Social Media Copyright Management using Semantic Web and Blockchain

Roberto García
Computer Science and Eng. Department
Universitat de Lleida
Jaume II, 69
25001 Lleida, Spain
rgarcia@diei.udl.cat

Rosa Gil
Computer Science and Eng. Department
Universitat de Lleida
Jaume II, 69
25001 Lleida, Spain
rgil@diei.udl.cat

ABSTRACT

Solutions based on distributed ledgers require sophisticated tools for data modelling and integration that can be overcome using semantic and Linked Data technologies. One example is copyright management, where we attempt to adapt the Copyright Ontology so it can be used to build applications that benefit from both worlds, rich information modelling and reasoning together with immutable and accountable information storage that provides trust and confidence on the modelled rights statements. This approach has been applied in the context of an application for the management of social media re-use for journalistic purposes.

CCS CONCEPTS

• **Networks** → Social media networks • **Information systems** → Semantic web description languages • **Social and professional topics** → Copyrights; Centralization / decentralization

KEYWORDS

Copyright, Ontology, Rights Expression Language, Blockchain, Distributed Ledger, Ethereum, Semantic Web, Linked Data

ACM Reference format:

Roberto García and Rosa Gil. 2019. Social Media Copyright Management using Semantic Web and Blockchain. In *Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services (iiWAS2019). ACM, New York, NY, USA, 5 pages.* https://doi.org/10.1145/3366030.3366128

1 Introduction

Distributed Ledger technologies are being applied to almost any conceivable domain, from logistics [5] to renewable energy production management [10]. In all cases, in addition to the technologies specific to the distributed ledger itself, developers face issues related to information management and integration similar to those faced by Web applications. In this regard, Semantic Web technologies are well-positioned to solve this kind of issues, and have been even used to model the key concepts in one of the main distributed ledgers, Ethereum's EthOn [12].

This paper presents our vision and a first attempt to adapt an existing proposal based on Semantic Web technologies for the copyright management domain, the Copyright Ontology [6], to the distributed ledger space, particularly Ethereum. The approach is based on a revision of the Copyright Ontology that moves its foundations to the schema.org vocabulary [8]. The idea is that this facilitates the adoption of the Copyright Ontology thanks to the ampler use of schema.org. Additionally, to tie it better to copyright management solutions based on Distributed Ledger technologies, the Copyright Ontology is also connected to EthOn.

The rest of the paper is presented as follows, Section 2 presents the Copyright Ontology focusing on its 3 subparts though the corresponding subsections. Then, Section 3 presents a preliminary implementation of an architecture combining the Copyright Ontology, Semantic Web technologies and blockchain. This implementation addresses the management of social media re-use for journalistic purposes. Finally, Section 4 presents the conclusions and future work.

2 Copyright Ontology

The copyright domain is quite complex, so the Copyright Ontology attempts its conceptualisation in three steps, which allows facing this process in an incremental way.

First, the objective is the more primitive part, the Creation Model, detailed in Section 2.1. Second, there is the model for the rights part, the Rights Model described in Section 2.2, and finally a model for the available actions, the Action Model, which is built on top of the two previous ones and presented in Section 2.3.

The final objective is to develop a computerized model of the copyright domain that facilitates the automation of rights management applications [6].

2.1 Creation Model

The objective of this part of the Copyright Ontology is to model the different forms that a creation, the subject matter of copyright, can take along its life cycle. The resulting model should be capable of

iiWAS2019, December 2–4, 2019, Munich, Germany

ACM ISBN 978-1-4503-7179-7/19/12…$15.00
https://doi.org/10.1145/3366030.3366128

modelling complex lifecycles like the one presented in Table 1. An overview of the sample lifecycle in Table 1 from the point of view of the different forms the original creation acquires through it is shown in Figure 1.

