UNIVERSITY OF GOTHENBURG

Exploring Challenges in Embedding Metadata of Licence Information in Digital Work

Bachelor of Science Thesis in the Programme Software Engineering and Management

HAOQING YINHE
JING LIU

Exploring Challenges in Embedding Metadata of Licence Information in Digital Work

HAOQING YINHE
JING LIU

Exploring Challenges in Embedding Metadata of Licence Information in Digital Work

Haoqing Yinhe, Jing Liu

Department of Computer Science and Engineering
University of Gothenburg
Gothenburg, Sweden
gusyinha@student.gu.se
gusjingli@student.gu.se

Abstract: Recently, there has been an increasing trend in online digital content sharing. The attribution of copyright has become an important aspect of creative work production. However, there has been little research regarding identifying and using the licence information. This paper presents a qualitative case study that explored the technological and social challenges of developing an end-user tool that embeds licence information metadata in digital work. The study also sought suggestions from experts and end users to address the challenges. The case study mainly investigated the area through qualitative interviewing. A preliminary prototype of the end-user tool has been developed in order to validate the suggestions.

Keywords: Licence, embed metadata, attribution, digital work, Digital Rights Management, Rights Expression Language

1. Introduction

Nowadays technologies allow us to easily obtain and share a piece of creative digital work through various applications. While a lot of people who publish digital work online are happy to see it used in new and creative ways, there is a need for users to associate the work with the original creator. Such act of associating is called the attribution of copyright. Most of the time, the users are required to attribute and acknowledge the original creator, when a work is being used or appears in another work, depending on the specific permissions granted by the creator.

Licences are known to be a standard legal means that allows creators to communicate access and usage permissions information to the users. Identifying rights holders is an important aspect of licences. For example, Creative Commons releases licences that focus on ensuring attribution of

the rights holders. Knowledge of licences has great importance for enhancing attribution in online digital materials.

A long-standing problem has been to identify and to keep track of the licence information metadata, when one wishes to use or to redistribute a piece of obtained digital work. For example, one has to keep track of information such as:

1. "Which author is it?"
2. "How to attribute this author?"
3. "Under which licence was this work released?"

A considerable amount of research has been focusing on how to interpret the licence information metadata. For example, the Rights Expression Languages (REL) has received considerable attention over the last decade. The REL is originally designed to express the access and usage rights for digital materials. It is able to express the licence information and attribution information in machine-readable metadata.

However, less attention has been paid to the process during which the metadata is handled and used. A standard procedure for users to keep track of the licences information and attribution information has been creating a database for storing these metadata. The question remains: is it possible to embed the metadata into the digital materials, and can it enhance attribution?

The purpose of this paper is to explore the challenges in developing an end-user tool that can embed and retrieve the machine-interpretable licence information metadata in digital work.

The research question of this study is:

● What are the technological and social challenges of building an end-user tool, which can embed metadata of licence information in digital work? What are the suggestions to address these challenges?

This paper is structured as follows: Section 2 presents the theoretical background of the study; Section 3 explains the research methodology; Section 4 presents the results of this research; Section 5 includes a discussion of the results; Section 6 presents the recommendation of this paper, and Section 7 concludes the research.

2. Theoretical Background

2.1 Licence

A licence for digital work is an agreement issued by a rights owner to the prospective user of the digital content. It is a typical aggregation of one or more subsets of rights statements, i.e. a set of expressions of certain permissions, constraints, and obligations about the rights holder and the content. Based on Abelson et. al. (2008), typically a licence could specify issues such as: whether the licensed item can be redistributed or modified, and how to credit the original creator of the licensed item. A licence provides a standard means to express right statements. Attribution is an important aspect in many types of licences.

Over the last decade, there have been two main approaches [Ab08] to utilize licences for digital work. One approach was creating widely applicable licences in clearly human-readable forms. The other approach was increasing the reusability of licences, making it easy to be detected and interpreted by machines.

2.2 Rights Expression Language (REL)

A Rights Expression Language (REL) is a standardized machine-interpretable form of expressing usage and access rights to assets into metadata [Gu03]. RELs were developed for computer programs to read and interpret under what licence the work has been released, whether the work can be redistributed or modified, and how the original author should be credited for the work. Most of the RELs are based on the syntax of Extensible Markup Language (XML), and are able to handle flexible expressions, from formulating complex contracts to stand-alone rights expressions [Gu03]. There are a number of standards developed by various communities, and among the existing RELs, the Open Digital Rights Language (ODRL), eXtensible Rights Markup Language (XrML) and the Creative Commons Rights Expression Language (ccREL) are commonly used.

The REL provides the rights information about the content as metadata [Ia04]. For example, ODRL covers metadata about three core entities and their relationships: Users, Content, and Rights. Users can be any type of user, from a rights holder to an end-consumer. Content is any type of content at any level of aggregation. The Rights entity is an expression of the permissions, constraints, and obligations between the Users and the Content [Ia01]. The ability to express these three core entities makes REL a fundamental part of any DRM (Digital Rights Management) system.

