

# LEGAL ISSUES OF AGGREGATING AND CURATING INFORMATION FLOWS: THE CASE OF RSS PROTOCOL

Dimitra Karydi<sup>1</sup>, Ioannis Karydis<sup>2</sup>

<sup>1</sup>Attorney at law, Corfu, Greece, dimitra.karydi@gmail.com

<sup>2</sup>Department of Informatics, Ionian University, Corfu, Greece, karydis@ionio.gr

## Abstract

The ease of content distribution through the web in addition to the ubiquitous mobile devices that not only consume but also create content, have allowed for an ever increasing availability of information. Moreover, following the paradigm of social networking, this information is considered of variable flow rate and transient in nature. The resulting information volume has, paradoxically, lead to distracting and negatively impacting productivity and decision-making. To compensate, content curation has emerged as a prominent process engulfing a range of activities and procedures done to manage and enhance information and inform interested parties. In the context of this work the utilisation of widely employed syndication protocol Really Simple Syndication (RSS) is examined. During the curation key processes of collection, storage, customisation, inference extraction and redistribution of information flows' entities a litany of legal issues arise. As these issues present an increased complexity, this research focuses on the presentation of the associated to information flow curation legal issues as a starting point of further research, bridging thus the two disciplines.

## 1. Introduction

Nowadays, multi-sensor mobile appliances are widely available at affordable prices. These devices are, more often than not, capable of connectivity to the internet with prominent such examples the smartphones and tablets that usually include more than one such methodologies (i.e. wi-fi & cellphone network). In addition, these devices include a plethora of environment sensing capabilities, with still-image/video camera inarguably the most commonly used content creating sensor. Accordingly, it has become easier than ever to create and distribute content through the web, thus adding to the information already available.

Obviously, the resulting information volume is continually increasing. A mere example of the sheer volume produced by users daily is clearly shown by the statistics of the popular web video streaming service YouTube [Youtube, 2014] that receives 100 hours of video every minute [Youtube statistics, 2014].

The aforementioned ease of content creation and distribution in addition to the highly available increased WWW bandwidth, that can deliver large amounts of content very fast, has lead to the widespread adoption of the notion of content not as something static but as information that is in an ever changing state. Although this notion is far from new in research or specific domains [Golab and Özsu, 2003], its far-reaching effect on users' understanding is new. Accordingly, users are overwhelmed by information flows that have a variable flow rate and are of a transient character. A

number of existing protocols support the exchange of such information flows, e.g. the Really Simple Syndication (RSS) [RSS, 2014], Atom [Atom, 2014] and The Open Graph protocol [The Open Graph protocol, 2014] to name a few.

In contrast to common belief, the increase of available information and the treatment of information dissemination as a flow have led to the distraction of information consumers in addition to negatively impacting their productivity and decision-making mainly due to the poor penetration and, in some cases, low performance of data mining techniques that usually handle such information volumes.

To compensate for these unique information availability circumstances, web content curation has emerged as a prominent process engulfing a range of activities and procedures done to create, collect, manage, validate, enhance and customise information and subsequently inform interested parties of, relative to their needs, content. The curation process utilises human resources with domain specific knowledge while in cases these resources may be ICT-assisted in order to achieve the purposes of curation.

During the key processes of curation a litany of legal issues arise. Despite their severity, these issues present an increased complexity to implementing software engineers that might, due to the un-associated background, be unaware of the legal details pertaining to their actions.

### **1.1. Motivation and Contribution**

Bearing in mind the aforementioned extended capability of users' to create and distribute content and the widespread emergence of the curation process, the contribution of this work is the identification of legal issues pertaining to the common core practices of the information flow curation process.

The contribution is henceforth examined under the Greek legislation framework, while, where deemed necessary, directions as to other legislations are additionally provided.

It should be noted that in no case do the contents of this work constitute formal legal advice as each real case should be properly advised for, based on its own merits.

In the context of this paper, the term 'information' is also seen from a copyright law perspective, namely bearing the notion of a 'work' and not only pure data, as it will be presented further below. Moreover, this paper will focus mainly on the authors' rights (copyright law) while a very brief reference will be made to the related rights.

The rest of the paper is organised as follows: Section 2 describes background information and related work on the key themes of the work and Section 3 details the activity of storing the sources' information during the curation process. Next, Section 4 presents the activity of knowledge mining in curated informational flows, while Section 5 details the redistribution of the curation resulting information. Finally, the paper is concluded in Section 6.

