Copymarks:

A suggestion for simple management of copyrighted material

Jon Bing
Norwegian Research Center for Computer and Law
Faculty of Law, University of Oslo[1]

1 Introduction

In this paper, there will be suggested a simple way managing some of the issues related to copyright in the Web environment, especially with reference to electronic agents. Though the idea is simple, it rests on certain presumptions. In introducing the issues to be addressed, a very simple discussion of the copyright issues will be taken as a basis. There are numerous issues not taken into account, the introduction is an indication of the problems of the author in taking hold of one strand in the discussion for further development. Hopefully these shortcomings of the introduction will in some respects be relieved by the following discussion. This, therefore, is some sort of disclaimer or warning that the introduction by no means pretends to be an adequate presentation of the copyright issues related to the digital environment, but only a way towards the central idea of the paper.

Two introductory observations are made. First, the difference between a paper based and digital environment in copyright term in "reading" a protected literary work, a text or a document. Second, the current prevailing way of offering a literary work – a text or a document – on the Web.

In the *conventional environment*, a document is made available to the public in two steps.

First, the document is reproduced, typically in an edition corresponding to approximately what is perceived as sufficient to meet the demand for the document. The act of reproduction is the basic act protected by copyright; it presumes the consent of the copyright holder.[2] There will typically be a publishing contract assigning the reproduction right to a publisher.

Second, the reproduced documents are offered for sale through some retail system. The distribution is a separate exclusive right of the copyright holder. Through a publishing contract, this is also assigned to the publisher.

The interested member of the public – the user – will purchase a copy of the work. In reading the document, the user will not make any actions, which are relevant under copyright law. The copy is sold – typically – with the consent of the copyright holder, given to the retailer by the publisher based on the publishing contract. The user simply opens the book or journal, reading the text. There may be actions taken by the user, which requires us to consider them with respect to the exclusive rights of the copyright holder. One such act would be the reproduction by the user of the copy purchased, typically using a photocopying device – such an act of reproduction will again require either consent of the copyright holder or a statutory license.[3] Another such act would be the further distribution of the copy purchased,

---

[1]   PO Box 6706 St Olavs plas, NO-0130 OSLO, Norway

[2]   Alternatively, a statutory license replacing such consent, this will not be pursued in this paper, and is not essential to the argument being offered.

[3]   In European law, there will typically be limitation of the exclusive right with respect to copying for private purposes, cf Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society art 5(2)(b) "in

for instance by offering it for sale – this will involve the doctrine of "exhaustion of rights", and will be discussed below in another context.

The main point, however, is that the normal exploitation of the document purchased do not involve any acts relevant with respect to the exclusive rights of the copyright holder.

In the *digital environment* of the Web, this is rather different. A user will typically have an identification of a web site on which the document is uploaded – and in our discussion, we presume that it is uploaded with the consent of the copyright holder.

The identification may be from different sources, typically the user will

(1) have found an explicit URL to the cite in another published source, on the Web or otherwise;
(2) have found a hyperlink to the source in another document available on the Web, and which will make it possible for the user to access the document by double-clicking on the hyperlink typically emphasised by colour or underlining; or
(3) have found a hyperlink by using a search engine where the user has given a search term characterising some syntactic feature which the user thinks characterises a document of which he or she has an interest.[4]

The document requested is the communicated to the user. This is in practice done by copying and partitioning the document to fit into packets required by the Internet protocol TCP/IP, then communicating the packets through the Internet, where a packet in transit may be temporarily stored on the servers of the telecommunication operators until the packets reach the work station providing the user with Internet access. There the packets are re-assembled into a document, which then is available for the user, and held in temporary storage (the cache of the CPU) while being displayed on the screen. The document is also easily available for permanent storage by copying it to a local hard drive (or other machine readable storage medium), for local printing or for forwarding through the network to another user.

In the simplified diagram below, it is illustrated how the communication is achieved through a long series of reproductions. These are all relevant to copyright law, and presume the consent of the rightholder (or statutory license).

The issue of the *transient copies* has been discussed in the theory. Some of the representations of the documents – typically in the cache of the servers of the telecommunication operators or the cache of the workstation of the user – are of a temporary nature, and it has been argued that they do not have the permanence necessary to constitute a "copy", but are rather similar to mirror images, which reflect a work, but are no reproduction of that work. This discussion is now only of historic interest, the Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society art 2 states that "direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part" constitute copies.

---

respect of reproductions on any medium made by a natural person for private use and for ends that are neither directly nor indirectly commercial, on condition that the rightholders receive fair compensation which takes account of the application or non-application of technological measures …"

[4] Indexing web pages for a search engine may itself imply acts relevant with respect to the exclusive right of reproduction. As indexing typically involves simple forms of electronic agents, this is an example of what is addressed in the paper, and will be further pursued below.
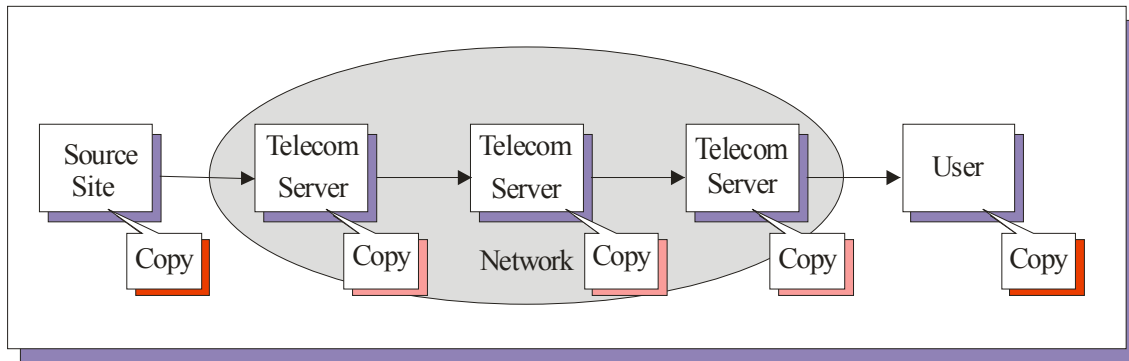
This implies that also the reproduction in the caches of the servers of the telecommunication operators in principle represent copies relevant with respect to the exclusive right for reproduction. However, art 5(1)(a) provides an exception to the exclusive right for "[t]emporary acts of reproduction … which are transient or incidental [and] an integral and essential part of a technological process and whose sole purpose is to enable … a transmission in a network between third parties by an intermediary". In the figure, this is indicated by a slightly paler shade for these reproductions.

There is, however, one more hurdle for qualifying these representations as "copies". As briefly stated, the material (a text document, an image, a sound track or other protected work) are partitioned into packets as required by the TCP/IP protocol of the Internet. Such a packet is certainly a "part" of the protected work, but for works, which require a voluminous representation in digital form (like a high-resolution colour photograph), such "part" may be sufficient small that its representation does not to constitute an act of reproduction of the work seen separately. To constitute a "copy", one may argue that a sufficient number of packets have to be present in the cache of the server at the same time to constitute a reproduction of the work. As the packets do not arrive exactly at the same time, and are typically only stored for less than a second before forwarded, there still may be some room for arguing that an act of reproduction does not take place. This, however, is only of theoretical interest as the Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society limits the exclusive right of reproduction with respect to such cases.

The main characteristic of the use of works in the digital environment therefore is – in contrast to the main characteristic for the use of the same type of works in a paper-based environment – that it presumes reproduction, an act that is the exclusive right of the copyright holder. Such reproduction can only take place with the consent of the copyright holder (or a statutory license).

One may observe that the digital environment also is ideally adapted for obtaining consent easy, typically by a wrap-click contract. These are recognised; see for instance Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market Section 3.[5]

However, tradition within the community of Internet users does not include any measure to provide contract or consent when simply accessing material made available on the net. Therefore, one may argue – in very formal terms – that accessing the material represents a massive infringement of the exclusive right to reproduction.

---

[5] Parts of the Directive does not apply to copyright, but this exclusion does not include Sect 3, see Annex.