According to Iannella (2001) and Gunter et. al. (2001), RELs are capable of modelling and expressing a wide range of complex rights statements due to its flexibility, such as usage permission, playback count constraints, time constraints, payment obligation. The example of application can be payment, time and date, identity of the user, membership in a club, identity of the device, return of usage reports, recent contact with a clearinghouse, etc. [GWW01]

2.3 Digital Rights Management (DRM)

Digital Rights Management (DRM) is a broad area that focuses on enforcing security and encryption during the creation, distribution, and usage of the digital content [Ia01]. Traditional DRM systems limit the distribution to only those who pay. Over the years, DRM systems also begin to utilise metadata to cover the description, identification, trading, protection, monitoring and tracking of all forms of rights usages [Ia01].

Digital Rights Enforcement (DRE) is the "public face" of what DRM exposed to the majority of end users [IH03]. It enforces the encryption and security of access. Iannella and Higgs (2003) have a good explanation of DRE, whose purpose is to ensure that the content is only used for the acquired terms and conditions. Traditional DRE technologies embed rights information within the content, for example usage agreements including the identity of party that downloads the content, or invisible watermarks on each page of a document [IH03].

There have been DRE solutions such as a secure digital container [Gu03], i.e. an application that encrypts the work together with its rights information and renders the contents only for authorized users. Microsoft's WMA format was known to implement this solution [Gu03]. But such approach limits access to the digital work.

3. Methodology

Case study methodology with exploratory purposes [Ro02] is applied in this research. The case study will take an interpretive perspective that attempts to understand phenomena through the informants' interpretation [KM99].

3.1 Research Sites

The research is done in cooperation with Commons Machinery AB in Sweden where the data sources are provided.

3.2 Data Sources

The case study was conducted with two groups of participants, the **expert group** and the **potential end user group**. There were three participants from the expert group: the first participant has a leader role in the ODRL Community Group [1], the second participant had been significantly engaged in projects that handled metadata for educational purposes in Creative Commons [2] and the third person had been engaged in metadata handling for the Europeana [3] project. Two professional photographers are in the potential end user group because they deal with digital work on almost daily basis.

3.3 Data Collection

The data collection followed the guide as demonstrated by Runeson and Höst (2009). A number of open-ended interviews were conducted in order to collect raw data for the research objective. The aforementioned five participants of two groups in the case study were the informants. Two sets of interview questions were designed, one for the experts and one for the potential end users. The interview questions were designed based on the semi-structured format where the researchers prepared some questions beforehand and there was space for improvisation [MN07].

All three interviews with informants from the expert group were video interviews conducted over the Internet due to geographical constraints. For the potential end user group, the researchers conducted one video interview over the Internet, and one face-to-face interview. The interviews ranged from 30 min to 50 min in length and all of them were recorded and transcribed. The interview text served as the primary data source during analysis.

3.4 Data Analysis

We conducted the data analysis using the qualitative data analysis method as demonstrated by Runeson and Höst (2009). A set of raw data was transcribed into text from the sound recordings from the interviews. These raw data were coded into grouped themes. The coding analysis mainly targeted identifying the following:

---

[1] ODRL Community Group http://www.w3.org/community/odrl/

[2] Creative Commons http://creativecommons.org/about/

[3] Europeana http://pro.europeana.eu/about/

- The potential challenges in embedding metadata of licence information in digital work from the experts' perspectives.
- The practical problems encountered in the current way in which attribution is handled from the potential end users' perspectives.
- The suggestions to address such potential challenges and practical problems.

After the challenges, problems and corresponding suggestions were properly identified, the outcomes were formalized into a list of software requirements for the implementation of the end-user tool. In the scope of this research, a preliminary prototype was developed. The prototype fulfilled part of the most significant requirements and served the purpose for validating suggestions.

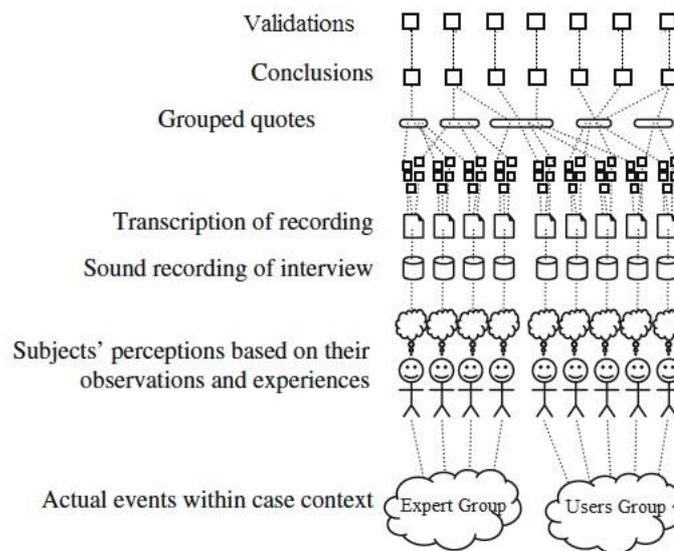The picture below illustrates the process of data collection and analysis in our research [RH09].