## **2. Background Information**

In this section we present background information related to information flows, content curation, the RSS protocol as well as an introduction to the Greek copyright law related to the theme of our work.

## 2.1. Information Flows

Information flows, or data streams [Muthukrishnan, 2005] are input data that have a definite beginning, no foreseeable end, variable flow rate and a transient character as to their data. The transient character refers to the availability of content from the source within a source defined time span at any given time of content request.

Following the aforementioned informal definition of information flows, it is clear that these continuously and rapidly produce information over time that leads to large amounts of data. Accordingly, the young and evolving data stream research discipline [Aggarwal, 2007] offers models and algorithms in order to manage such information.

In the context of this research, information flows refer to sources of information that make their content available following the paradigm of data streams.

## 2.2. Content Curation

Information curation is far from new a process. In general terms, curation is the process of selecting existing and/or creating, validating, enhancing and distributing information. In this sense, it is a perpetual process done by almost all humans every day.

Figure 1 shows an abstract information flow diagram of the key actions involved during curation while a more detailed lifecycle model of the curation process is available in [Higgins, 2008].

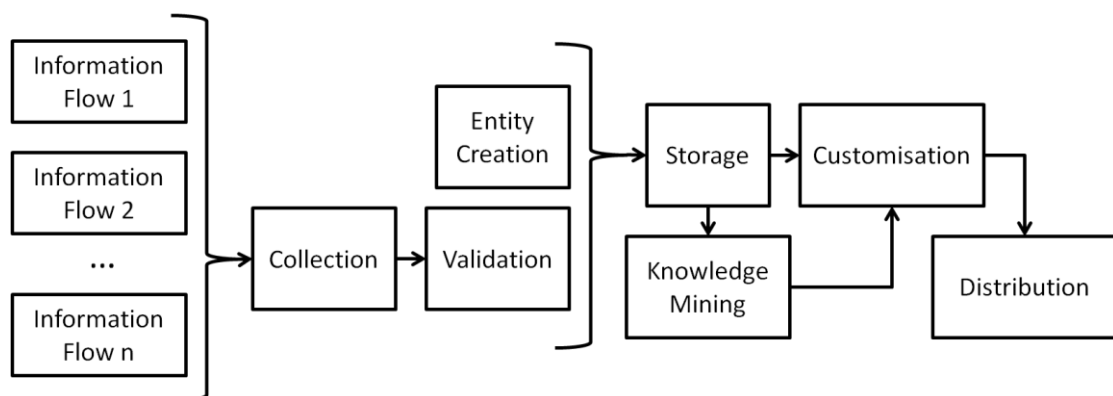


Figure 1. An abstract model of an information flow curation process

The notion of curation applied to digital information was originally used in the “Digital Curation: digital archives, libraries and e-science seminar” organised and sponsored by the Digital Preservation Coalition (DPC) and the British National Space Centre (BNSC) in 2001 [Beagrie, 2006]. Although the term is under constant evolvment, as may bear alternate meanings for different disciplines, most definitions revolve around maintaining and adding value to trusted information for current and future use [Yakel, 2007].

The aforementioned adding value step, shown in Figure 1 as “Knowledge Mining”, is one of the main processes of the curation procedure. During this step, informational entities are processed in order to extract the relation entities exhibit between them, adding thus new information [Wolff and Mulholland, 2013].

Digital content curation targets the long-term sustainability of information by (a) preserving and protecting information, (b) offering alternative methods for access to information, (c) promoting information re-use, (d) supporting information aggregation and (e) transforming information to ensure its meaningfulness, among other functions [Abbott, 2008].

In the context of this work, information flow curation refers to the curation process of informational flows.

### **2.3. RSS**

RSS represents a “lightweight XML vocabulary for describing metadata about websites” [King, 2003]. The acronym originally stood for RDF Site Summary and evolved to Rich Site Summary, although it is usually referred to as Really Simple Syndication. This is due to its widespread use [RSS Usage Statistics, 2014] for automated data syndication between feeds (sources or channels), that update content frequently, and content consumers [Sikos, 2011].

The utilisation of the XML file format ensured the interoperability of the information exchange allowing thus RSS to evolve into a very popular means of sharing content. Content consumers utilise software such as RSS/feed readers or RSS aggregators in order to acquire and present feed content [RSS, 2014]. In addition, the notion of feed subscription, by means of the aforementioned consumers’ software, alleviated the need for content consumers to manually check the source and thus extended further the popularity of RSS.

In the context of this work, the utilisation of RSS is not only supported as a means for information aggregation but is also assumed for the distribution step of the curated information.