Such an argument would be an obvious failure. The material is in the typical case uploaded by the rightholder to be exploited by users. An academic author communicates his or her papers to the scientific community to be read and commented upon by others. A commercial site may derive its income from banner advertisements, and the revenue may be directly linked to the number of "hits" which is registered on the site.

The solution would seem to qualify the act of uploading the material by the rightholder to his or her site as also implying consent for the necessary exclusive acts, *ie* the reproductions necessary to "read" a document. However, implied consent is very imprecise.[6]

To take an obvious example: A user finds a document, which is pertinent to a report being prepared for a client. In order to back up the conclusions of the report, the user prints out a copy of the document (a clear example of reproduction) and annexes this to his or her report before forwarding it to the client. It is very difficult to make any general statement with respect to whether the implied consent also embraces such reproduction and distribution of the resulting copy. In a paper-based environment, the user would have to photocopy the document, and this would imply consent of right holder. One may therefore argue that the implied consent should not give the user more extensive rights. On the other hand, in the paper-based environment there is no consent for reproduction, implied or explicit. Moreover, the users may be exploiting the document in a way, which one may argue is "loyal" to a right holder. The academic author, who has uploaded his or her document, probably will not find it offending that the document is annexed to the report, one may rather see this as a success in the dissemination of research results, adding to the reputation of the author.

The solution would be some digital rights management system. Currently, one is waiting for a generally accepted and easy to use DRMS to emerge. On short or medium term, there is little reason to believe that this will provide a solution with respect to the bulk of material being made available on the net.

At the same time, increased use of electronic agents emphasises the need to offer right holders a simple solution offering some sort of control of the work, something that may be seen as a very simplified and embryonic DRMS. Based on the work within the Alfebiite project, such a suggestion is put forward. It may be seen as an attempt of communicating some of the results of the analysis of the legal issues related to electronic agents to a wider audience, as well as making the theoretical ground work slightly more operational.

2   Copyright notices, incoterms, trustmarks *etc*

The suggestion put forward in this paper has many parallels. One such parallel is the copyright notice.

Today, the most important international treaty in copyright is the Berne Convention – the Berne Convention for the Protection of Literary and Artistic Work under the administration of the World Intellectual Property Organisation. It is important also because there is reference to the Berne Convention in the Trade Related Intellectual Property Services (TRIPS) agreement, where a member of the World Trade Organisation has to adhere to the major substantive articles of the Berne Convention.[7]

---

[6]   The theory of implied consent and the consequences for the digital environment is discussed in detail by Geir Kielland's report developed for the Alfebiite project, "Teorien om implisitt lisens" (Norwegian, under translation).

[7]   Cf TRIPS Part II art 9(1): "Members shall comply with Articles 1 through 21 of the Berne Convention (1971) and the Appendix thereto. However, Members shall not have rights or obligations under this Agreement in respect of the rights conferred under Article 6bis of that Convention or of the rights derived therefrom". The Berne Convention Art 6bis concerns moral rights. The original Berne Convention was concluded in 1886, but the TRIPS make reference to the revised text from 1971, often referred to as the Paris text.

One of the major principles of the Berne Convention is that a member country is not permitted to make intellectual property protection depend on formalities like registration, cf art 5(2) "The enjoyment and the exercise of these rights shall not be subject to any formality". The protection is extended to the work as it is created. This was the reason for the United States of America for more than a century *not* to become party to the convention, as protection in this country – following the tradition at least going back to the Stature of Anne (1709) was dependent on registration and deposit of copies by the US Copyright Office. Only in 1988 did USA ratify the Berne Convention.[8]

Due to the situation in which USA was not party to the Berne Convention, there was in 1952 created the Universal Copyright Convention under the auspices of United Nation Educational, Scientific and Cultural Organisation (UNESCO). The UCC permits formal requirements, but only those specified in the UCC itself. This is the "copyright notice" – the word "copyright" (optional) followed by a special character or sign which is a capital C inscribed in a circle, followed by the name of the copyright holder and the year of first publication, cf art III(1).[9] The notice should be clearly displayed on all copies of the work. The requirements are rather strictly interpreted, omission of a notice will result in loss of protection under UCC, and even small deviations from the stipulated requirements, may have the same effect.[10] A typical copyright notice might have the form

Copyright © Alfebiite project 2002

After the United States ratifying the Berne Convention, the UCC in practice has a minor role to play in legal terms, as protection is extended also in the United States without formalities, as in all the other 149 countries[11] having ratified the Berne Convention. Strictly speaking, the copyright notice is superfluous, but it is still used extensively. This is not because there are legal effects in a strict sense flowing from the use of the notices, but because the notice and the sign © have become associated with the notion of copyright protection. The notice is therefore a compact form of reminding a user of the fact that the work is protected by copyright law, and that there are exclusive rights of the copyright holder, which require consent if the user wants to perform certain acts with respect to the protected matter, as typically reproduction.

Copyright notices are also used on the Web. A web site or page bearing the notice has the same protection as a site or page without the notice. However, the notice will have the practical effect of reminding the user that the material of the site or page is protected by copyright law. And it may have some importance when interpreting the implied license associated with the uploading of the protected material by the rightholder, it will at least be an argument for that the implied license cannot be interpreted to mean that the rightholder has given away all the control of the work which copyright law offers him or her.

It is these two last elements, which are similar to what the copymarks suggested in this paper, may offer:

---

[8] Effective from 1 March 1989.

[9] Protection is assured "if from the time of the first publication all the copies of the work published with the authority of the author or other copyright proprietor bear the symbol © accompanied by the name of the copyright proprietor and the year of first publication placed in such manner and location as to give reasonable notice of claim of copyright".

[10] For instance, it has been permitted that due to the limited character set of early printers, the use of parenthesises rather than a circle was permitted – (C). These had to be "round" parenthesises, if one used square parenthesises, – [C] – the notice might be deemed invalid.

[11] December 2002, see http://www.wipo.org/treaties/documents/english/word/e-berne.doc.

- Information of a practical nature to the user that there resides a certain protection in the material,
- in addition, stating more explicitly the terms of the unilateral license implied by the uploading.

In electronic commerce, the use of *trustmarks* has become common. A trustmark is a visual symbol or logo, which a web site typically is given the permission to use by an organisation that requires the site to adhere to certain standards. Typically, the trustmarks are used for sites offering goods or services to consumers, and the associated standards govern the policy of the provider with respect to such issues as data protection, return of goods not meeting the announced quality standards and refund of payment, damage during transport, dispute settlement mechanisms, *etc*. The organisation permits the trustmark to be used only after it typically has checked that the business practices correspond to the standards, and it typically will audit these practices periodically. In addition, customers, for instance consumers, may complain to the organisation that the site does not adhere to the practices implied by the use of the trustmark. The sanction is generally limited to the withdrawal of the permission to use the trustmark, and the success of the strategy therefore will rely on the reputation, and the associated value, of the trustmark.

The trustmark has some characteristics shared with the copyright notice. However, it awarded by a third party – the organisation licensing the trustmark – which checks that practices correspond to some published standard. The user may easily access these rules through the web, and in this way learn of these policies. The trustmark is not a contract proper, it is a unilateral declaration of the provider of the goods or services offered that it would adhere to certain policies. However, it is presumed that if a contract is entered with a user, this contract will correspond to these policies. If there can be demonstrated to be a conflict, typically that the contract limits the rights of the purchaser more severely than indicated by the policies associated with the trustmark, there may be a legal issue in the interpretation of the contract – and, of course, the permission to use the trustmark by the provider may be at stake.

Trustmarks and the associated policies represent a strategy of self-regulation. The Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market Chapter III Art 16 encourages the use of codes for self regulation of electronic commerce. Trustmarks are currently perhaps the most common strategy for such self-regulation in electronic commerce.

