

Rights Data and Identifiers in the Text Publishing Industry

Piero Attanasio

30 March 2022

Copyright Infrastructure Webinars
WIPO

Summary

Few historical notes

The principles

The implementation

Focus on identifiers

Next steps

Conclusions

***The history: when
metadata and identifiers
crossed author's rights***

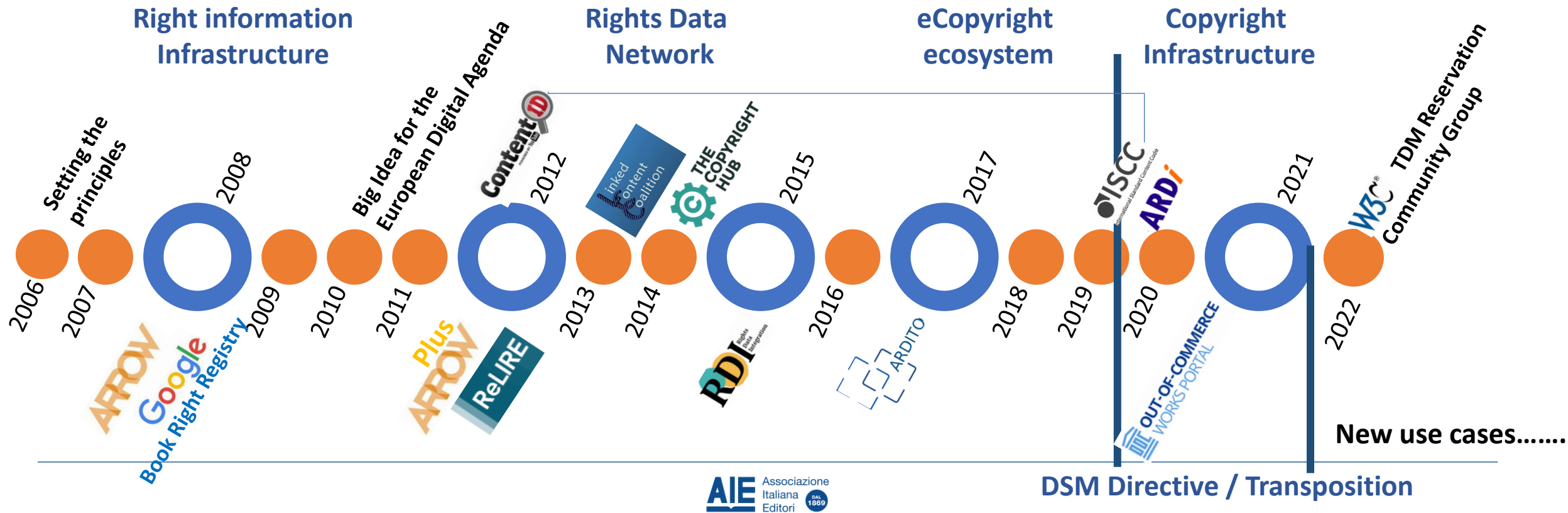
Lawyers vs. IT engineers in the book sector