### **2.4. The basics of Greek Copyright Law**

The rights of the authors on their original intellectual creations are protected under Greek Law 2121/1993 (hereinafter Greek Copyright Law/GCL) [FEK A’ 25/1993; Greek Copyright Law, 2014]. Accordingly, protection is ascertained to the author (creator) of an original work, such as a literary work or a musical work, which deprives others from engaging in certain uses of that very work, for a defined period of time, without the author’s consent [Articles 1, 2, 3, 4, 29, GCL]. The Greek copyright system is oriented towards the ‘droit d’ auteur’ system of civil law countries. Generally speaking, the author of the work is the first owner of the copyright in it, although there are provisions prescribing differently [Chapter 2, GCL]. It has to be stressed that copyright protection is not vested in ideas but only on their original way of expression [Article 2(1), Berne Convention for the Protection of Literary and Artistic Works, 1886, Paris Act 1971 (hereinafter Berne Convention)].

The protected subject matter under Greek Copyright Law, i.e. the term ‘work’, refers to any original intellectual literary, artistic or scientific creation, expressed in any form [Article 2 par. 1, GCL]. In other words, for a work to be protected under Greek law, it has a) to be the result of intellectual endeavour, b) to be original, c) to be expressed in any form and d) to fall within the notion of a literary, artistic or scientific creation. The Greek legislator provides for a list of protected works, such as texts, musical compositions, theatrical works, audiovisual works, that is not an exhaustive one but one indicative. Apparently, the Greek copyright system, follows the flexible open-list scheme of the French law contrary to the United Kingdom’s rigid close list which confers protection only on the eight categories specified in the English Copyright Act [Aplin & Davis, 2009; Article L 112-1 & L 112-2, French Intellectual Property Code 1992; Section 1(1), Copyright Designs and Patents Act 1988 (hereinafter CDPA)].

As regards to originality, except for the computer programs [Article 2 par.3, GCL; the Council Directive 91/250/ECC] no definition is provided in the Greek Copyright Law. For the recognition of originality, Greek legal theory and jurisprudence lean towards the principle of ‘statistical uniqueness’, according to which a work is original if, under similar circumstances and with similar aims, no other person could potentially create the same work or a very similar one [Despotidou, 2011; Μαρíνος, 2000; Κοριατοπούλου-Αγγέλη, 2008; Greek Court Judgments ΕφΑθ 4793/2009, ΕφΑθ 885/2009, ΕφΑθ 5863/2008]. Moreover, the work should at least show a minimum level of ‘creative level’ and a ‘particular individuality’, stemming from the creator’s intellectual endeavor, which distinguish it from routine, common, self-evident and similar known creations [Despotidou, 2011]. Aesthetic merits or the purpose of a work are not taken into consideration for the recognition of originality [Κοριατοπούλου-Αγγέλη, 2008; Greek Court Judgment ΜΠρΤριπ 261/2012].

The Greek copyright legislation does not adopt the English law originality criterion where a work is original when this is the result of the author’s ‘skill, labour and judgment’ [*Walter v. Lane*, AC 359 (1900); *Cramp v. Smythson* AC 329 (1944); *University of London Press v. University Tutorial Press* 2 Ch 601 (1916)], nor the American older test of ‘sweat of the brow’ [*Jeweler’s Circular Publishing Co. v. Keystone Publishing Co.*, 281 F. 83 (2d Cir.); See also *Feist Publications, Inc. v. Rural Telephone Service Co.*, 499 U.S. 340 (1990)]. Furthermore, it does not entirely embrace the traditional French subjective approach of originality which required for a work to bear the imprint of the author’s personality in order to be original [Gervais, 2002; *Witt v. Pachot* (1986)]. The notion of originality under Greek law is a combination of the aforementioned theories requiring a minimum level of ‘creative level’ and a ‘particular individuality’ in a work [Κοτσίρης, 2005; Μαρíνος, 2000].

Authors are granted a bundle of exclusive and absolute rights, namely economic and moral rights [Articles 1, 3, 4, GCL], where economic rights are linked to the economic use of the work while moral rights refer to the special personal link between the author and his work [Dworkin, 1995; Stamatoudi, 1997; Swack, 1988]. These are two independent sets of rights where moral rights remain with the author even after the transfer of the economic rights [Article 4 par.3, GCL].