It is reported that the European Commission is planning a single trustmark scheme to ensure minimum standards of on-line consumer protection. By displaying an easily recognizable trustmark on a Web Site, providers of goods or services would be able to show consumers they meet certain protection standards, such as data privacy requirements, commitment to price and information transparency and information about how to deal with complaints.[12]

The International Chamber of Commerce (ICC) maintains a manual of contractual clauses for the sale of goods *etc* known as *INCOTERMS*. These are much abbreviated statements. The statement "CIF (INCOTERMS)" is interpreted to imply that the contract with respect to "cost, insurance and fright" of the goods is governed by the CIF clause in the detailed manual, provisions which are based on the custom in international trade, but made explicit and detailed in the manual.

---

[12]   Reuters, 29 October 2002.

Originally, abbreviations developed when the telegraph made it possible to communicate speedily, but at a cost per character. The solution was abbreviations of typical clauses like CIF, FOB (free on board), CFR or C&F (cost & fright) *etc* which characterised the underlying substance of the clause. Based on custom alone, such abbreviations might in detail be interpreted differently and create conflicts.

Therefore, in 1936 the ICC created the "International Commercial Terms" (known by the abbreviation INCOTERMS). Each INCOTERM refers to a type of agreement for the purchase and shipping of goods internationally. There are 13 different terms, each of which helps to deal with different situations involving the transport of goods. For example, the term FCA is often used with shipments involving Ro/Ro or container transport; DDU assists with situations found in intermodal or courier service-based shipments. There is a simple syntax for the terms: Terms beginning with "F" refer to shipments where the primary cost of shipping is not paid for by the seller, terms beginning with "C" deal with shipments where the seller pays for shipping, *etc*.

The INCOTERMS have been endorsed by the United Nations Commission on International Trade Law (UNCITRAL), and there are authorised definitions available in 31 languages.[13] The current version is INCOTERM 2000.

In many countries, the INCOTERMS were implemented in national legislation, giving them a default interpretation as specified in the law, and corresponding to the ICC manual. Today, the term are typically not defined in legislation, but is given legal effect through the reference by the parties in a binding contract.

For *electronic data interchange* (EDI) similar schemes were developed from the perspective of trade facilitation. As stated in the introduction to the United Nations Directories for Electronic Data Interchange for Administration, Commerce and Transport:[14]

> "When data are interchanged between trade partners by means other than paper documents, e.g. by teletransmission methods including direct exchange between computer systems, a common 'language' should be used with an agreed mode of expressing it, i.e. common protocols, message identification, agreed abbreviations or codes for data representation, message and data element separators, etc. If a universally-accepted standard is not used, the 'language' has to be agreed bilaterally between each pair of interchange partners. Taking into account the large number of parties exchanging data for an international trade transaction and the ever increasing number of potential users of teletransmission techniques, it is obvious that such a bilateral approach is not viable."

The basis for any trade data interchange is the United Nations Trade Data Elements Directory (UNTDED), where data elements are uniquely named, tagged and defined, and where the representation of data entries is specified both as regards expression and syntax. From this directory, data elements required to fulfil specific documentary functions are selected.[15]

The EDI schemes are, in contrast to the other examples sketched in this brief overview, designed for communication between computers. The messages are typically generated automatic by a computer system triggered by a certain event.

---

[13] A brief introduction is published by ICC on its website, cf http://www.iccwbo.org/incoterms/understanding.asp.

[14] Cf http://www.unece.org/trade/untdid/welcome.htm

[15] The Nordic Legal Committee on trade facilitation played an important part in developing Uniform Rules of Conduct for Interchange of Trade Data by Teletransmission (UNCID), which today is part of UN Trade Data Interchange Directory (UNTDID). The NRCCL is rather proud of its co-operation with this Committee in developing the first draft of UNCID.

For instance, the number of items in stock by a retailer falls under a defined threshold, and the inventory system requests that a message should be generated to the wholesaler to supplement the stock. The message is received by the system of the wholesaler, generating an acknowledgement to the system of the retailer, and generating a request for in-house systems, which governs the loading of the truck taking goods in the direction of the retailer. The in-house systems will also take care of logistics, for instance co-ordinating the scheduling of trucks with respect to the total volume of items ordered.

Basic to the EDI scheme would be a legal framework, an interchange agreement, which would ensure that the exchange of messages between the computer systems of the parties result in legally binding contracts. The interchange agreement will also define the risk for an economic loss resulting from some form of malfunction of the system, or unexpected failures.

3 Access control

One may also mention the different attempts at introducing access control in the internet environment. To some extent, this has been discussed in copyright terms.[16] In our perspective, such access control may be seen as part of a digital rights management system. However, of course there are access controls also implemented for other measures. The more conventional is probably the scrambling of a broadcasting signal, making it necessary for the user to acquire a decoder unit and a smart card with an embedded chip through a contract with the broadcaster. The equipment constitutes a set-top box for the reception of television. In Europe, a directive requires member states to introduce a provision in their national legislation making it a crime to circumvent such a control.[17] It is a crude access control, and not very relevant in the perspective of this paper.

More interesting are the parent control systems based on filtering or rating.

Filtering systems are often based on a rather rough identification of words or phrases indicating the subject discussed. However, the natural language is a complex and sophisticated construct, and it is difficult indeed to make an appropriate rule on this basis, as the huge literature on text retrieval will disclose.[18] The occurrence of homographs is only one of the more obvious problems – though "tit" frequently is used as a synonym for the female "breast", it also is the name of a small bird. A even more severe problem is that of context, again using the word "breast" as an example, it often will occur in a text of an explicit sexual nature, but also in a medical discussion of breast cancer. In our context, it may be taken as an illustration of the problems an electronic agent would have in trying to interpret the natural language of a web site, and the risks it would imply to have the agent act on rules based on the occurrence of terms or phrases. Until we more satisfactory have solved the challenging problem of natural language understanding, such filtering techniques hardly can be considered for access control in our perspective.

The Platform for Internet Content Selection offers several *rating systems*. The rating systems typically rely on self-rating by the content provider according to "maturity level"

---

[16] Cf Jane C Ginsburg "From Having Copies to Experiencing Works: The Development of an Access Right in U.S. Copyright Law", *Columbia Law School, Public Law & Legal Theory Working Paper Group, Paper Number 8,* New York 2000, at footnote 20, with citations and Jon Bing"The New or Evolving 'Access Right'", *ALAI 2001 USA, Proceedings of the ALAI Congress June 13-17, 2001,* Columbia University, School of Law, New York City 2002:288-315. "Le 'droit d'accès': renouvau ou évolution", *ALAI 2001 USA, Actes du Congrès de l'ALAI 13-17 juin 200,* Columbia University, School of Law, New York City 2002:316-345.

[17] Cf Directive 98/84/EC of the European Parliament and of the Council of 20 November 1998 on the legal protection of services based on, or consisting of, conditional access**.**

[18] Cf Jon Bing *Handbook of Legal Information Retrieval,* North-Holland, Amsterdam 1984 on the causes to recall and precision failure-

(aimed to exclude access to pornographic or very violent content) or some other scale.[19] The self-rating may be replaced by rating by a panel. The web browser may be enabled to employ one or several of these rating systems to screen the web sites to be displayed.[20] This system is related to the issues discussed in this paper, as it allows the content provider to indicate to the user some property or quality related to the content. In a way, it may be considered as some sort of mark-up language (see below), and certainly adds a meta-layer to the document. It does not offer any control, however, on behalf of the content provider – it offers the *user* some control of which material to access (or more practical, what material a parent will allow younger members of the household to access).

4   Mark-up languages and metadata

4.1   The Dublin[21] Core initiative

The Dublin Core Metadata Initiative (DCMI) is an organisation promoting the adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources. In this paper, we are concerned with metadata, but not in the general sense – we are mainly concerned with metadata that can be interpreted by electronic agents. The Dublin Core offers a set of metadata, and the original set includes

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

The initiative endeavours to define a set of metadata elements that is interoperable, and can be associated with sources in different formats and different forms. It may be appropriate to maintain that the initiative is closely related to a librarian approach, and that the definitions offered are not sufficient for the purposes of a DRM. The element most related to the perspective in this paper, is the last – "Rights" – which is explain as:

> "Typically, Rights will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights

---

[19]   The RASCi system offered four categories of disfavoured content: Nudity, sex, language, and violence and five settings within each category, cf Majorie Heins "Filtering Fever", *Filters & Freedom 2.0, Free Speech perspectives on Internet Content Controls* (2. ed), Electronic Privacy Information Center, Wshington 2001:3.