Fig. 3.1 Modified version of data collection and analysis process [RH09]

3.5 Strategies for Validating Findings

Qualitative Validity

In order to ensure the qualitative validity, we took the following approaches during the period of coding:

- When coding from transcripts, a theme was only established if same traits of information appeared in different sources. We triangulated different data sources of information.
- Detailed descriptions were provided for the findings to avoid misinterpretation caused by too few words or by narrow perspectives.
- We tried to reflect the reality from multiple perspectives and did not avoid negative or discrepant information about the study.

Qualitative Reliability

According to Creswell (2009), we set up four checks to ensure the consistency of coding between different researchers.

- Checked and made sure that the transcripts were correctly transcribed from the recordings.
- Ensured that the codes interpreted the correct meaning of the participants and documented any coding conflicts in interpretation.
- Compared the results of what different researchers had coded from the transcripts.
- Agreed on a certain level of consistency of coding, i.e. the extent to which different coders used same or similar codes for different passages of text.

4. Results

The results of this study are threefold (see Fig. 4.1): challenges (C), problems (P) and suggestions (S). The source of the challenges is the expert group, the source of the problems are the potential end user group and the source of the suggestions are all informants, which aim to target both challenges and problems respectively. Each of the three categories and the items under it are elaborated in the text after Fig. 4.1.

Suggestions (S) → Potential Challenges (C) and Practical Problems (P) ↓	S1. Embedding metadata	S2. Reusing existing tools	S3. Integrating with CMS	S4. Improving user experience using logos	S5. Using GUI pop-ups	S6. Embedding during content creation	S7. Verifying the authenticity of the source metadata	S8. Resolving conflicts in composite work	S9. Storing metadata separately	S10. DRM issues	S11. Use simple constraints
C1. Increasing end-users' awareness		X	X	X	X						
C2. Increasing the adoption of the end-user tool		X	X								
C3. Integrating the end-user tool with existing software		X									
C4. Embedding licences with too many constraints						X					X
C5. Attributing multiple right holders in composite work								X			
C6. Rework effort									X		
C7. Verifying the authenticity of the source metadata							X			X	
P1. Informal attribution standard	X		X		X	X					
P2. Users not providing attribution	X			X	X						
P3. Low awareness of licensing			X	X	X						
P4. Lack of attribution support in CMS	X		X								
P5. Issues using a watermark	X										
P6. Issues keeping track the attribution	X										
P7. Verifying the authenticity of the source metadata							X			X	

Fig. 4.1 Mapping of Suggestions to Challenges/Problems

4.1 Challenges

Regarding improving the experience of attributing the author and keeping track of the licensing information, the following challenges are identified through the interviews with those who have expertise in the related area. The challenges include both social perspective and technological perspective.

C1. Increasing End-users' Awareness

A major challenge reflected by the expert group was that people who are using the digital work lack the general awareness of copyright and licences. It appears that the users lack the understanding of what licences are and how they work, or that they are not interested in knowing the licence and rights information of the digital contents. One informant mentioned that most people are reading the related rights information simply out of "curiosity", rather than using it for genuine purposes. He considered the challenge would be to make the licence and rights information "needed" by people, as opposed to "just-fun-to-have".

Moreover, the lack of awareness has led to the lack of community support. One informant mentioned the maintenance problem of an existing open source project: OpenAttribute[4], a project that aimed to ease the process of giving attribution for using pictures licensed under Creative Commons licences. However, this software stopped being maintained due to lack of funding. The informant referred to the nature of such project: only run by volunteers.

It appears that respecting copyright and giving attribution has not yet been part of the social norm. As one informant commented, it eventually might become one of the social norms, but people's awareness could only increase over long periods of time, when we are able to inform them in an understandable fashion.

"And one of the social norms of the future will be to respect copyright. But we can't do that until we get them to show people what copyrights actually are. We can't tell them, we can't show them yet in a nice way, without a thousand pages of legal documents."

C2. Increasing the Adoption of the End-user Tool

The informants suggested increasing the adoption of the end-user tool would be a challenge. If the end-user tool is independent software, it may

---

[4] OpenAttribute http://openattribute.com/

be hard to get the user to actually use the software. Meanwhile it can also be difficult to make the end-user tool as a solid extension to a mainstream proprietary application that creates digital content, such as Microsoft Word and PowerPoint, since these applications sometimes do not support 3rd party modification.

Also few demands and requests from the community affect the tool adoption. Regarding keeping track of the licensing information, the informants reflected that there have been few demands from the communities to actually initiate it or explicitly requesting for a solution. An informant presented it in this way:

"I just don't know many people that are asking for (tracking licences) other than the librarians who are cataloguing resources themselves."