The answer to the machine is in the machine

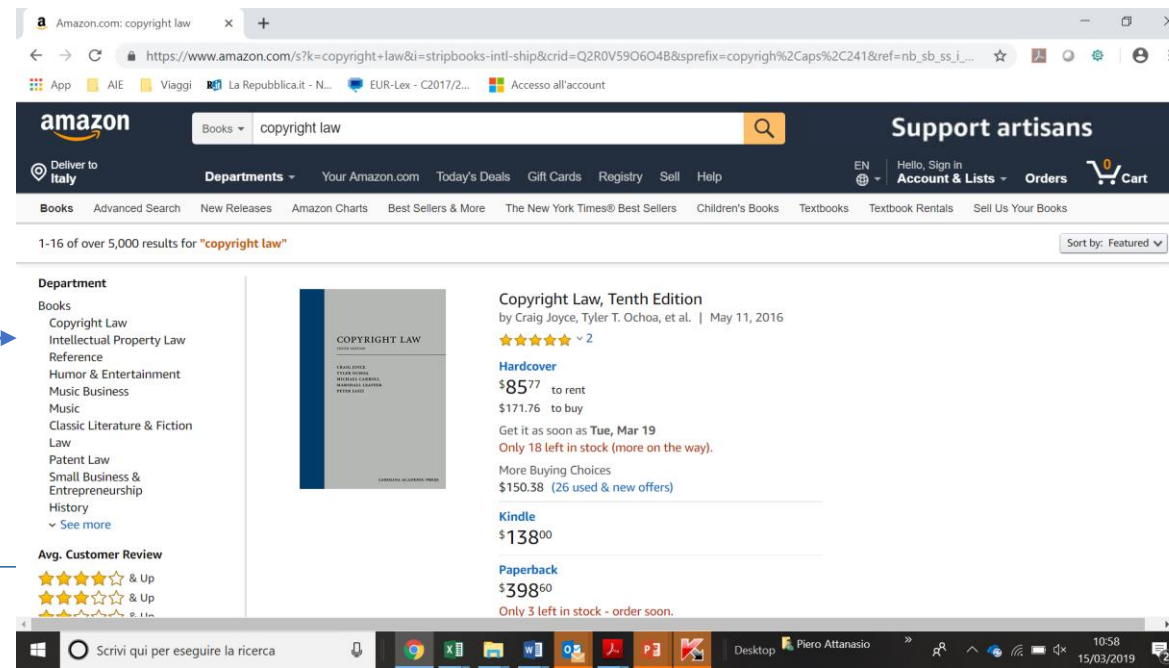
C. Clark, In *The Future of Copyright in the Digital Environment*
(ed. by P. Bernt Hugenholtz), 1996

Objective: Fill the gap in the digital content value network and connect online content to rights information, by building complementary digital rights data networks



The analogy: the data infrastructure in the book sector

E-commerce started from books because – in mid 90ies – an independent, open, standard based data-infrastructure existed only in the book sector



Amazon, as a new comer, was facilitated by the openness of the data infrastructure
We need similar principles for the rights infrastructure

Setting the principles



Rights data management should be

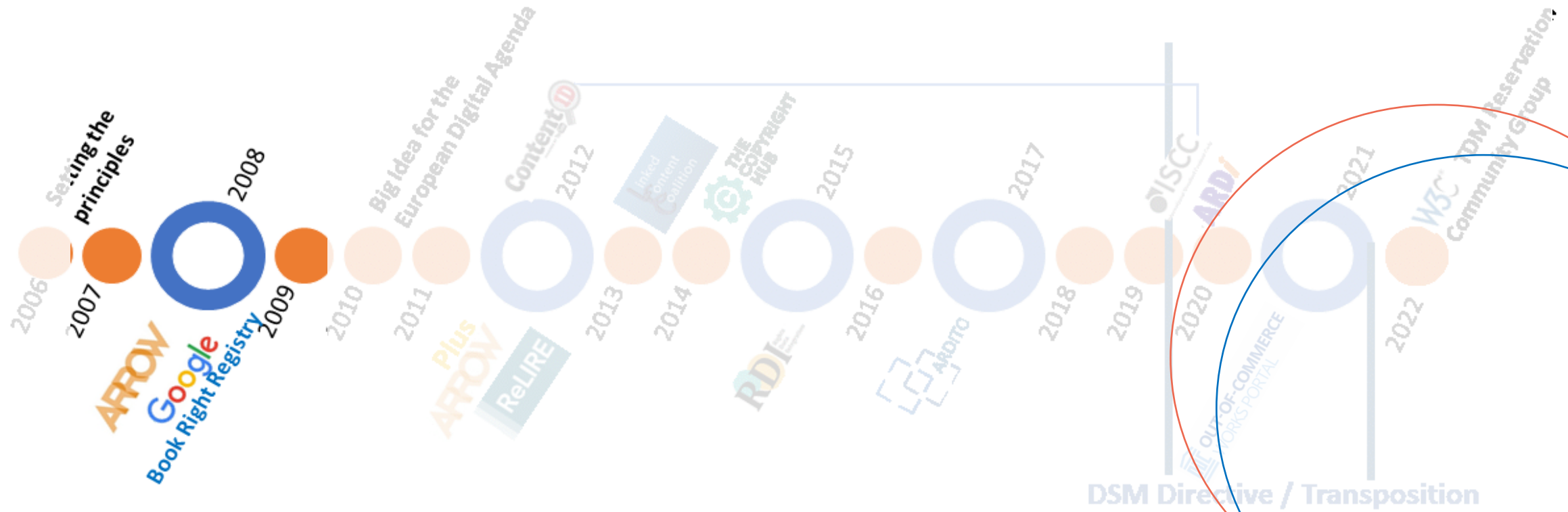
- 1) **Separate and independent** from rights management
- 2) **Based on standards** for identification, communication, description of all components of right transactions
- 3) **Based on neutral solutions** in respect to **business models**
commercial vs. not-for-profit,
collective vs. direct licencing,
subscription vs. pay per use vs. open access
and **content types**
text, image, audio-video, music, etc.
- 4) **Data should remain distributed**