Table 1: Creation lifecycle sample: broadcast of a serial adapted from a literary work

a) A creator adapts the original **literary work** Alexandre Dumas' "The Count of Monte Cristo", to produce a **serial**.
b) The resulting adaptation to a serial is realised as a **script** that is **performed** by some actors, including Gerard Depardieu, and recorded into a **motion picture**.
c) This motion picture is finally transmitted to users who can tune the resulting **broadcast**.

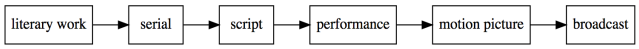

Figure 1: Overview of the creation lifecycle sample in Table 1.

To capture all these different "shapes" creations take along their life-cycle, a set of concepts are proposed and framed in the context of more generic schema.org [8] conceptualization:

- *Schema.org Intangible*[1]
  - **Work**: is a kind of intangible that represents what we recognise, from the commonality of content between and among the various expressions of the work [9], as an original creation. It includes literary and artistic works, music, pictures and motion pictures, but also computer programs or compilations, like databases.
- *Schema.org CreativeWork*[2]
  - **Manifestation**: the materialisation of a Work in a concrete medium, a tangible or digital object.
  - **Recording**: the materialisation of a Performance in a concrete medium, a tangible or digital object.
    - Includes **Sound Recording** and **Audiovisual Recording**.
  - **Instance**: the reproduction, i.e. copy, of a Manifestation, a Recording or another Instance.
- *Schema.org Event*[3]
  - **Performance**: the expression in time of a Work. Performers or technical methods might be involved in the process.
    - **Improvisation**: the expression in time of a Work that has not been previously materialised in a Manifestation.
  - **Communication**: the transmission of a Work among places at a given time. It is a process performed when the public is not present at the place and or time where the communication originates. It includes broadcasts, i.e. one to many, but also communications from a place and at a time individually chosen, like Internet streaming.
    - **Live Communication**: the direct communication of a performance without a mediating non-ephemeral recording.

The previous concepts are the building blocks of the Copyright Ontology Creation Model. However, to capture creation value chains, the model should also include how these different creation shapes relate among them.

Figure 2 presents these relationships that capture the flux from **Work**, towards something that can be consumed by end-users, e.g. an **Instance**, a **Communication** or a **Performance**, despite **Recordings** or **Manifestations** might also be consumed, e.g. a painting.

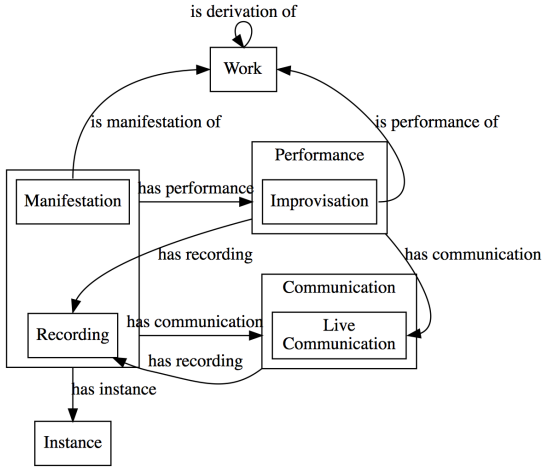

Figure 2: Copyright Ontology Creation Model

With this model, it is possible to represent the value chain scenario presented in Table 1, as shown in Figure 3.

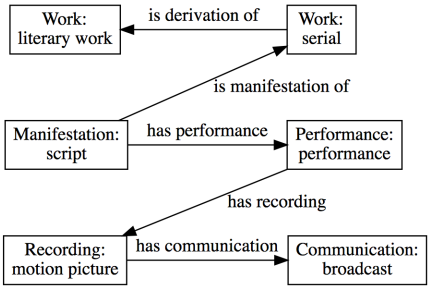

Figure 3: Overview of the creation lifecycle sample in Table 1

2.2 Actions Model

Once the Creation Model has been established, the foundations for representing creation value chains are laid. However, it is necessary also to capture the dynamics that make creations move from one stage to the next along these chains. The Action Model captures the dynamic part of the Copyright Ontology, the actions performed by

[1] https://schema.org/Intangible
[2] https://schema.org/CreativeWork
[3] https://schema.org/Event

creations value chains' actors. Actions are the main building-block of the Copyright Ontology, in the sense that usually all representations based on this ontology will be built around one or more actions that connect all the parties and entities participating or involved in them.