[20]   For MS Internet Explorer, the controls may be found under the tabs "tools", "internet settings", and "contents".

[21]   "Dublin" refers to a workshop in Dublin, Ohio, March 1995.

> element is absent, no assumptions may be made about any rights held in or over the resource."

The element therefore offers a slot for a statement on the rights associated with a source, but it is rather coarse with respect to the level of definition of the license as envisages in this paper. However, this element, of course, may be further developed. Several of the other elements have been further refined,[22] but – as far as the author is aware – this is not the case of the element "Rights".

Indeed, it has been claimed that the element "rights" is "fraught with confusion".[23] Creative Commons has therefore, adopted a somewhat more structured representation:

```
<dc:rights>
    <Agent rdf:about=http://me.yoyo.dyne.name/>
    <dc:title>Yo-Yo Dyne<dc/title>
    <dc:date>1010-10-01<dc:date>
</Agent>
</dc:rights>
```

In this, the rightholder is a person with an URL and a name, the Dublin core element "title" has been used to link to the work created by the rightholder, and a date of birth is given. This is suggested to be identical to a copyright notice (see above).[24]

The Dublin core elements have been widely recognised, and are often included in other "semantic" representations of web pages, making it possible to map one type of metadata onto another type, making the elements interoperable. The Dublin core set of metadata may therefore also be a platform for creating a set of interoperable statements of rights, following the suggestions below.

4.2   SGML and other mark-up languages

In this paper, it would not be appropriate to include a general introduction to mark-up languages. The uses for mark-up languages are diverse, the basic notion rather simple.

A mark-up language adds a meta-level to – typically – a document. It consists in principle of two parts:

- A set of definitions, in which certain codes are defined (elements and attributes).
- An application interpreting these codes, and acting upon identifying a code in the document.

A naïve example would be a mark-up language for a text. The text has a certain structure, perhaps similar to this document: Headings of different levels, paragraphs with or without indentation, bulleted texts, blank lines, *etc*. A mark-up for the level 1 headings might simply be </h1> for start of heading level 1, and <h1/> for end of heading one. An application may interpret these codes for layout purposes, and make headings level 1 appear in bold face New Roman Times point 12, colour red, with leading consecutive numbers, starting with "1" for the first heading in the document. However, another application might interpret the same codes differently, for instance assigning another font, centring the heading, *etc*. Alternatively, it might not be interpreted for layout purposes, but being a program updating a data base system, the text would be partitioned into separate elements for retrieval purposes.

---

[22]   Cf http://dublincore.org/documents/dces/.
[23]   Creative Commons, cf http://creativecommons.org/learn/technology/metadata/extend.
[24]   Creative Commons will express this in RDF, see below.

This is probably sufficient to convey the basic idea of mark-up languages. There are several standards of such languages; the basic standard is, perhaps, the Standard General Mark-up Language (SGML). On this is based the HyperText Markup Language,[25] well know to all users of the Web as the abbreviation "html" occurring in the URL of a site.

If logging onto the Alfebiite home site, a home page is displayed with some nice graphics, the top being reproduced below:

Using the "source"-function of the "view"-option of the browser, the html-coding is revealed, and the first few lines are:

```
<html>
<head>
<title>ALFEBIITE Home Page</title>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
<script language="JavaScript">
<!--
function MM_preloadImages() { //v3.0
  var d=document; if(d.images){ if(!d.MM_p) d.MM_p=new Array();
    var i,j=d.MM_p.length,a=MM_preloadImages.arguments; for(i=0; i<a.length; i++)
    if (a[i].indexOf("#")!=0){ d.MM_p[j]=new Image; d.MM_p[j++].src=a[i];}}
```

The codes are easily to identify, if not to interpret, but reading the text, one gets the idea that images, including the one reproduced, are loaded onto the page. The code – or meta-tag – "title" corresponds to the example above, though the syntax is different.

It may be argued that html is a development based on SGML. There are further developments, for instance the Extended Mark-up Language (XML). Currently, standards are developed using XML to replace the EDI-messages for electronic commerce in open systems, typically trade on the Web. XML makes it possible to define new codes. Reading the definition, a human will often understand what the code is trying to accomplish. The codes may also be interpreted by an application, provided its programming is appropriate. In the same way as with respect to EDI-messages, parties may define vocabularies describing structures and semantics for data exchanges, with fixed sets of elements and permitted attributes. In principle, such systems can be built by anyone using XML as a tool.

The potential of XML for electronic commerce is considered huge. In principle, the legal aspects are similar to that of systems based on EDI, but the systems may be more flexible. In this context, these possibilities are not pursued.

The main point in this brief presentation is the introduction of meta-tags (codes) in mark-up languages, which may be understood by humans, but are developed to be interpreted by computer systems, and therefore is a simple formalism with a defined semantic. A computer system appropriately programmed may therefore take appropriate action encountering such meta-tags.

---

[25] The Generalized Markup Language was developed by Charles F Goldfarb of IBM in 1969, who also in 1975 wrote the Document Composition Facility (DCF). It was not only the source for the Document Type Description for SGML in ISO 8879, but also it was the source for the World Wide Web's HTML document type as well. Cf Charles F Goldfarb "The Roots of SGML – A Personal Recollection" at http://www.sgmlsource.com/history/roots.htm (December 2002).

The relevance for electronic agents is probably self-evident. An electronic agent would interpret the meta-tags, in this way acquire information, and possibly take action. One possible action would be to generate its own message in the same formalism and communicate this to the agent or site from which the information is acquired. The result will be similar to the example of the communication between the systems of the retailer and wholesaler sketched above, but potentially much more sophisticated. In the Alfebiite project, it is presumed that the interaction, for instance, may take the form of a negotiation.

5 Legal recognition of meta-tags

Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ("Directive on electronic commerce") governs several legal aspect. One of them is the Liability of intermediary service providers (Sect 4). It is not the purpose of this paper to discuss the liability of intermediaries, the section of the directive establishes principles for when intermediaries such as telecommunication operators providing mere conduit, or operators providing caching in proxy servers or hosting, are *not* to be held liable for the information transmitted, cached or hosted.

Proxy servers are used in the Internet to optimise traffic. As indicated in the introductory figure, material from the original site is communicated in a relay race from node to node in the network until arriving at the workstation of the user. Each time a user requests access to the original site, the material is communicated. If several users request the same material, it will be communicated as many times through the net, generating a traffic load. It the requesting users are located relative to the available links in such a way that the data has to flow along the same links, such repeated request may create congestion. Services in the net measure this traffic, and make an automatic decision to store a copy of the material (the site) closer to the requesting users. In this way, traffic may be optimised.

In principle there material will be reproduced on the proxy server, but the Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society limits the application of the exclusive right in such a situation, cf the discussion above at sect 1. However, the material may be illegal for other purposes – it may be uploaded without the consent of the rightholder (but these cases are excluded from the discussion in the paper), it may contain slander, pornography *etc*.

The Directive of Electronic Commerce specifies in art 13(1) the conditions for the operator of the caching service to avoid liability:

Where an information society service is provided that consists of the transmission in a communication network of information provided by a recipient of the service, Member States shall ensure that the service provider is not liable for the automatic, intermediate and temporary storage of that information, performed for the sole purpose of making more efficient the information's onward transmission to other recipients of the service upon their request, on condition that:

(a) the provider does not modify the information;

(b) the provider complies with conditions on access to the information;

(c) the provider complies with rules regarding the updating of the information, specified in a manner widely recognised and used by industry;

(d) the provider does not interfere with the lawful use of technology, widely recognised and used by industry, to obtain data on the use of the information; and

(e) the provider acts expeditiously to remove or to disable access to the information it has stored upon obtaining actual knowledge of the fact that the information at the initial

> source of the transmission has been removed from the network, or access to it has been disabled, or that a court or an administrative authority has ordered such removal or disablement,

We are not suggesting discussion of this article, but would like to indicate items (b) and (c) – a condition for avoiding liability is that the provider of the caching service comply with the conditions on access and the rules for updating – for the latter it is indicated that this should be "specified in a manner widely recognised and used by the industry". The preamble to the Directive does not explain what types of "manners" the article refers to, but we argue that the wording imply that certain elements included for interpretation of computer programs only, have been made relevant for determining the liability of the providers of a cache service.

