain they would not have prevailed	23%
Saw procedure as too complicated	20%
Did not know to whom to turn	14%
Did not know how to proceed	11%
Were afraid their account would be closed	8%
Were afraid of being "blacklisted"	5%
Were afraid of receiving a fine	3%
Preferred attempting to share content again	3%
Would have filed a contestation had they been able to contact an intermediary	2%

Base : Internet users ages 15 and above who have shared audio or video contents and have not disputed the last encountered blocking

Hadopi

INTERNET USERS WHOSE SHARED CONTENTS HAVE BEEN BLOCKED CHANGE THEIR PRACTICES ACCORDINGLY

Actions taken following the latest blocking of a shared audio or video content

Action	%
Stopped sharing the type of content for which they received a blocking decision	53%
Used other platforms to share content	44%
Took action to circumvent blocking	34%
Took more care to cite sources when sharing content (credits, mentions, sources, etc.)	55%
Sought information about the rules on content sharing	47%

Base : Internet users ages 15 and above who have had a shared audio or video content blocked (latest event to date)

Hadopi
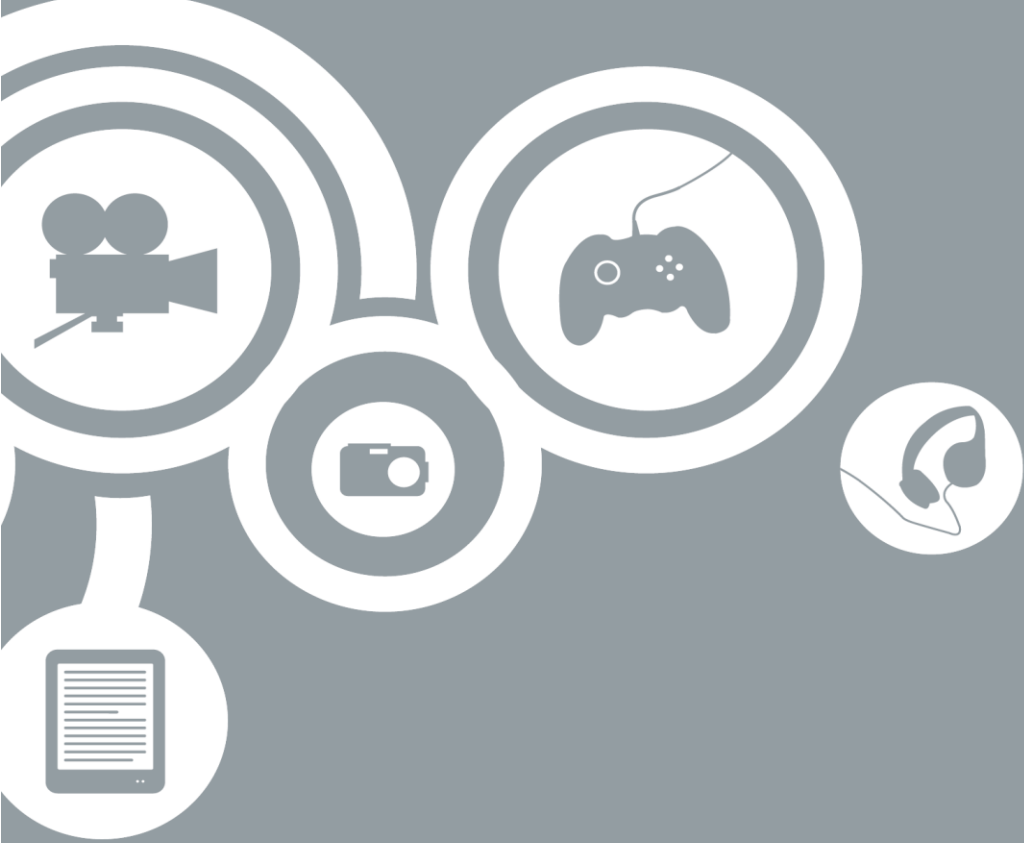
THE EXPECTATIONS

Hadopi

HETEROGENEOUS SITUATIONS AND POLICIES AMONG RIGHTHOLDERS

- The audiovisual and musical industries already use fingerprint systems extensively:

  - The **video sector** favors the blocking of protected content: reported issues relate to the modalities of supply of the fingerprints and the functioning of the tools, which remain at the discretion of the platforms.

  - The **musical sector** prefers to monetize content: reported issues mainly concern the unreliability of the reporting and the conditions of remuneration, which remain opaque.

- Other sectors such as **publishing, video games and photography** do not yet use content recognition technologies on a large scale. The reasons are numerous: lower economic weight, depth of the catalogue... In these sectors, there are needs for **expertise and consultation** in order to find out which tools should be implemented.

Hadopi

DISPARATE POLICIES AMONG PLATFORMS

- Some platforms such as YouTube, Facebook, Dailymotion have implemented fingerprint systems on a very large scale, **the effectiveness of which has been established**.

  - ➤ These solutions are **continuously improving**: anti-circumvention strategies, reduction of the duration limit under which a content can be recognised, recognition in real time (live content).