This "Action-Oriented Conceptual Modelling" [1] approach has been followed since early versions of the Copyright Ontology [6]. The importance of Actions for knowledge representation is also highlighted by other initiatives, like schema.org that includes them as part of their proposed vocabulary [4]. This has facilitated the integration of the Copyright Ontology with schema.org, as detailed next.

The most relevant set of actions in the Copyright Ontology includes those that "move" creations across their different "statuses" identified in the Creation Model. For instance, making the abstract idea represented by a Work perceivable as a Manifestation involves the *Manifest* action. The most relevant actions in this set are presented in Figure 4 as arrows that link the original and resulting creation "statuses" and illustrate the connection between the Creation and Actions Models. Other examples are the *Perform* action to generate a Performance from a Manifestation or the *Record* one to fix a Performance into a Recording.

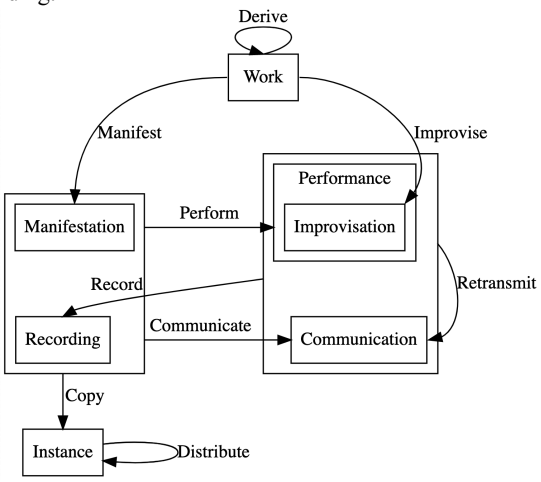

Figure 4: Copyright Ontology actions in the context of the Creation Model.

These actions are also linked to the rights identified in the Rights Model that govern them, as detailed in the following subsection. In addition to this set of actions that are modelled to capture what rights or licenses authorise or prohibit (like *Copy* or *Perform*), there are additional actions to manage how the previous actions (or the related rights) are granted (like *Agree* or *Disagree*) or what is asked in return (like *Attribute* or *Pay*).

In all cases, actions share a common set of relationships with other entities participating in them. These relationships, adopted from linguistics, are called case roles or semantic roles [13]. Semantic roles in linguistics link the different parts of a sentence to its main component, the verb, which usually represents an action.

These same roles are also used in other ontologies and vocabularies like schema.org [4]. In fact, all actions in the Copyright Ontology are particularisations of schema.org's Action and in some cases have direct equivalents in schema.org.

The set of semantic roles is presented in Table 2, organised in sets related to the broad questions that can be made about actions (like Who? What? When?...) and linked to schema.org when the role is reused from that vocabulary.

Table 2: Semantic roles relating actions to the entities participating in them (those starting with "schema:" correspond to schema.org)

Role Kind	Main Role	Description
who	schema:agent	The direct performer or driver of the action
	schema:participant	Other co-agents that participated in the action indirectly, for instance, a recipient
what	schema:object	The object upon which the action is carried out
	schema:result	The result produced in the action
where	schema:location	Where an action takes place
	schema:fromLocation	The original location of the object or the agent before the action
	schema:toLocation	The final location of the object or the agent after the action
when	schema:startTime	When the action started or the time it is expected to start
	schema:endTime	When the action finished or the time it is expected to end
	pointInTime	The point in time when the action happens
	duration	The amount of time the action requires to complete
with	schema:instrument	The object that helps the agent perform the action
why	aim	The reason or objective of the action
how	manner	The way the action is carried out
if	condition	Something that must hold or happen before the action starts
then	consequence	Something that must hold or happen after the action is completed

For instance, for a particular performance as the one shown in Figure 5, it is possible to capture the details of the *Perform* action using some of the semantic roles presented in Table 2.