*Coherence between
principles and
implementations*

Separate and independent rights data

The Arrow system and the Google BRR were both based on this principle:

- Data not controlled by the organisations managing the rights



Based on standards

Identifiers: The Content-Id experience (proprietary) and the need for standard solution: ISCC

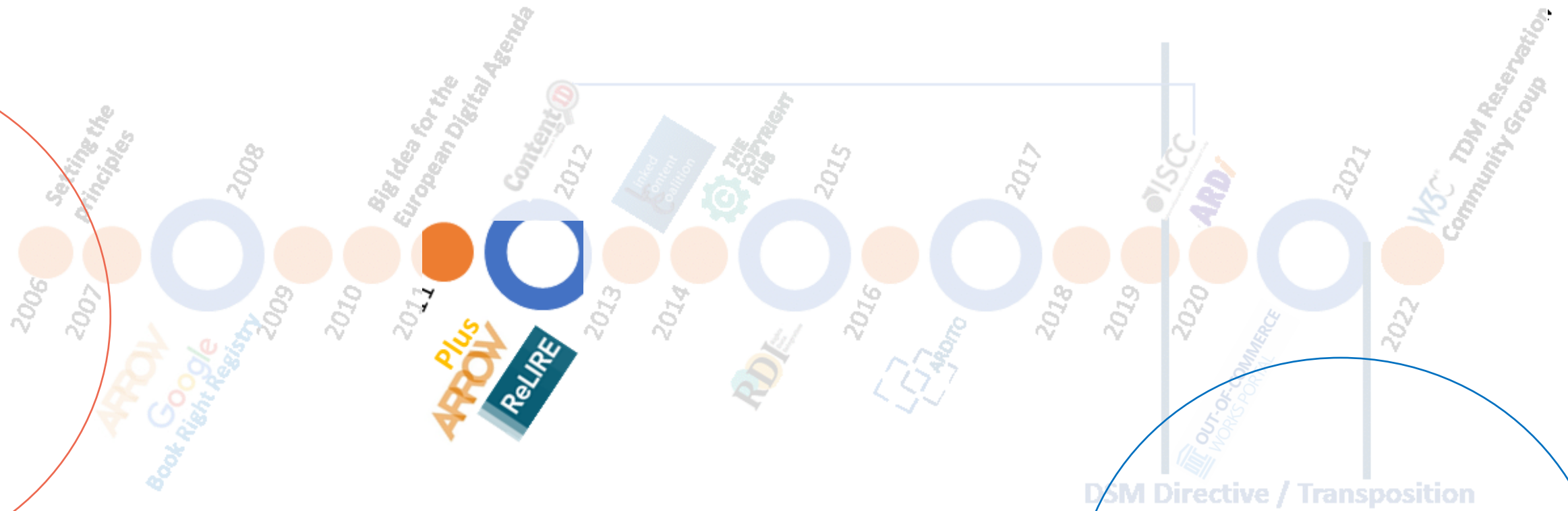
Metadata: LCC set a model

Communication: The TDM reservation protocol by the W3C working group



Solutions neutral to business models

Arrow Plus, serving digitisation programmes by libraries,
and **ReLiRe in France**, serving re-publication of out of commerce works by publishers
used the same approach