What material is to be stored in the cache of the proxy server, is not decided on a case-by-case basis by the operator, this obviously is not feasible. There are programs which analyse the traffic flow, and which – according to the criteria in the program – "decides" what is to be reproduced. In order to comply with the rules, for instance of updating, the provider of the service will have to ensure that the programs observe such rules. Therefore, we here have an example of the liability of the operator (a legal or a physical person) depending on whether the programs are designed to follow the rules specified in a manner widely recognised by the industry.

The operator of the original site, and the rightholder, will know, of course, that the network includes proxy servers and that the material may be reproduced and accessed on a serve outside his control. There may be situations in which he wants to ensure that such intermediary storage does not continue beyond a certain date. This may be written in clear text in the document, but he will know that the program deciding to reproduce the material on a proxy server, will not be able to interpret natural language, and that no human is involved in the decision.

Therefore, there are possibilities to address the program itself; one such way is the meta-tags of HTML. A meta-tag specifies information about the document, and has no effect on the appearance of the web page. It is intended for use by other programs, such as search engines, web browsers or the programs of proxy servers.

For instance, the meta-tag

> HTTP-EQUIV="FieldName"

specifies the name of a HTTP response header field. Widely recognized values include:[26]

> *Expires*: The date and time after which the document should be considered expired. An illegal date, such as "0" is interpreted as "now." Setting the Expires attribute to 0 may thus be used to force a modification check at each visit. Dates must be given in RFC850 format, in GMT. For example:
>
> ```
> <META HTTP-EQUIV="expires" CONTENT="Sun, 18 Dec 2002 09:32:45 GMT">
> ```

It may easily be argued that this represent information on the updating of the site from which the material is reproduced, that is "specified in a manner widely recognised and used by industry". The operator of the proxy server may use a program that disregards such a tag, and retain the material beyond the specified date. But the operator does this at a risk, according to

---

[26] Cf HTML Tag Reference, http://developer.netscape.com/docs/manuals/htmlguid/tags3.htm#1697602.

the Electronic Commerce Directive Art 13(1)(c), the operator will then not enjoy the exclusion for liability.

Another example of a similar way of addressing programs is a *robots.txt* file. This is mainly designed for search engines, which is a simple type of electronic agents. The program constituting the search engine will look into the root domain for a file named "robots.txt". The first part specifies the robot, while the second part consists of directive lines, disallowing the robot to index files or directories. A simple example would be

```
User-agent: googlebot
Disallow: cheese.htm
```

This is addressed to the robot "googlebot", and direct this not to index the file "cheese.htm".[27]

If one does not have access to the root domain, it is possible to use a *robot meta-tag* in HTML. The Robots META tag is placed in the HEAD section of your HTML document: The format is quite simple:

```
<HTML>
<HEAD>
<META NAME="ROBOTS" CONTENT="NOINDEX,NOFOLLOW">
<META NAME="DESCRIPTION" CONTENT="THIS PAGE ....">
<TITLE>...</TITLE>
</HEAD>
<BODY>
...
```

There are four directives that can be placed in the meta-tag: index, noindex, follow, and nofollow. "Index" tells a robot that it is all right to index an page, with "noindex" being a prohibition, and "follow" tells the robot that it is all right to follow a link, with nofollow being a corresponding prohibition.

One will note that none of these tags *etc* will have the effect of compelling a robot or a program for proxy servers to follow the rules implied. The disloyal operator may instruct the robot or the program to disregard the rules set out in the formalism. Today, this will not have any legal consequences – with the possible exception for the liability of operators of proxy servers indicated above. The meta-tags and robot.txt possibilities have not been developed to address legal issues.

However, even the existing modest possibilities may be relevant for governing search engines or the services of a proxy server. It may be argued that the Electronic Commerce Directive art 13(1)(b) is an indication of a *strategy of legal policy,* making elements like meta-tags relevant for determining liability. These are elements not addressed primarily to human users, but to the programs operated by such users, requiring such programs to take into account certain specifications "widely recognised and used by industry". This strategy is suggested developed to a simple system for DRM in this paper.

6   Copymarks

6.1   *Effect of unilateral clauses*

The suggestion includes the use of a unilateral clause that is imposed by the rightholder, binding any third party using the material uploaded and made available to the public. There is no process concluding a contract – this would require an offer (by the right holder) and an acceptance (by the user) along the lines set out in the Electronic Commerce Directive Sect 3.

---

[27]   Cf http://www.searchengineworld.com/robots/robots_tutorial.htm.

This possibility is certainly not excluded by this suggestion. It is presumed that mechanisms for concluding on-line contractual licenses will be subject to rapid development, also extended to negotiations between autonomous electronic agents.[28] The copymarks may be an interim solution, something that may be deployed very easily, without the infrastructure necessary to support a properly functioning DRM.

The legal effect of unilateral contractual clauses has been subject to controversy within the area of information and telecommunication technology. A "classic" example of such clauses is those known as "shrink-wrap licenses". These are licensing terms printed on the box containing the disks *etc* of a computer program, stating that the user will be bound by these clauses if the seal is broken. There has been reluctance to accept such licensing practises; there are no negotiation between the provider of the programs and the user, and the user is therefore left with a choice of not using the program at all, or accepting the terms. It has been argued that recognising such licenses by the law would be to allow the providers to print their own law, as the licensing terms would replace the default regulation of the copyright legislation which otherwise would govern the relation between the parties.[29]

However, the scepticism is based on a situation where the unilateral clauses *restrict* the legal position of the user compared to the situation under the default legislative regime. The copymarks are designed *not* to restrict the rights of the user compared to the default law, rather the copymarks only make explicit the default regulation for "copyred", while for the two other categories suggested, the user is granted more extensive use of the material than flowing from the default legislation.

There is obviously a problem in comparing the legal content of the license to the background legislation, as the national law governing the copyright issue has to be determined according to the rules of the interlegal law (jurisdiction and choice of law). This paper will not look into the interlegal issues in this respect.[30] There will certainly be variations in national legislation with respect to the default situation, but for the European region the law in this respect will be co-ordinated by the Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, and therefore not represent any grave problem.

For the two other alternatives suggested in this paper, the rightholder unilaterally extends greater rights to the user than flowing from the default legislation. It is generally deemed that such conditions will bind the user. If the user finds these conditions less than acceptable, the user may fall back on the (more restrictive) default legislation. The situation is therefore rather different from the situation with respect to the traditional unilateral shrink-wrap licenses that seek to *restrict* the user compared to the default legislative situation.

6.2   Creative Commons and unilateral licenses

Creative Commons[31] was established in 2001, and is an effort fronted by several well-known academic figures, like Professor Larry Lessig. It was supported by the Centre for Public Domain, and was helped on the way by the Berkman Center for Internet and Society (Harvard Law School). It is co-located with the Stanford Law School for Internet and Society, directed by Glenn Otis Brown, a prominent intellectual property lawyer.

---

[28]   As for instance set out in the "Lovely Rita" scenario developed for illustration within the Alfebiite project.
[29]   In Norwegian law, the legal history of the relevant provisions of the copyright legislation makes it clear that shrink-wrap licensing normally will not be binding upon the user.
[30]   There have been many contributions to this issue in the last few years, though there still is no consensus on the jurisdiction or choice of law question. One of my own contributions is "Copyright in Electronic Commerce and Private International Law"; Jan Kaestner (ed) Legal Aspects of Intellectual Property Rights in Electronic Commerce, Verlag CH Beck, Munich 1999:33-42.
[31]   http://creativecommons.org/

The organisation explains some of its background:

> "Until 1976, creative works were not protected by U.S. copyright law unless their authors took the trouble to publish a copyright notice along with them. Works not affixed with a notice passed into the public domain. Following legislative changes in 1976 and 1988, creative works are now automatically copyrighted. We believe that many people would not choose this "copyright by default" if they had an easy mechanism for turning their work over to the public or exercising some but not all of their legal rights. It is Creative Commons' goal to help create such a mechanism."