As mentioned, actions are the key building blocks when modelling statements about rights using the Copyright Ontology. For instance, a *Manifest* action gives birth to the rights associated with a new creation. These events should be recorded and benefit from distributed ledgers as they provide a way to keep track of them in a secure and auditable way. Consequently, the Copyright Ontology, and particularly actions and the entities participating in them, are also connected to distributed ledger concepts, particularly to Ethereum concepts through the EthOn ontology [12].

Concretely, copyright actions, when registered in the Ethreum blockchain, will be captured through instances of the EthOn class contracts:Event. For instance, *Manifest* actions are captured by a subclass of the previous class called ManifestEvent. On the other hand, instances of the class ExternalAgent defined by EthOn, an agent (person, group, software or physical artefact [7]) that controls

an Ethereum account, can be used to model the actors involved in these copyright actions.

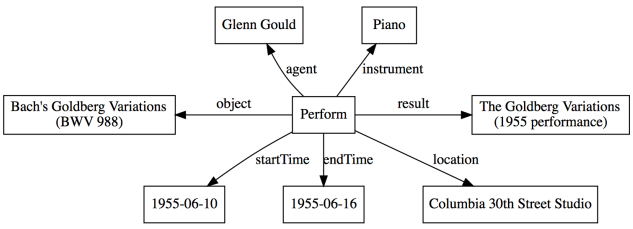

Figure 5: Perform action model using semantic roles

2.3 Rights Model

This is the last part of the Copyright Ontology, build on the foundations provided by the other two parts, the Creation Model and the Actions Model. Rights constraint what actions are favoured or restricted based on different legal regimes. Thus, they restrict how value chains are built by moving creations among them, encouraging certain behaviours and discouraging others, for instance, punishing them.

Though there is not a global copyright framework that uniformly applies worldwide, there have been many attempts to harmonise the plethora of country-level regulations into a global system through the World Intellectual Property Organisation [14].

The Copyright Ontology captures this overarching set of rights in its Rights Model. It includes economic plus moral rights, as promoted by WIPO, and copyright-related rights. The following are the most relevant rights in the context of copyright value chain, those related to the production and commercial aspects:

- **Reproduction Right**: regulates *Copy* actions that produce replicas of a given object, i.e. Instances as defined in the Creation Model. Examples of reproduction are the mass production of CD copies from an audio recording master, to scan a book to produce a digitalisation of it or to download a digital file into the local hard disk.
- **Distribution Right**: governs *Distribute* actions geared towards providing users with previously made copies incorporated into tangible products. The ownership of the corresponding physical support can be transferred permanently, i.e. the act is *Sell*, or just temporally, i.e. *Rent* if there is a significant economic compensation or *Lend* if not.
- **Public Performance Right**: controls *Perform* actions of Works when they are made in public, i.e. before an audience.
- **Fixation Right**: regulates the materialisation of a Performance into a fixation object through a *Record* action. Common kinds of Recording are Sound and Audiovisual Recording, which are governed by the corresponding **Sound Recording Right** and **Motion Picture Right**.
- **Communication Right**: governs the realisation of *Communicate* actions of works, including wire or wireless means and those realised from a place and at a time individually selected. This right is concretised into **Broadcasting Right**, when the communication is simultaneously made to a plurality of people as a *Broadcast*, and **Making Available Right**, when the communication is individually chosen like video-on-demand streaming through a *Make Available* action.
- **Transformation Right**: controls the Derive action that generates new works from previously existing ones. The results of this kind of action are considered new works, and not mere reproductions because they contribute something new. This right is concretised into the **Adaptation Right** and the **Translation Right**. *Adapt* creates a new work of a different type than the original one, e.g. a film from a novel. *Translate* generates a work of the same type but in a different language.