C3. Integrating the End-user Tool with Existing Software

An informant emphasised the lack of existing technical support in reading the embedded metadata. At the moment, there is hardly a mainstream music displayer or picture viewer that supports displaying licensing metadata information embedded in a file. An informant described this integration challenge as a "Catch-22", i.e. applications that embed and applications that read/retrieve metadata would depend on each other's pre-existence, neither one can get started before the other one gets into use.

"But then we're in a Catch-22 about applications that embed that information won't be used until there's applications that consume that information. And there won't be applications that consume it until there are applications that embed it. So it's one of those problems that someone just needs to push forward."

Currently the infrastructure of embedding self-defined metadata exists. (Such as camera info, creation date etc.) But there are hardly any existing tools that specifically target metadata about the licence information.

C4. Embedding Licences of Too Many Constraints

One challenge identified was that the various licences and rights expressions were too flexible and contained too many constraints. The REL was originally invented to provide a means of expressing the usage and access rights to digital contents, and to formulate business models [Gu03]. But over the years, it appears that too flexible rights expressions can give rise to "weird business models" that do not make practical sense to the user. According to one informant, the business models that REL

tried to express sometimes turned out to be "too flexible" and it became impossible to validate all the constraints:

"Back in 2002 we had lots of really weird business models being proposed for using content. They were trying to test it out how good REL works. They came up with business models like, you know, you could buy this video and play it every second Wednesday and when you visited your grandmother's house after 10pm at night.... and the REL were designed around that flexibility and in fact it turns out to be, I think, too flexible. [...] When you have a language that can do anything, then [if you] try to validate it, it's almost impossible."

Another informant raised similar issue that a CMS platform can have: giving users too many options complicates the usage of licences, since users must have good background knowledge on how licences work in order to use the system.

C5. Attributing Multiple Right Holders in Composite Work

One technological challenge is to express licence information metadata for composite digital work, i.e. a work that is composed of several other subordinate works from different authors or licensed under different licences.

One informant gave an example to demonstrate a possible consequence of such problem. For archives and museums, if there is no clear licence statement for the composite work, the institute may be forced to lock up the work and restrict access to it in order not to violate the corresponding copyright laws. This is bad not only for the sharing of digital work but also for the promoting of cultural heritage.

Another informant suggested other situations where parts of the contents can be used for certain purposes but the whole content can be used for other purposes due to different licence permissions. In this case there might be conflicts in rights expressions. Thus expressing multiple licences and attribution in a composite work, resolving rights conflicts, and having "interoperable licensing expressions" could be a future challenge.

According to the informants, currently there is no proper technology for machines to read multiple licence information in composite work. There are different opinions on this issue as well, as an informant has expressed that multiple licences would only make the user interface more complicated. The common resolution is enforcing only one licence for each piece of work.

C6. Rework Effort

As the licence metadata information is to be embedded in the actual digital contents, there are concerns about the amount of effort in adding this new information and in updating the existing infrastructure. This concern is more serious when the amount of digital contents to be reworked is huge. One informant gave an example: Europeana, a web interface that is managing the access to 27 million pieces of digital files from different cultural institutes. Services like Europeana may have big concerns on the transaction costs in reworking, and updating all the contents to the new standard.

C7. Verifying the Authenticity of the Source Metadata

Several informants have frequently mentioned authenticity verification of the source metadata. It means one should not be able to easily get rid of or modify the metadata without authentication. Building a trust mechanism that authenticates the embedding of licence appears to be a hard challenge. Creating the licence is simple and straightforward but it can be hard to embed it and make it available in a secure way. As one informant commented on the existing metadata:

"It's easily to get rid of, because mostly it's header files, and you just remove the header and it's still the image, nobody would know where this is from."

However this issue should not be construed with the access security, which restricts access rights to the digital content. It is a challenge that we cannot simply add a mechanism to encrypt the content and to prevent unauthenticated modification, because at the same time we do not want to prevent the sharing of the content. For example, one informant had expressed worries that any automatic licence compatibility checking might be misused as DRE (Digital Rights Enforcement). One should always be aware of such difference and not to construe the authentication with measurements that restrict the distribution of content.

The informant also stated that a not-well-thought-out mechanism in automatic licence compatibility checking might not care for the fair use scenario, i.e. the use of content under certain circumstances without the need for prior permission from the author.

4.2 Problems

These results were obtained from the interviews with two independent photographers who played the role of both digital work owners and

prospect end users of the tool. These are the practical problems, which are encountered regarding the way digital work and attribution information are handled in the current situation.

P1. Informal Attribution Standards

The informants revealed that sometimes when their work was consumed or distributed, the attribution was given in very informal ways. For example, attribution given in oral format was not acceptable by the photographers. There also exist many forms of "agreement" for granting permissions, which depends on the specific context.