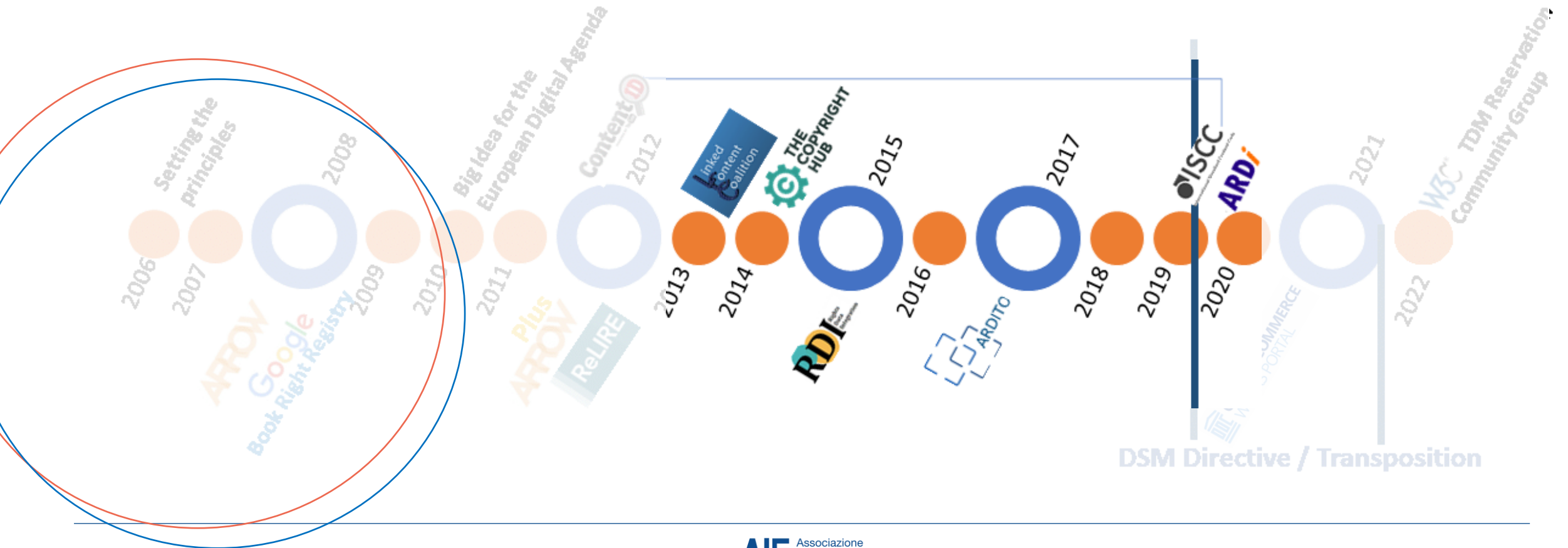


Solutions neutral to content genres

Methodology → Infrastructure → Demonstrators → Tools

LCC → Copyright Hub → RDI → Ardito / ISCC / ARDI

All developed for multiple content types



Distributed systems

BRR = centralised / Arrow = distributed

Content-Id = centralised / ISCC = distributed

Copyright Hub, RDI, Ardito, ARDi... = all distributed



Focus on identifiers

The situation in the publishing industry



Excellent in **product identifiers**



Advanced in **actionable identifiers**

i.e. identifiers associated with a “resolution mechanism” that allows linking the id with services



Absent in **work identifiers** (ISTC failed)

- Too focused on product trade (the ISTC as a mechanism to cluster the ISBN and sell books), rather than on rights data
- Lack of resolution mechanism enabling new services



Growing deployment of **name identifiers**



Emerging **rights declaration identifiers**



Interested in **content derived identifiers**

Not an alternative to the existing identifiers

International Standard Content Code

The reasons for the difference with other cultural sectors

Prevalence of direct rights management – Marginal role of collective management

Simplified supply chains in the licensing business and thus less need for interoperable work identifiers

Collective rights management traditionally linked to reuse of the «product» (not the work)