Above, the use of the copyright notice has been briefly introduced. When the US become party to the Berne Union, art 5(2) prohibits any formal requirements to be protected by copyright, including the use of the copyright notice, which depended on the UCC rather than internal US law.

Attribution. You let others copy, distribute, display, and perform your copyrighted work — and derivative works based upon it — but only if they give you credit.
Noncommercial. You let others copy, distribute, display, and perform your work — and derivative works based upon it — but for noncommercial purposes only.
No Derivative Works. You let others copy, distribute, display, and perform only verbatim copies of your work, not derivative works based upon it.
Share Alike. You allow others to distribute derivative works only under a license identical to the license that governs your work.

The idea is related to the Free Software Foundations GNU[32] General Public License,[33] which uses a license containing terms to secure that those utilising the open source code for derivate works will not be able – without violating the terms of the license – to impose exclusive rights on the result.

The license conditions are spelt out in detail for the Creative Commons alternatives. There is also a small "expert system" which guides the user through a series of questions, and in this way determines which type of license is most appropriate for the user's needs.

Creative Commons offers their licenses in three or four levels. The first level is the logo, which may be attached to a web page. The second level is a popular description or explanation of the content of the license, developed to be understood by a user without any legal background. The third level is the license itself, drafted with all the cunning used in US contract law. Currently, these licenses have been drafted for US law, but it is the plan to have them developed for different jurisdictions, the two first levels (logo and explanation) being common, and the third level, the license itself, possibly differing between jurisdiction to lock properly into background law.

The fourth level is the license in a mark-up language. This will state in terms, as introduced above, the nature of the license to be read by browsers. An example may be:[34]

```
<! -- Creative Commons License -->
<a href="http://creativecommons.org/{license URL}
<img alt="Creative Commons License" border="0"
```

---

[32]   GNU is a recursive acronym for ``GNU's Not Unix''.
[33]   http://www.fsf.org/licenses/gpl.html
[34]   Copied from http://creativecommons.org/learn/technology/usingmarkup.

```
scr=http://creativecommongs.org.images/public/somerights /></a>
<br />
This work is licensed under a
<a href=http://creativecommons.org/{license URL}>Creative Commons License
</a>
<!-- /Crestive Commons Lincense -->
<!--
<rdf:RDF xmlns=http://web.resource.org/cc/"
         xmlns="http://web.resource.org/cc/"
         xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
Work rdf:about="">
<lincense rdf:ressource="http://creativecommons.org/{license URL}" />
</Work>
</rdf:RDF>
-->
```

The Creative Commons scheme is obviously borne from the same type of concerns basic to this paper. Indeed, the scheme became known to the author when doing the research for this paper, after it had been already published in a draft form. This is mentioned to emphasise the independent origin of the scheme of Creative Commons and the scheme suggested in this paper.[35] There may other similar schemes, which are not known to the author – and one should appreciate that the basic idea, perhaps, is rather obvious for those concerned with intellectual property law and the shortcomings offered by the "band aid" of implied licenses. However, Creative Commons should be given the credit the initiative deserves for making the idea operational and easily accessible.

A difference between the objective of this paper and Creative Commons may be found, perhaps, with respect to Digital Rights Management. As is indicating in the final section, the paper envisages the formalism developed associated with copymarks (see below) to used directly by electronic agents negotiating for the purchase of a license. Creative Commons emphasise that their tools act as information aids, not instruments of control.

> "We want to help copyright holders notify others of their obligations and freedoms, and to help everyone find places on the internet where creative reuses are encouraged."

It may be suggested that the difference in more one of policy than substance, Creative Commons emphasising that their information is about the rights to reuse, while a DRM is about the conditions for reuse (or other utilisation). Both approaches may be seen as a way of defining the normative position of the right holder, and the difference then is perhaps more the attitude towards using the normative position to grant rights to third parties.

It also would seem that the licenses of Creative Commons currently have not been developed to meet the possibility of autonomous electronic agents – which is only appropriate for licenses operational in the current digital environment.

---

[35] As mentioned below, a form of copymarks was adopted by the Norwegian Research Center for Computers and Law as early as 1995, but only consisting of a logo hyperlinked to a web-page setting out the license in natural language.

6.3   Copymarks – introducing the suggestion

The copymark suggestion has three related elements:[36]

- A logo to be displayed on a website or page, in a way similar to a copyright notice, a trustmark or the symbols developed by Creative Commons, see above.
- An associated webpage setting out in clear text the conditions of the copymark license, and some practical information, like contact address of the author *etc*.
- A meta-tag for the HTML-coding of a page, similar to that of robot meta-tag briefly mentioned above, directing electronic agents to respect the terms of the license as expressed in the formalism of the meta-tag.

All elements will be part of an attempt on behalf of the rightholder uploading material to the Web, to better control this intellectual property. These elements will therefore constitute rights management information, as understood by the World Intellectual Property Organization Copyright Treaty (WCT) Art 12(2),

> (1) Contracting Parties shall provide adequate and effective legal remedies against any person knowingly performing any of the following acts knowing, or with respect to civil remedies having reasonable grounds to know, that it will induce, enable, facilitate or conceal an infringement of any right covered by this Treaty or the Berne Convention:
> (i) to remove or alter any electronic rights management information without authority;
> (ii) to distribute, import for distribution, broadcast or communicate to the public, without authority, works or copies of works knowing that electronic rights management information has been removed or altered without authority.
> (2) As used in this Article, "rights management information" means information which identifies the work, the author of the work, the owner of any right in the work, or information about the terms and conditions of use of the work, and any numbers or codes that represent such information, when any of these items of information is attached to a copy of a work or appears in connection with the communication of a work to the public.

It also will fall within the scope of the Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, cf Art 7(2), which is near identical to the clause cited above:

> *Article 7 – Obligations concerning rights-management information*
>   1. Member States shall provide for adequate legal protection against any person knowingly performing without authority any of the following acts:
>   (a) the removal or alteration of any electronic rights-management information;
>   (b) the distribution, importation for distribution, broadcasting, communication or making available to the public of works or other subject-matter protected under this Directive or under Chapter III of Directive 96/9/EC from which electronic rights-management information has been removed or altered without authority,
>   if such person knows, or has reasonable grounds to know, that by so doing he is inducing, enabling, facilitating or concealing an infringement of any copyright or any

---

[36]   The two first elements were introduced for the web-site of the Norwegian Research Center for Computers and Law in the early versions approximately 1997. They have not been maintained, but an example may be found at http://www.jus.uio.no/iri/forskning/lib/papers/dp_norway/dp_norway.html.

rights related to copyright as provided by law, or of the sui generis right provided for in Chapter III of Directive 96/9/EC[37]

2. For the purposes of this Directive, the expression "rights-management information" means any information provided by rightholders which identifies the work or other subject-matter referred to in this Directive or covered by the sui generis right provided for in Chapter III of Directive 96/9/EC, the author or any other rightholder, or information about the terms and conditions of use of the work or other subject-matter, and any numbers or codes that represent such information

The first subparagraph shall apply when any of these items of information is associated with a copy of, or appears in connection with the communication to the public of, a work or other subjectmatter referred to in this Directive or covered by the sui generis right provided for in Chapter III of Directive 96/9/EC.

This implies that some extra legal protection is offered to the rightholder in using copymarks, the user will not be permitted to remove or alter any of this information without liability under the national legislation implementing the Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, and through this the WCT.

6.4   The Copymark Logo

The logo is, of course, just a suggestion for illustration purposes. It is inspired by the copyright notice, but distinct from this notice as it places the "C" within a hexago. The three types of unilateral licenses suggested, are "copyred", "copyorange" and "copygreen", the inspiration for these colours are obviously traffic lights, "copyred" corresponding to the most restrictive license, "copygreen" to the most permissive license, and "copyorange" to a warning that terms that are more specific are specified in the license.