P2. Users Not Providing Attribution

This appears to be a frustrating problem encountered by many photographers. The users of the digital work could have forgotten to give attribution to the owner of the work. One informant explained that she had to be very "picky" and always sent out explicit reminders. She also tried to Google her name or her picture to track down where and how her work has been used. Another reason for not giving attribution could be the difficulty in identifying the author, especially when the work has been widely spread out. Sometimes the file information metadata is used to identify ownership when there is a conflict.

P3. Low Level of Awareness of Licence

It was noticed that the informants were not familiar with the term "licence" regarding digital work and that they never used it before.

P4. Lack of Attribution Support in CMS

Digital work owners had concerns about the attribution support in the mainstream social media. They consider it a risk to put up serious work on content management systems such as Facebook. The informant phrased the problem in this way:

"Whenever I upload anything onto the social media, I know people are probably going to use my pictures in a way. So I'm aware of that. I mean they're not the serious work."

P5. Issues Using a Watermark

Both informants had mentioned using a watermark on the photos as a way to give attribution. Whether one should use watermarks or not remains a controversy. On one hand the owner of the work could gain the necessary publicity. On the other hand, the visible watermark could damage the quality of the original work.

P6. Issues Tracking Attribution Information

One informant mentioned the difficulties in keeping track of attribution of digital work. Currently she tackles this problem and keeps track of the information in a manual way by saving the attribution information in the filename or folder name, but she reflected that it was time-consuming when the amount of work became large.

P7. Verifying the Authenticity of the Source Metadata

Preventing the attribution from being easily modified or gotten rid of is a concern which was voiced from a digital work owner point of view. For example, a watermark could be removed. Moreover, if a screenshot is taken of a picture, the attribution requirement embedded inside might be lost.

4.3 Suggestions

The comments from the expert group and the end user group are categorised as suggestions in this section. The suggestions respond to the challenges and practical problems.

S1. Embedding Metadata

Most informants are positive about embedding licence and attribution information in the digital contents. One expert agreed that it facilitates identifying and tracing the origin of contents. He suggested that embedding metadata is able to improve the usage and popularity of REL, favouring the REL communities.

Having contents and their licensing information together at all times is suggested to be a succinct way of managing contents, because it is difficult to manage transformations between the upstream and downstream use [IH03]. It could decrease dependencies to third party databases, reducing the effort of manually fetching associated rights

information. Such metadata systems are a significant part in managing Content, Parties, and Rights within the content value chain [IH03], i.e. the greater the metadata systems can improve the end-user experience, the greater the value created.

Also the end users expressed the need for some form of "personal logo" that shows credit information without affecting the visual contents. Such "invisible watermark" is in fact a form of embedded metadata of attribution information.

S2. Reusing Existing Tools

A major suggestion was that the development of software that embeds metadata of licences and attribution should be built on existing projects, and/or reuse existing software tools. It can be tools that provide technical infrastructure and support, or tools that have gained popularity among users.

An informant considered it to be easier than developing a new project from scratch, due to established standards, support, and stable user groups. The informant instructed how to develop such software:

"It should be a part of what they already do. You know if they are blogging then it should be part of their WordPress. If they are creating a presentation, it should be a part of PowerPoint."

The informants suggested a list of tools, for example the media viewer VLC, the blogging service WordPress, and Microsoft PowerPoint etc.

S3. Integrating with CMS

Most mainstream content management systems (CMS) in social networks focus on the downstream and upstream use over the digital content [IH03]. Creating licence-handling functionalities for CMS would not only improve the upstream use experience of the end users, but also improve the downstream use experience of the publishing organisations. One informant suggested that the prospective software could take advantage of the CMS user group and promote the use of licences.

Iannella (2010) saw consequences caused by the explicit arrangement under which contents are shared in social networks, because the activities such as sharing of content between close friends might sometimes ignore attribution [Ia10]. Integrating software that embeds licence information with CMS in social networks can make the attribution more explicit.

S4. Improving User Experience with Logos

Two informants suggested using graphical logos for right statements, instead of interpreting terms and conditions in textual format. Having a graphical, transparent, and "up-in-front" reminder helps people gain confidence of which content they can or cannot use. The informants suggested that increased data transparency results in better decision-making.

For instance, Creative Common uses a succinct way of expressing terms and conditions: using graphical badges to represent legal codes. The CC badges serve as a constant reminder for the users what specific permissions are granted for a certain piece of digital work.

Iannella and Finden (2010) has done an experiment study on the privacy icons in social network that evaluates user's interpretation of using graphical representation for the privacy options. The final results provide a sound set of graphical icons for indicating privacy preferences [IF10].

S5. Using GUI Pop-ups

The prospective software could actively send out straightforward notifications to the user, reminding which licence is embedded in the material. This strategy can also increase users awareness of giving attribution.

An example use case is that the system uses a pop-up dialogue to inform the user about the licence information, when the user attempts to modify an image which is licensed under a CC BY-ND (Attribution-NoDerivs) licence.