The author of an original work, or anyone authorised by the author, has a) the right to fix and reproduce the work by any means and in any form, in whole or in part, b) the right to translate the work, c) the right to make an arrangement, adaptation or other alteration of the work, d) the right to distribute the original work or its copies to the

public in any form via sales or otherwise, e) the rental or public lending right, f) the public performance right, g) the broadcasting or rebroadcasting of the work to the public by any means or process, h) the right to make available the work to the public by any means of process and i) the right to import copies of the work [Article 3, GCL].

As for the moral rights vested in the author, he has a) the right to decide the time, place and way in which the work will be accessible to the public (publication right), b) the right to be acknowledged as the author of the work (paternity right), c) the right to prohibit any distortion, mutilation or other modification of the work including the presentation of the work in any way that prejudices the reputation of the author (integrity right) d) the right to access the work irrespective of the ownership over the physical embodiment of the work, e) the right to cancel a contract transferring the economic right or an exploitation contract or license of which his work is the object, subject to payment of material damages to the other contracting party, for the pecuniary loss he has sustained when the author considers such action to be necessary for the protection of his personality because of changes in his beliefs or in the circumstances [Article 4, GCL].

In general terms, the duration of copyright protection under Greek Copyright Law is 70 years *post mortem auctoris* [Chapter 5, GCL]. Greek Copyright Law also affords protection to databases provided they are original, meaning the author's own intellectual creation, by virtue of the selection or arrangement of their contents [Article 2 par. 2(a), GCL]. Moreover, *sui generis* protection (database right) is provided to databases. According to Article 45A GCL, which embodied the EU Directive 96/9/EC on the legal protection of databases (hereinafter Database Directive), protection is granted to the maker of a database not on the basis of the originality in selection and arrangement but due to the substantial investment put into the obtaining, verification or presentation of the contents of the database. Greek Law 2121/1993 includes provisions with regard to related rights, also known as neighbouring rights.

### **3. Storing Sources' Information**

This section discusses the notion of content aggregation. In the context of this research, the aggregation refers to the purposeful action of collecting information as a preparation task in order to identify the super-set of information elements that the curation procedure with subsequently process.

The aggregation step is an integral part of the curation process as it is required in order for the next steps of the curation procedure, namely the knowledge mining and distribution.

The specifics of the aggregation, as far as the ICT procedures are concerned, can greatly vary according to the domain of application. Nevertheless, a coarse categorisation is possible based on whether the collected information is planned to be stored indefinitely long or for just the exact time required in order to process it. Accordingly, storage of the collected information can be either persistent or transient. In both cases, the resulting mined knowledge is stored, and the collected data, in the former case are also stored, while in the latter are discarded immediately after the knowledge mining.

Based on the aforementioned decision as to the persistence or transiency of the collected data, the curation process, during its final step, can always distribute the knowledge mined while in the former case the collected data as well.

### **3.1. Legal Aspects of Information Storing**

Copyrighted works are protected not only in the analogue but also in the digital world. Greek legislation has incorporated the EU Directive 2001/29 on the harmonization of certain aspects of copyright and related rights in the information society and has ratified the WIPO Internet Treaties [Article 81 Greek Law 3057/2002; Greek Law 3183/2003; Greek Law 3184/2003]. Accordingly, protection is afforded to works transmitted throughout the internet that fulfil the aforementioned criteria of copyright law. For instance HTML coding and metadata may be protected as literary works, sound files are sound recordings comprising of musical works and are accordingly protected, while digital newspapers articles are literary works duly protected [Pedley, 2003].

Unless copyrighted works are lawfully obtained, e.g. via online sales, Internet users should bear in mind copyright rules when downloading works. Storing copyrighted material falls within the reproduction right of the copyright owner since a copy of the work, even if this is a single one, is made and saved in the computer's memory. Article 3 par.1(a) of the Greek Copyright Law, in conformity with Article 2 of the EU Directive 2001/29/EC, prescribes that the copyright holder has the exclusive right to authorize or prohibit any direct or indirect, temporary or permanent reproduction of his work by any means and in any form, in whole or in part. Thus, copyright holders are vested with the reproduction right of the work also in the digital world. So, the general principle is that there is no implicit licence to reproduce material available on the Internet [Pedley, 2003] and therefore authorization is required.