These rights are conceded to the author or promoter of the creation by the mere act of bringing the work into existence. Consequently, whenever an author generates a new creation, the author automatically gets all Economic Rights, and also the Moral Rights depending on the jurisdiction. This fact can also be captured using the Copyright Ontology as detailed in the Actions Model. From this initial situation, it is possible to *Transfer*, or at least license, the Economic Rights to third parties giving rise to content value chains, as also detailed in the Actions Model. Consequently, the registration of this event in a distributed ledger becomes a pivotal point in the integration with this kind of technologies, as detailed in the next section.

3 Semantic Web and Blockchain for Copyright Management

The Copyright Ontology, through the combination of Semantic Web and blockchain technologies, has been applied during the development of the Rights Management module in the context of the EU H2020 project InVID[4], which focuses on the verification of social media video content for the news industry [11].

The Rights Management module is structured as detailed in Figure 6. For the client-side there is mainly the **Rights Management App**, a Web application for journalists and social media content owners. After journalists have verified the trustfulness of the content with journalistic interest, they can use this application to prepare a reuse request addressed to the content uploader. The uploader also uses this application to assert if she is also the content owner. In this case, she can inspect the reuse request, accept, reject or negotiate its conditions. Additionally, users interact with the **uPort App**, a mobile application developed by uPort[5] and available for Android and iOS that helps users generate and manage their identities, required to use the blockchain capabilities as detailed next.

On the server-side, the main component is the **Rights Management Services**, which provides the RESTful API that facilitates the integration of the Rights Management module with the rest of the InVID Platform. There is also the **Identity Provider**, which centralises authentication of InVID users and leverages the **Identity Attestation** service that provides attestations compatible

[4] https://www.invid-project.eu

[5] https://www.uport.me

with the uPort application about platforms users identity. Attestations about users' e-mails and social networks memberships (Google, Facebook and Twitter) are stored in the users' devices thanks to the uPort application. This way, they can be managed by the user themselves in a self-sovereign way, and linked to their Ethereum accounts.

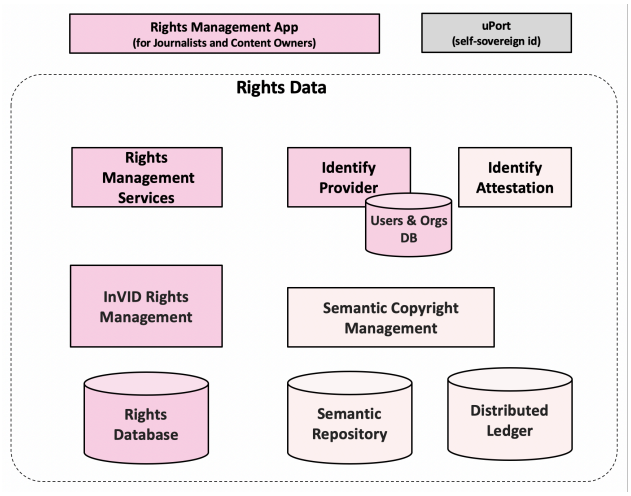

Figure 6: InVID Rights Management Module architecture

Beyond the previous components providing the external APIs, there is the **InVID Rights Management** component that implements most of the services and coordinates the rest of the module components. Its state is kept by the **Rights Database**, though copyright-related expressions, like re-use requests or re-use agreements, are modelled using the Copyright Ontology and stored though the module's **Semantic Copyright Management** services into a **Semantic Repository**. In combination, they also provide the reasoning capabilities that facilitate checking if the existing agreements already authorise an intended re-use.

Finally, there is the **Distributed Ledger** component implemented using an Ethereum Smart Contract [2]. From re-use request creation, through all negotiation steps and until an agreement is reached, if it is so, all these steps are recorded in the Ethereum blockchain through a smart contract. These blockchain transactions are signed by the involved parties using the accounts provided by uPort, which are linked to their identities through e-mail and social network membership attestations.