S6. Embedding During Content Creation

Two informants saw the importance having the licence information collected and embedded during the creation of the content. The creator of content needs to provide information regarding issues such as who is credited, who are the parties that are involved, and what types of permissions they want to use for downstream use. The informants reflected that it would not only help support the management of entities in the content value chain [IH03], but also would simplify the user interface, minimizing the end-user's input. Another informant suggested a simple user interface in CMS that allows user to select and set a licence for their content during the creation of the post.

S7. Verifying Authenticity of the Source Metadata

To respond to the challenge of verifying authenticity of the source metadata, two informants suggested having a certain form of trust mechanism through which the rights information is embedded. An example would be a trusted third party who authenticates and confirms the user's identity when they publish the content.

However, one informant acknowledged that by encrypting and securing the licence information initially during the content creation, this trust mechanism should not be misused or misrepresented in preventing the content from being freely distributed. Another informant implied that this could be a potential cause that leads to the lack of support from the community.

S8. Resolving Conflicts in Composite Work

An informant emphasized the importance of resolving conflicts where multiple media sources are put together to create a composite piece of work. The informant envisioned a possible functionality in a CMS where the system automatically checks uploaded contents for conflicts in rights and licences. The system notifies the user to change the way the content is created or to renegotiate the licence. The user is then able to see whether and how a composite of contents can be put together to create a new work, according to their different licensing terms.

In addition, According to Iannella (2010), when policy conflicts are identified, the appropriate resolution may involve input from the user. Therefore the relevant research must consider human interaction with these policies [Ia10].

S9. Storing Metadata Separately

The traditional way of handling metadata, i.e. storing the licensing metadata information in a separate database, should remain as an important solution for convenient and collective management of the metadata, especially when there is a huge amount of metadata that needs to be managed. An informant has reflected that having such databases that only store this metadata information would make data more accessible, safer, and cleaner and it is beneficial when users only wish to fetch the metadata information of the digital work instead of the content itself.

S10. DRM Issues

The suggestion from the informants regarding DRM boiled down to two aspects: one is to ensure the usage and the free distribution of digital work, the other is to ensure the licence and attribution information travels with the content. It appears that the licence information should be embedded in a way that does not affect the free distribution and usage of the content. One informant suggested that even though REL is a fundamental part of any DRM system, it could still be perfectly adapted to a DRM-free environment.

S11. Using Simple Constraints

Regarding the challenge of too much flexibility in constraints when expressing rights with REL, an informant suggested to implement the software based on a small subset of the terms and conditions. The informant suggested that having simple constraints when expressing the rights statements helps building up a succinct business model. As he gave an example, one could limit the constraints down to critical ones such as region, jurisdiction about time, or devices.

4.4. Validation of Suggestions

We developed a prototype in order to examine the validity of the suggestions. A preliminary version was built after a two-week iteration. It aimed to realise basic functionalities according to the informant's expectations.

The prototype was based on Adobe's Extensible Metadata Platform (XMP) Toolkit SDK, which provides a variety of functions that allows embedding formatted metadata into files. A list of preliminary software requirements was captured from the suggestions given above. Considering the time and resources, the first iteration only fulfils RQ1 to RQ6.

*RQ1. The application should be able to read *.xmp file into XMP Meta objects.*
*RQ2. The application should be able to embed XMP Meta objects into images (preliminarily *.jpg and *.png).*
RQ3. The application should be able to read previously embedded XMP metadata in the image file.
RQ4. The application should support ccREL syntax.
RQ5. The application should be based on Adobe XMP Toolkit SDK.
RQ6. The application should be able to run locally (offline).

As shown in Fig. 4.4.1, the prototype is able to embed a piece of Creative Commons licence information into a JPG image. The licence (CC BY-SA 2.5 SE) is written in ccREL. The visual content of the image remains identical. The embedded XMP metadata can be found in Appendix 3.

Fig. 4.4.1 Comparison between file contents after embedding XMP metadata

The future contributors are recommended to continue implementation with RQ7 to RQ14.

RQ7. The application should integrate with WordPress as a plug-in.
RQ8. The application should use icons.
RQ9. The application should use GUI pop-ups.
RQ10. The application should support embedding licence at content creation.
RQ11. The application should have a trust mechanism that verifies the authenticity of the source metadata.
RQ12. The application should resolve licence conflicts in composite work.
RQ13. The application should be implemented in a DRM-free environment.
RQ14. The application should avoid complex constraints on rights statements.

5. Discussion

The results show that the approach of embedding metadata of licence information in digital work can address many practical problems. The primary intention is to improve end user's experience of identifying and attributing the original author of a piece of digital content. Moreover, for each technological and social challenge identified during development of the end-user tool, there is at least one corresponding suggestion. It appears that embedding metadata of licence information has a promising future.

As discussed by Iannella and Higgs (2003), the rights management technologies were only focused on utilising metadata and REL for managing the downstream use. However the upstream use, i.e. the flow of content from consumers to the publishing organisation, has somehow been neglected. This has been confirmed in the results, i.e. too few metadata systems favour the identification of the creator in regards of attribution information. We speculate that an end-user tool that utilises REL will not only increase REL's usability, but also improve consumer's upstream use experience.