Nevertheless, due to the way technology works, Greek Copyright Law in compliance with Article 5 of the EU Directive 2001/29, provides for an exception to the reproduction right for certain temporary acts which are transient or incidental in nature, which constitute an integral part of a technological process and whose sole purpose is to enable: a) a transmission in a network between third parties by an intermediary or b) a lawful use of a work or other protected subject-matter, and which have no independent economic significance [Article 28B, GCL]. Accordingly, the acts of caching, including second level caching and proxy server, routing and browsing, that fulfil the aforementioned requirements as well as the 'three-step test' [Article 28C, GCL; Article 9 par.2, Berne Convention], fall within this exception [Τσίγκου, 2008].

Materials that are not copyrightable either because they do not meet the relevant prerequisites, or the duration of copyright protection has expired, rest in the public domain and can be freely stored and used as regards to copyright law.

Moreover, certain exceptions/limitations of the economic rights are set forth by the Greek copyright law, within an exhaustive list [Μαρίνοϋ, 1995], in an attempt to draw a balance between the rights of the copyright owners and the public interest, particularly for reasons of social policy and promotion of free flow of information [Chapter 4, GCL; Μαρίνοϋ, 2000]. Acts that fall under these exceptions/limitations, although copyright infringing in principle, are ultimately excused without the need to get permission or pay remuneration. Nevertheless, these exceptions/limitations should

comply with the ‘three-step test’, meaning that they can only be applied in certain special cases which do not conflict with a normal exploitation of the work or other protected subject-matter and do not unreasonably prejudice the legitimate interests of the rightholder [Article 28C GCL; Article 9 par.2 Berne Convention].

One of the strictly specified exceptions/limitations is the private use exception, according to which reproducing a copyrighted work is permitted when this is made for private use [Article 18, GCL]. The underlying rationale for having this exception is the difficulty to control such acts as much as the promotion of individuals’ cultural engagement [Μαρίνος, 2000]. The meaning of private use refers to the copying made by the user for his own utilization or for his narrow circle of the family and his immediate social circle [Article 3, par.2, GCL]. There is no private use, and therefore copyright infringement takes place, when the reproduction of the work aims at the distribution of the copies to the public or at the public performance of the work. In other words, the sole aim of the reproduction should be the private use of the copy and not any pecuniary exploitation of the copy. Moreover, making many copies of the work comes up against the ‘tree-step test’ which points that there are limits to the private use exception [Μαρίνος, 2000]. It is also noteworthy that the law does not permit the reproduction of electronic databases for private uses [Article 3 par. 4, GCL].

In addition, the law refers to the reproduction for private use of a lawfully published work. Yet it is debatable whether it is important for the exception of private use how the user obtained the carrier of the copyrighted work or if the reproduction is made from a lawful copy of the work [Μαρίνος, 2000].

In all, when it comes to storing copyrighted material, this may be justified under the private use exception. But that means that the copy of the work will not be made available to the public in any way, for instance via uploading it on the web. Taking into account that the ultimate goal of information curating is sharing the results of such process and enhancing the flow of information, be this material copyrighted or not, the exception of private use of the copyrighted works is presumably not a viable defence. Put it simply, it seems that in the process of curating copyrighted material, the private use exception does not offer a safe harbour for digital curators who are opting in sharing the material obtained.

#### **4. Knowledge Mining**

Knowledge mining refers to the process of applying methods and algorithms, on the collected data, in order to discover new, interesting, not obvious and very hard to manually extract knowledge. Such processes could be the identification of relations that organise flow entities in groups (classification & clustering), the selection of unusual flow entities (anomaly detection), the prediction of information in flows (association rule mining) as well as many other.

The organisation of flow entities in groups can be of great significance to the ICT-assisted curation as it can both produce multiple organisations of existing entities based on different criteria and in addition categorise newly received into the existing categories. The selection of unusual entities is also very interesting to the ICT-assisted curation process as it identifies entities that are not common and thus highly important to the curation process. Moreover, the ability to discern association rules in flow



entities can support the curator's work by allowing the conditional prediction of entities to arrive and thus improving the curated content's customisation to the petitioner's request.

Accordingly, the curation process may not only include the collection and validation of information flow entities, but also their enhancement by incorporation of additional relevant information. This added information is the result of the knowledge mining and can mostly be produced by occupying a bird's-eye view position as to the information flow's contents and the application of the aforementioned methods and algorithms.

#### **4.1. Legal Aspects of Knowledge Mining**

It is hard to say in advance whether and which copyright rules could possibly be violated on the process of mining new information from collected copyrighted materials. Every case is decided upon its on specific technical and legal merits. Generally speaking, where 'mining' refers to the process of applying methods and algorithms on the collected copyrighted materials in order to reach conclusions and deduce new information, it is unlikely that copyrighted infringement could be maintained since the outcome of the mining process does not reproduce the work or conflict with any of the other rights of the copyright owners on the copyrighted works. In other words, this ICT process is using the copyrighted material aiming to reach somewhat theoretical observations and results, for instance the recurrence of a word in an news article, activities that do not seem to interfere with the absolute rights of the copyright holders.