e.g. management of reprographic rights - the ISBN was all what RROs needed

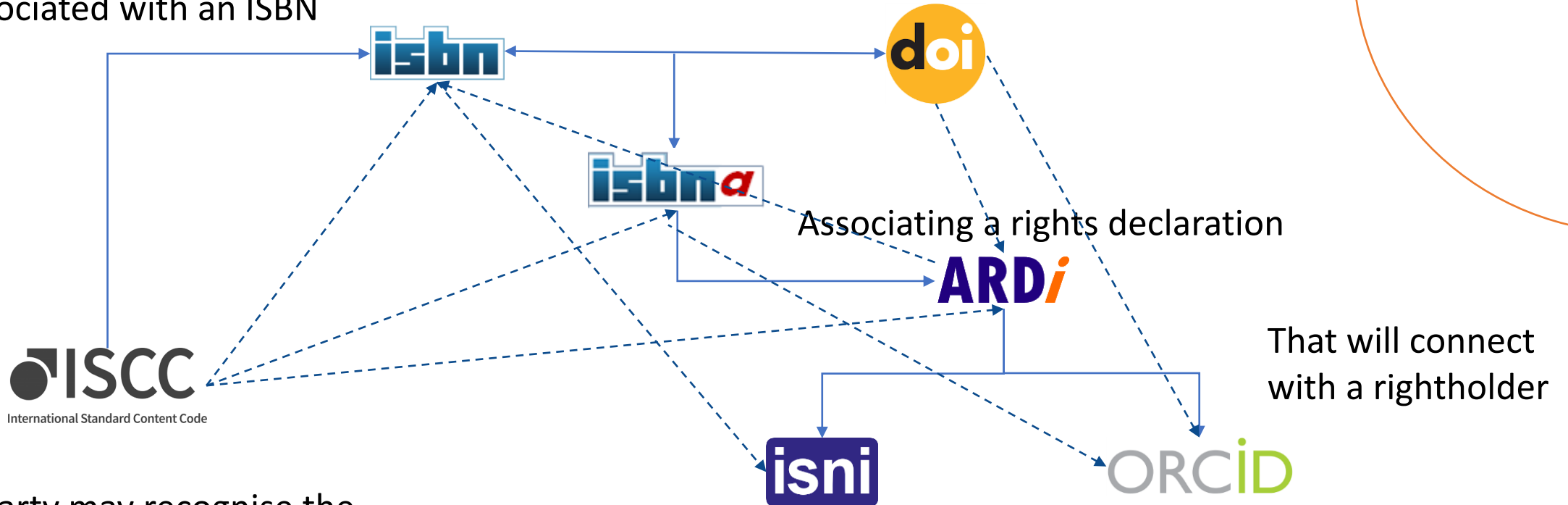
Importance of the back list, which makes it hard reshaping bibliographic databases

Data from Italy: 1,4M titles in commerce; 1,1M sold at least one copy in 2021

A vision for the future: the network of identifiers

Combining an ISBN with a DOI

Associated with an ISBN



A third party may recognise the content through a ISCC

Any other relation / link is possible

Next steps

From the digital society to the data society

Data determine market equilibrium in any business, including copyright licensing

“The quintessential task of many digital platforms is making predictions of various sorts (...) Data is the oil that powers these predictions” (Calvano and Polo, 2020).

“Platforms can use this information asymmetry to facilitate interaction and increase welfare for users. These data externalities attract users to the platform” (Martens, 2020)

“The collection and use of big user data enables [platforms] to continuously improve the quality of their offerings” (Fast et al., 2021), which “may result in monopolistic market power of platforms which they can use for their own benefit, at the expense of users” (Martens, 2020)

The European case: rights data needs in the DSM Directive

Art. 4: TDM exception: rightholders can reserve their rights

Rightholders to communicate with the miners, including through “machine readable means”

TDM Reservation Protocol in the **W3C Community**

Art. 5, par. 2: The “prevalence of licence”, when implemented, only if licences are “visible”

Rightholders to communicate with the educational establishments

Coexistence of direct and collective rights management

“Visibility services” *besides* right management

Art. 8-11: Complex data infrastructure to manage the new licensing mechanism for out of commerce works

Rightholders and Cultural Heritage Institutions to speak each other

Set up of the EU-IPO portal

Services required for the correct determination of the out-of-commerce status

Art. 17: Licensing works to online content-sharing service providers or Notice and Stay Down mechanism

Rightholders to communicate with platforms, sending “sufficiently substantiated notice” about their works

ISO standardisation process for the ISCC and promising applicability for this purpose

Conclusions

Conclusions: towards an *Open Copyright Data* Infrastructure

A definition from the book publishing industry: the copyright infrastructure is:

**A set of technologies, open standards, tools and services to
connect
content identifiers with rights data declarations
using different methodologies**

Tools may encompass web resolution systems (DOI, Handle, URN...), Copyright hub or alike, etc.
It is not a set of authoritative registries to solve the attribution problem
It is not the place where right negotiations happen: it just deals with data
It should be *open* to guarantee^(*) an open publishing market

^(*) Open ≠ for free

Thank you

Piero Attanasio
piero.attanasio@aie.it

AIE Associazione
Italiana Editori 