The results show that the attribution problem is a mainly a social problem, rather than a technological one. It appears that the lack of awareness of licence and attribution is a fundamental challenge. It resulted in a series of other challenges, e.g. difficulties in pushing into use of the end-user tool, and lack of technical support in existing software. A number of practical problems discovered in this paper also have their roots in the lack of awareness for licences or attribution, e.g. informal attribution standards, and users not providing attribution.

The results also show that increasing user's awareness of licence and attribution is a long-term process. As one informant argued, respecting copyright and attribution has not become a social norm. However, we see a decent number of suggestions that can respond to the challenges and problems related to the lack of awareness.

6. Recommendations

While there is an increased interest in digital content sharing, this study is useful in developing strategies to increase user's awareness for attribution in the long run.

We recommend future research to further explore more challenges and investigate in wider perspectives. A complete end-user tool is also recommended. The implementation could take into considerations of the suggestions as discussed in this paper. In particular, the development is suggested to focus on the following:

- Support authentication;
- Support multi-licences in composite digital work;
- Compare interpreting licence information with different REL standards (ccREL, ODRL, etc.);
- Develop strategies to increase awareness for attribution;
- Transform image-watermarking solutions to embedding metadata;
- Promote usage by integration with mainstream social network CMS.

7. Conclusion

This case study mainly explored the challenges of developing a tool for embedding licence metadata and attribution information metadata in digital work, from both social and technological perspectives. The study resulted in a list of challenges, practical problems, and suggestions in terms of the implementation of such software.

The challenges identified by the experts are: increasing end-users' awareness, increasing the adoption of the end-user tool, integrating the end-user tool with existing software, embedding licences of too many constraints, attributing multi right holders in composite work, rework effort, and verifying the authenticity of the source metadata.

The practical problems identified from the end user's behaviours are: informal attribution standard, users not giving attribution, low awareness of licences, lack of attribution support in CMS, issues using a watermark, issues keeping track the attribution, and verifying the authenticity of the source metadata.

The suggestions given by both experts and end users in order to address the challenges and the problems are: embedding metadata, reusing existing tools, integrating with CMS, improving user experience using logos, using GUI pop-ups, embedding during content creation, authenticity, resolving conflicts in composite work, storing metadata separately, DRM issues, and using simple constraints. Each suggestion responds to one or more challenges or problems above, as shown in the mapping table in the Results section.

In addition, a preliminary prototype has been developed in order to evaluate the suggestions. The prototype fulfils basic software requirements captured from the suggestions. Some of the essential

suggestions are justified to be valid. The future contributors are recommended to perform further justifications and evaluations.

7.1 Limitations

Due to the limited time and scope of this research, only five informants were chosen with convenient sampling. For instance, only two independent photographers were chosen in the potential end user group. If resources permit, we would have included organisational-level users such as online galleries or newspaper agencies, who might have more professional insights and domain knowledge about digital work.

Acknowledgments

8. Bibliography

[Ab08]   Abelson, H. et. al.: CcREL: The Creative Commons Rights Expression Language. W3C Member Submission. 2008

[Cr09]   Creswell, J.: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 3rd Edition. SAGE Publications Inc., 2009

[GWW01] Gunter, C.; Weeks, S.; Wright, A.: Models and Languages for Digital Rights. InterTrust Star Lab Technical Report STAR-TR-01-04, 2001

[Gu03]   Guth, S.: Right Expression Languages. In (Becker, E. et. al. Eds.): Digital Rights Management. LNCS 2770, Springer-Verlag Berlin Heidelberg, 2003; pp. 101–112

[Ia01]   Iannella, R.: Digital Rights Management (DRM) Architectures. D-Lib Magazine, Vol. 7, No. 6, ISSN 1082-9873, 2001

[IH03]   Iannella, R.; Higgs, P.: Driving Content Management with Digital Rights Management. IPR Systems Pty Ltd, 2003

[Ia04]   Iannella, R.: The Open Digital Rights Language XML for Digital Rights Management. Information Security Technical Report, Vol. 9, No. 3, 2004: pp. 47-55

[Ia10]   Iannella, R.: Towards E-Society Policy Interoperability for Social Web Networks. International Journal of E-Adoption (IJEA), 2(1), doi:10.4018/jea, 2010: pp. 13-29

[IF10]   Iannella, R.; Finden, A.: Privacy Awareness: Icons and Expression for Social Networks. 8th International Workshop for Technical, Economic and Legal Aspects of Business Models for Virtual Goods Incorporating the 6th International ODRL Workshop. Namur, Belgium, 2010

[KM99]   Klein, H. K.; Myers, M. D.: A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. MIS Quarterly, Vol. 23 No. 1, 1999: pp. 67-94