  - ➤ **Points of vigilance** raised by platforms: the number of existing management rules should not become overwhelming, the provision of fingerprints by rightholders should be increased and, in case of a dispute, platforms would rather like to have a limited role such as simply connecting rightholders and users.

- **Other sharing platforms still refrain from deploying content recognition tools** and continue to impose on rightholders cumbersome and inefficient notification procedures.

  - ➤ **Efforts are still to be made**: fortunately, content recognition solutions are being developed by third parties for platforms that cannot (or do not want to) develop tools in-house.

Hadopi

VIDEO MAKERS AND YOUTUBERS: A NEW TYPE OF ACTOR?

- Video makers and "YouTubers" have an **ambivalent status** with regard to content recognition technologies: initially considered as users, they have contributed to the success of platforms such as YouTube, and today they also tend to be regarded as rightholders.

  - ➢ On the one hand, they are calling for **more transparency** in the blocking and demonetization rules. They also want **copyright exceptions** to be more efficiently taken into account: film critiques and makers of parodies or documentaries using short excerpts still risk losing all revenues in case of a claim made by a rightholder.

  - ➢ On the other hand, they wish to be able to **benefit from tools allowing the control of their contents**: YouTube has for example gradually taken their expectations into account by providing them with a solution, called Copyright Match, which allows the protection of their contents but is however less sophisticated than Content ID.
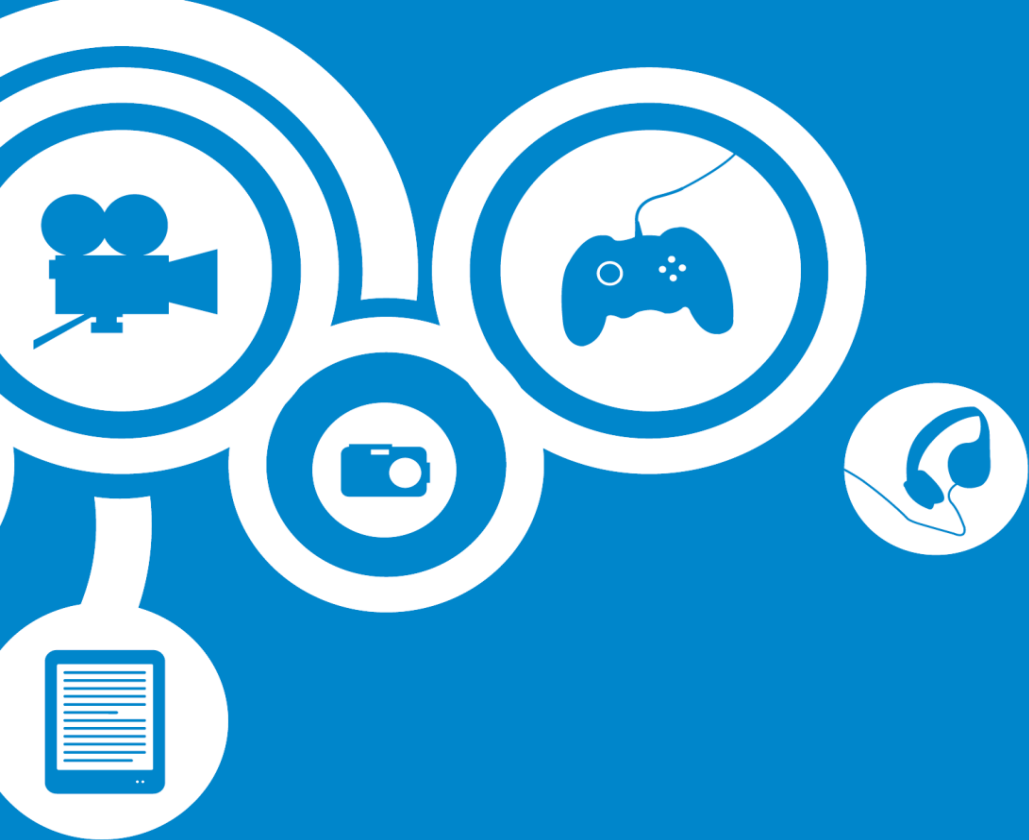
MAIN LEARNINGS WITH REGARD TO ARTICLE 17 OF THE DIRECTIVE

Hadopi

NEW BALANCES FOR MORE EFFECTIVE RIGHTS ON PLATFORMS

- Article 17, after being subject to many heated debates prior to its vote, actually **formalizes and perpetuates a pre-existing situation**. It will therefore allow better copyright protection but will also improve balances between protection and uses.

- Article 17 makes **existing citation and parody exceptions** compulsory and gives a central role to the **dispute settlement mechanism** (claims or blockings must be justified, disputes must be handled without undue delay and have to be controlled by a person).

- The new system **requires more transparency and calls for concerted actions and expertise** in order to define solutions that could correspond to the best efforts expected from platforms, based on the information that rightholders will have the duty to provide.

- At the European level, the European Commission is working on the **adoption of guidelines** that will provide recommendations for the application of this article in each EU country.

- In France, the new audiovisual law being discussed in 2020 has assigned a **central role to Arcom, a regulator** that will have to **analyse and assess existing content identification solutions** and that will be able to play a role in the context of **dispute resolution**.

WWW.HADOPI.FR

Hadopi