Due to storage costs of most blockchains like Ethereum, the representations of the agreements are actually stored in the IPFS peer-to-peer distributed and immutable file system [3]. The semantic representations of the agreements, based on the Copyright Ontology *Agree* action, detail the re-use conditions using other actions and semantic roles, like *Make Available* or *where*, and are serialised as JSON-LD[6]. Then, what is stored in the blockchain is the hash of this representation. This way, re-use agreement terms are immutably stored and time-stamped, so it is possible later to prove the existence of these agreements, the agreed terms and the identities of the involved parties. The whole process is illustrated in the Rights Module blockchain online guide[7].

4 Conclusions and Future Work

The Copyright Ontology, combined with semantic technologies, provides a formal representation of key copyright concepts that can be used to enable rights management through distributed ledgers and blockchain technologies. This approach has been validated in the context of the EU H2020 InVID project, where the Ethereum blockchain is used as a way of keeping track and providing proof about re-use agreements between journalists and content owners of social media videos with journalistic interest.

Future work focuses on deploying a completely decentralised application for copyright management beyond the particular scenario addressed in InVID. Thus, it will not just address re-use agreements, but the whole value chain of copyright actions defined in the Copyright Ontology. Moreover, the objective is that users directly interact with the blockchain from their browsers, avoiding the centralised intermediary API currently used in InVID. In this case, semantic reasoning capabilities will be provided by independent software agents listening to copyright events and reconstructing the semantic data from the corresponding IPFS serialisations.

ACKNOWLEDGMENTS

This work was partially supported by the EU H2020 project "InVID - In Video Veritas, Verification of Social Media Video Content for the News Industry" (687786).

REFERENCES

[1] Ågerfalk, P.J. and Eriksson, O. 2004. Action-oriented conceptual modelling. *European Journal of Information Systems*. 13, 1 (2004), 80–92.
[2] Antonopoulos, A.M. and Wood, G. 2018. *Mastering Ethereum: Building Smart Contracts and DApps*. O'Reilly Media.
[3] Benet, J. 2014. IPFS - Content Addressed, Versioned, P2P File System. *arXiv:1407.3561 [cs]*. (Jul. 2014).
[4] Brickley, D. 2014. Announcing Schema.org Actions. *Official blog for schema.org*.
[5] Dobrovnik, M. et al. 2018. Blockchain for and in Logistics: What to Adopt and Where to Start. *Logistics*. 2, 3 (2018), 18.
[6] García, R. 2010. *A Semantic Web Approach to Digital Rights Management*. VDM Verlag.
[7] Graves, M. et al. 2007. FOAF: Connecting People on the Semantic Web. *Cataloging & Classification Quarterly*. 43, 3–4 (Apr. 2007), 191–202. DOI:https://doi.org/10.1300/J104v43n03_10.
[8] Guha, R.V. et al. 2016. Schema.org: Evolution of Structured Data on the Web. *Communications of the ACM*. 59, 2 (2016), 44–51. DOI:https://doi.org/10.1145/2844544.
[9] IFLA Study Group on the Functional Requirements for Bibliographic Records ed. 2013. *Functional Requirements for Bibliographic Records, Final Report*. De Gruyter Saur.
[10] Mengelkamp, E. et al. 2018. A blockchain-based smart grid: towards sustainable local energy markets. *Computer Science-Research and Development*. 33, 1–2 (2018), 207–214.
[11] Mezaris, V. et al. eds. 2019. *Video Verification in the Fake News Era*. Springer International Publishing.
[12] Pfeffer, J. 2018. *EthOn-an Ethereum ontology*. Technical Report 0.2. Consensys.
[13] Sowa, J.F. 1999. *Knowledge Representation. Logical, philosophical and computational foundations*. Brooks Cole Publishing Co.
[14] WIPO 2004. *WIPO Intellectual Property Handbook: Policy, Law and Use*. WIPO Publications.

[6] https://json-ld.org

[7] https://rights.invid.udl.cat/blockchain/guide