The logo is similar to a system introduced at the Norwegian Research Center for Computers and Law for its website in 1995. The logo is here somewhat differently designed, the "copygreen", for example, is given a design indicated below:

As stated, the graphic design is not very important, the main consideration taken is that it should easily associate with "copyright" in the mind of the user, and a similarity to the well-

---

[37]   The *sui generis* data base protection.

known copyright notice is used for this purpose. For this purpose, it may be that the "C" itself is sufficient. One should note that Creative Commons has suggested a similar way of indicating licensing terms using a logo. Their alternative may be better than the one suggested here, as the logos have a symbolism designed to indicate their contents. On the other hand, they are rather dissimilar, therefore – perhaps – not alerting the user to their common basic function. The similarity between the logos is seen as an advantage of those suggested here, but there obviously is no particular preference invested in the design.

The word of the colours are spelled out, this may be to be recommended as a black-white display or print-out would not give the necessary distinction using colours alone.

The logo itself does not really explain anything. It is hoped that is by association leads the user consider copyright protection, but this is not necessarily so. A copyright notice, perhaps, does explain somewhat more by adding the word "copyright" to the logon ©. A similar addition could be made to the copymarks. On the other hand, the interpretation of the copyright notice requires the user to access the Universal Copyright Convention. In the digital environment, the logo, of course, is a hyperlink to the second element, an associated webpage.

6.5   *The associated webpage*

It is suggested that an associated webpage should spell out the terms of the unilateral license in more detail. As an example is included the copyright notice of my own paper "Data protection in Norway", as displayed if clicking on the green copymark:

One will notice that this is a license, which does not restrict the user. The user is granted the right not only to display the paper, but also to reproduce the paper in machine-readable or paper form. The user is also granted the right to secondary distribution.

In the introductory section, the example was limited to the exclusive right of reproduction. However, the rightholder also has an exclusive right to *distribution* of copies. This right is generally exhausted[38] in a copy purchased by the user. Therefore, the purchased book may be sold or given away. In the digital environment it is not the "original" copy downloaded by the user that is subject to secondary distribution, it may indeed be argued that to use the term "original copy" is not meaningful with reference to how protected material is made available in the digital environment. Therefore, if the user shall be permitted to make a copy available to a third party – even within the ambit of the typical "copying for private use" traditionally excluded from the exclusive right of the rightholder – one has to refer to the implicit license. It will be uncertain how far to stretch the implied license, especially with respect to commercial use. The suggested terms for the green copymark make it clear that there are no limitations apart from the moral rights.

One should emphasise that this is but an example for illustration. If the suggestion was to be deployed, one would have to examine the exact wording of the page, to make sure it is as clear as possible, and that the necessary rights are granted.

The operative legal clause is rather brief. A traditional copyright notice is included, though – as discussed above – it adds little to the substantive protection of the material.

In addition, there are other data on the paper, for instance an explicit identification of its URL if the user wants to point to the paper, and a contact address for the rightholder, making it comparatively easy for the user to contact the rightholder. Such contact may be for discussing licensing terms, but mainly – at least in this case – for discussing the issues of the paper.

The red copymark is very restrictive, for instance along the lines suggested below:

> **The author has authorised this paper for publication only in**
> NRCCL's Document Collection on the WWW.
>
> The user **may not** download this paper nor print copies of the paper for any purpose.
>   If the user wants to make further use of the paper, one should contact the right holder on the address stated below to obtain a license.

The red copymark only licenses the user for displaying the material on the screen, and permits no reproduction or re-distribution. For any further use of the material, the rightholder has to be contacted, but such contact is facilitated by the addresses given on the page. It is suggested that the red copymark would correspond closely to what follows by the default copyright legislation, and should be drafted to reflect this, but being somewhat more practical in its terms than the abstract language of the law. It may be seen as somewhat more restrictive, as it does not permit copying for private use. However, such copying presumes that the original has been made available to the public by the consent of the copyright holder, and this consent is here limited. It may therefore be argued that this represent the most restrictive license permitted under the general background law.

Similarly, the orange copymark gives an in-between license:

---

[38]   Exhaution is limited to exhaustion within the European Union, cf Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society Art 4.

In this case, it is suggested that the unilateral license gives the user the permission to reproduce that paper in machine-readable or paper form, but only for his or her private use. This will, it is believed; functionally give a license that correspond to the use which the default legislation gives a user purchasing a paper copy, but due to the nature of the digital environment, the user is granted somewhat more extensive right for reproduction than necessary with respect to a paper copy.

    In considering these suggestions, it is again emphasised that they are for illustration only. It is not claimed that they are appropriate with respect to the general default legislation, and the drafting would obviously have to be carefully considered if the suggestion was to be realised. However, with these disclaimers, the illustration should be sufficiently realistic to allow the suggestion to be considered.

6.6   *Meta-tags or other formalism to be interpreted by agents or other programs*

6.6.1   *XrML*

XrML[39] (eXtensible rights Markup Language) is a language to specify "rights". XrML is an XML-based syntax for specifying rights and conditions to control the access to digital content and services. XrML had its roots in Xerox Palo Alto Research Center. Digital Property Rights Language (DPRL) was first introduced in 1996. DPRL became XrML when the meta-language (used to construct the language) was changed from a lisp-style meta-language to XML in 1999.

---

[39] *XrML 2.0 Technical Overview*, Version 1.0**,** March 8, 2002.

License
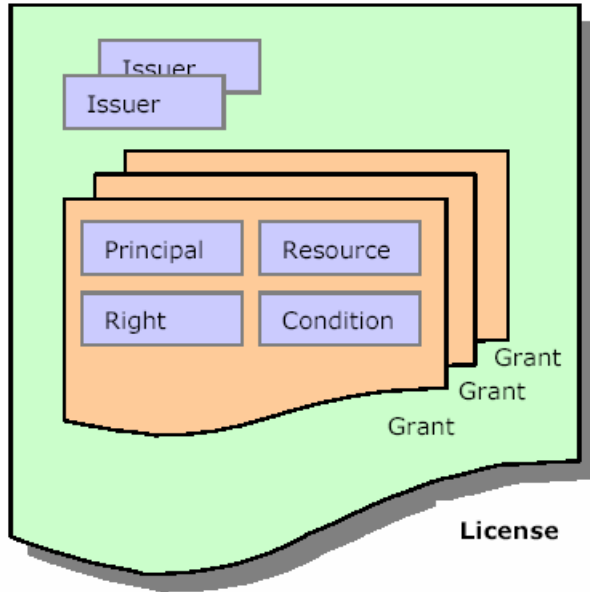
The structure[40] presumes an "issuer", which issue a grant, this grant identifying the principal to which the grant is extended, the resource for which the grant is extended, the right that is extended, and conditions on which the grant has been extended. Structurally, a grant consists of the following:

- The principal to whom the grant is issued
- The right that the grant conveys to the specified principal
- The resource against which the specified principal can exercise or carry out
- this right
- The condition that must be met before the right can be exercised

The principal may be given the right to issue further grants (sub-licensing, in copyright terms), and will then be an issuer. Certificates may be used to ensure the identity of the principals.

In this introduction, the formalism cannot be presented in a way making justice to the approach. We are mainly interested in elements that may be of use for the copymarks introduced above.

The "condition" may include specification for the interval in which the license is valid. It presumes some sort of enforcement mechanism that checks that the principal still holds a grant at the time the interval expires. The "grant" element may specify territory for which the license is valid, payment against the material being made available, *etc*.

The "content extension" replaces the abstract element "right", and includes a number of items relevant in with respect to copyright law:

Type of right	*Definition*
File management rights	
read	Represents the right to read the work from the repository.

---

[40]   The figure has been copied from *XrML 2.0 Technical Overview*, Version 1.0**,** March 8, 2002 sect 3.0.