[MN07]   Myers, M.; Newman, M.: The Qualitative Interview in IS Research: Examining the Craft. Information and Organization, 17(1), 2007: pp. 2-26

[Ro02]   Robson, C.: Real World Research. Blackwell, 2nd Edition. Wiley, 2002

[RH09]   Runeson, P.; Höst, M.: Guidelines for Conducting and Reporting Case Study Research in Software Engineering. Empir. Software Eng. 14, 2009: pp. 131-164

Appendix 1. Pre-designed Interview Questions for the Expert Group

1. Could you tell us about your current position and responsibilities in the organization?
2. Are you engaging in any specific activities that handle metadata of digital work licences? Which one is the most important? Why?
3. How do you keep track of the metadata and rights statements?
4. Does the organization have any strategies to make it easy for user to keep track of the licence and attribution information of the digital work?
    Probe: How would you increase people's understanding and awareness of keeping track of the licences and attribution?
5. How do you intend to realize multi-copyright in one work?
6. If there is a tool to embed the metadata and rights statement into the digital objects, are there any particular functionalities or requirements that you'd be looking into?
    Probe: How could it possibly influence the rights holders and users? How do you speculate the consequences? And why?
7. If we are going to develop a prototype that we described, could you provide any suggestions to address the issues that you described in the development of our prototype?
    Probe: Multiple-copyright issue?
8. From your perspective, what are the challenges when embedding the licence and attribution as metadata in the digital work?

Probe: Any comments on security issues?

9. Do you have any further comments?

Appendix 2. Pre-designed Interview Questions for the End User Group

1. Could you introduce yourself?
2. Could you describe the creative work activities you are engaging at the moment?
3. What's your opinion towards copyright issues?
4. How do you release your work on the Internet? Do you release them under any licences?
5. What's your most frequently used photography-sharing platform? (CMS such as WordPress, Flickr, etc.)
6. How do you comment on their copyright support?
7. Have you ever found out your work being reused somewhere without acknowledgement? Tell more?
8. What do you usually do in order to prevent this from happening?
9. What's problematic about it?
10. If there were a software tool to help protect the rights holders, what would you expect out of it?
11. Have you ever used any tools like this to protect your own copyrights?

Appendix 3. Example XMP Metadata

```
<?xpacket begin=" id="?>
<x:xmpmeta xmlns:x='adobe:ns:meta/'>
 <rdf:RDF xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
  <rdf:Description rdf:about="
            xmlns:xapRights='http://ns.adobe.com/xap/1.0/rights/'>
   <xapRights:Marked>True</xapRights:Marked>
  </rdf:Description>
  <rdf:Description rdf:about="
            xmlns:xapRights='http://ns.adobe.com/xap/1.0/rights/'>
   <xapRights:WebStatement rdf:resource='http://theurl'/>
  </rdf:Description>
  <rdf:Description rdf:about="
            xmlns:xapRights='http://ns.adobe.com/xap/1.0/rights/'
            >
   <xapRights:UsageTerms>
  <rdf:Alt>
   <rdf:li xml:lang='x-default' >This work is licensed under a &lt;a
rel=&#34;license&#34;    href=&#34;http://creativecommons.org/licenses/by-
sa/2.5/se/&#34;&gt;Creative Commons Attribution-ShareAlike 2.5 Sweden
License&lt;/a&gt;.</rdf:li>
   <rdf:li xml:lang='en_US' >This work is licensed under a &lt;a
rel=&#34;license&#34;    href=&#34;http://creativecommons.org/licenses/by-
```

sa/2.5/se/">Creative Commons Attribution-ShareAlike 2.5 Sweden License.</rdf:li>
    <rdf:li xml:lang='en' >This work is licensed under a Creative Commons Attribution-ShareAlike 2.5 Sweden License.</rdf:li>
     </rdf:Alt>
      </xapRights:UsageTerms>
    </rdf:Description>
   <rdf:Description rdf:about="
              xmlns:dc='http://purl.org/dc/elements/1.1/'>
     <dc:title>
   <rdf:Alt>
    <rdf:li xml:lang='x-default'>TheTitle</rdf:li>
    <rdf:li xml:lang='en_US'>TheTitle</rdf:li>
   </rdf:Alt>
      </dc:title>
    </rdf:Description>
   <rdf:Description rdf:about="
     xmlns:cc='http://creativecommons.org/ns#'>
     <cc:license rdf:resource='http://creativecommons.org/licenses/by-sa/2.5/se/'/>
    </rdf:Description>
   <rdf:Description rdf:about="
     xmlns:cc='http://creativecommons.org/ns#'>
     <cc:attributionName>TheAuthor</cc:attributionName>
    </rdf:Description>
   <rdf:Description rdf:about="
     xmlns:cc='http://creativecommons.org/ns#'>
     <cc:morePermissions rdf:resource='http://permissionurl'/>
    </rdf:Description>
  </rdf:RDF>
</x:xmpmeta>
<?xpacket end='r'?>