Transport rights	
Loan	Represents the right to lend a work to another principal for a specific period of time. While the work is on loan, the original copy cannot be used.
Derivative work rights	
Edit	Represents the right to make changes to a work to create a new work based on the original work. Edit is like extract in that it creates a new work. It differs from extract in that it confers the right to make changes to the work.
Embed	Represents the right to include the work as part of a composite work. An embed operation places a copy of the work inside the composite work.

This tiny extract gives, perhaps, an idea of the richness of the XrML formalism. It enables full control of the material being made available to a "principal". But it would seem to presume the establishment of a repository that is controlled according to the terms in the grants by some sort of enforcing mechanism. There has been developed a trust model as part of the approach, in which several alternatives are open.

XrML is therefore a language for Digital Rights Management. It would in itself be an interesting task to evaluate the formalism and the control made possible through this formalism with the legal acts relevant to copyright law. No doubt this has been part of the development of the XrML, but a mapping of the legal concepts to the definitions has not been available.

In our context, it would seem that XrML certainly would serve the needs necessary for control. But it presumes that material is made available under the control of XrML, and that the users also avail themselves the system. It would presume that one co-operates within some agreed upon model for DRM, and this will obviously make the simplistic copymarks superfluous. For the interim situation, a more simple solution would have to be found.

6.7 Resource Description Framework (RDF)

RDF is a framework for expressing metadata, using a very simple basic structure of three parts:

- The subject, identified by its URL
- The predicate: the type of metadata, which may be title, creator or any other element, which for instance is specified in the Dublin Core (see above)
- The object: The value of this type of metadata, which may be the name of title of a story, the name or the creator, *etc*.

RDF is expressed in XML, and comes together to create the emerging "semantic web", that is a World-Wide Web where elements have been described according to the same semantic categories. This would make it possible for different programs to access the metadata and utilise the defined elements. It is easy to see that if documents (or "resources") are described by RDF, it would be easy to require that "resources" be retrieved that in the type "title" has the value "Jeff Noon", in this way retrieving only documents authored by Jeff Noon. Today, one only would be able to do a more superficial search, asking for documents in which the words "Jeff Noon" appeared adjacent, or something similar, not differentiating between documents *on* Jeff Noon and document *by* Jeff Noon.
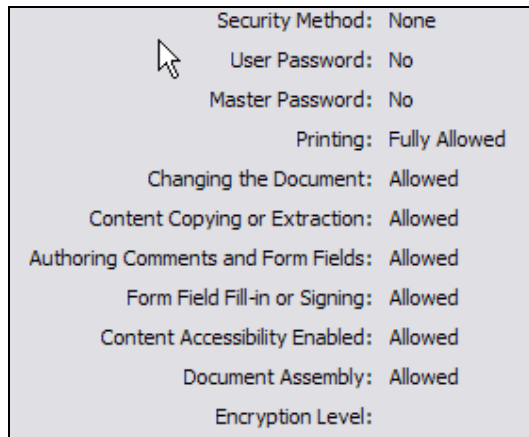
Creative Commons (see above) realises its machine-readable level of the license using RDF, embedding the RDF in HTML, and also refining some of the Dublin Core metadata elements (see above), especially the dc:rights.

The Semantic Web is still in the making. It is associated with many of the same challenges as indexing documents using a thesaurus or any other pre-defined vocabulary of some sophistication. As long as values have a high inter-personal invariance (like "title" or "author"), often called "strict criteria", it will be relatively easy to apply. However, including values with more inter-personal relativity – like "subject" (according to a pre-defined set of subjects), often called "assessment criteria" – one will expect to encounter the same challenges of inter- and intra-indexer consistency as other indexing efforts in the past.[41]

6.7.1   HTML meta-tags

The suggestion is to set up the browser in such a way that it would respect the copymarks embedded in meta-tags (or another formalism to be read by the agent). Ideally, the source site should check with the browser of the user before communicating the data to ensure that – at least for other settings than copygreen – the browser is enabled for the interpretation of the formalism specifying the terms of the license. If another program than a browser is to be activated by the user, correspondingly the check should be made.

This is to some degree related to the security settings used by viewers. For instance, the Adbobe Acrobat viewer will permit the user to read files, but will impose restrictions on the possibility of the user to manipulate the data being viewed, cf the example of an Acrobat menu below, which is an example of very lenient settings.

Security Method:	None
User Password:	No
Master Password:	No
Printing:	Fully Allowed
Changing the Document:	Allowed
Content Copying or Extraction:	Allowed
Authoring Comments and Form Fields:	Allowed
Form Field Fill-in or Signing:	Allowed
Content Accessibility Enabled:	Allowed
Document Assembly:	Allowed
Encryption Level:	

In our example, it is suggested to use the formalism developed within Alfebiite to specify the rights and obligations of the user. The formalism may have to be developed to handle this function, but it has a power of expression beyond the simple settings indicated above.

Also, one should consider developing the meta-tags beyond simple static statements of licensing terms. One may easily see that the author offers the material under a copyred license, but would extend the license to more generous terms in the case a payment was made. The meta-tags would then have to be developed to an agent residing with the material, and which could be engaged by the agent of a user in a negotiation with several dimensions, the licensing terms (the type of copyright relevant acts permitted, the duration of the license, *etc* would be one dimension, the amount of payment another obvious dimension – but there may be further less obvious dimensions).

---

[41]   For an entertaining as well as scholarly review of such efforts over time, see Umberto Eco *The Search for the Perfect Language,* Blackwell, Oxford 1995.

This paper ends only with the suggestion.[42] The mere outline of the interpretation of the different copymarks is indicated below:

The most restrictive copymark – copyred – includes the maximum functions that one would like to make operational. The license permits the user only to display the material (for sound, this would correspond to streaming), not to download the material to local hard disk or print the material locally. Also, copyred should be interpreted not to allow intermediate storage on proxy servers.

The meta-tags are directed towards agents, including simple agents like those of a search engine. However, copyred should not prohibit indexing by the agent of a search engine.

The prohibition of intermediate storage on proxy servers could be implemented, it is suggested, using the "expire" meta-tag, where setting the value for today may force an agent to refrain from copying the material.

New meta-tags would have to be introduced for "not download locally" and "not print locally". These would then have to be included in the browser to be interpreted directly by the browser of the user, disabling the button for "print" and "store".

The "copyorange" do not require any meta-tags for agents, browsers *etc*. The restricted acts are all directed to the user – the user is not allowed to exploit the material for commercial purposes under the unilateral license, but obviously the programs will not be able to determine this. The unilateral license will here have to rely on the loyalty of the user. But this will be combined with the possible sanctions in copyright law. If use is made of the material beyond what is permitted in the unilateral license, it will not only be a breach of contract, but actually an infringement of copyright. This can be claimed with greater certainty than in a situation where the terms of the license has not been detailed natural language, and gives the rightholder a somewhat stronger position in principle. In practice, it obviously will be difficult to enforce such licenses, as it is today difficult to enforce infringement of material made available under an implicit unilateral license.

If the suggestion of copymarks is deemed to be attractive, it would be necessary to carry out a feasibility study on how to implement these three meta-tags for browsers and agents, and define their syntax.

7 Summing up

This paper has explored some of the current ways in which digital material is handled with respect to copyright, especially the use of notices, meta-tags *etc* to control the material.

It has been observed that the Directive on Electronic Commerce seems to qualify meta-tags as relevant for determining the legal situation of the operators of proxy servers. It is indicated that this may be seen as the emergence of a legal policy that may be further developed.

A suggestion is made to introduce copymarks, which have three elements:

- A logo to be used on web-pages
- An associated natural language explanation of the content of the unilateral license implied by the copymark
- A formal representation of the conditions, implemented as meta-tags in the html-representation of the document, and to be read and interpreted by agents, browsers *etc*.

The suggestion include three levels of unilateral license, from a very restrictive to a very permissive license, the three categories being termed "copyred", "copyorange" and "copygreen". The category "copyred" is the one that require meta-tags similar to the current "restrict" tag, and disabling the print- and store-functions of the browser.

---

[42] But further work is considered by Alfebiite partners.