Nevertheless, on the occasion that the 'mining' process is one involving any of the exclusive rights of the copyright holder of the work, for instance, a substantial part of a work is reproduced throughout this procedure, then permission should be obtained. It is rather unlikely that a curator's mining activities aim to be only for private use whereas the free flow of information is usually his goal, so the aforementioned private use/exception does not seem to fit in this situation.

Moreover, in case 'mining' of copyrighted materials means the creation of derivative works, then the permission of the copyright owner of the initial work is required. According to Article 3 par.1(c) GCL, only the copyright holder has the right to authorize or prohibit the arrangement, adaptation or other alteration of his work, meaning the making of derivative works. Thus, if the 'mining' process leads to the creation of a derivative work, which should be also original and fulfil the criteria prescribed in Article 2 GCL, then in principle, authorization of the copyright owner of the initial work is needed for the pecuniary exploitation of the derivative work, such as its communication to the public via the Internet. If the derivative work is kept only for private use, then permission is not necessary [Μαρίνος, 2000; Τσίγκου, 2008]. On the occasion that the derivative work is a mere reproduction of the initial work, then the reproduction right of the copyright owner is infringed. As regards to the similarity between the initial and the derivative work in case of a copyright infringement allegation, it is difficult to provide for a clear-cut rule given that every work has somewhat been influenced by previous ones. Therefore, each case is examined on its own merits where not only quantitative but also qualitative factors of the two works are considered [Μαρίνος, 2000].

It should be noted that a curator's derivative work itself, if original, can be protected under copyright law. Considering that the process of curating goes further than just aggregating information, in more elaborated activities such as customisation and knowledge mining, derivative works may come out of this process.

In all, considering the process of 'mining' new information from a copyright law perspective, a case by case assessment should be followed. It is worth noticing that in the U.K., in June 2014, various changes of the copyright law will take effect. One of them is a new 'text and data-mining' exception permitting computer-based analysis of copyright material for non-commercial research without the requirement of prior permission from the copyright holder. In this way, researchers on non-commercial projects will be able to reproduce materials for the technical process of data mining, provided they can lawfully access the relevant works [Intellectual Property Office, 2014].

## **5. Redistribution**

The redistribution part of the curation process, although not obligatory, is essentially almost always performed, except for the rare cases that the curation, in its entirety, addresses the need for private information consumption.

In all other cases, the curator's work is concentrated in informing interested parties of the relative to their needs content after the content's creation or collection and validation.

Although the methodology for the redistribution can take more than one forms, in the context of this work, we examine one of the prominent and widespread such methods, the Really Simple Syndication (RSS) protocol, as described in Section 2.3.

### **5.1. International legal approach of RSS feed**

As already mentioned, RSS constitutes a protocol for information exchange that can be utilized for obtaining and redistributing content on the Internet after the curation process has taken place. It has to be stressed that from a legal standpoint, there is no certainty on how to deal with the various issues that may appear with respect to RSS functionalities and copyright. No straightforward answers can be given, as it had happened every time technology came up with something new. Let's not forget the Napster case and the ongoing challenge of file sharing systems. When Napster and its centralized system of file-sharing among its users was found guilty of contributory and vicarious infringement of copyright, decentralized systems of file sharing such as Grokster and Kazaa emerged where contributory or vicarious liability was harder to be established [*A&M Records, Inc. v. Napster, Inc.*, 114F, Supp.2d 896 (ND Cal. 2000); *Metro-Goldwyn-Mayer Studios Inc et al. v. Grokster Ltd et al.*, 259 F. Supp. 2d 1029 (CD Cal 2003)]. Thus, every case should be decided on its own merits.

Generally speaking, RSS feed users, such as curators, should be wary of their acts concerning copyright rules. They should comply with relevant copyright provisions, the terms and conditions of the provider's RSS feed and where not available, they should ask for proper legal advice on their particular case.

The fact that RSS content is provided in a syndication friendly form, does not necessarily mean that this very content is not protected under copyright law. Under Article 3 par. h GCL, copyright infringement takes place when making available a work on the Internet without the rightholder's permission [Section 20 CDPA; Section 106(3) 17 U.S. Copyright Act 1976]. The curators' argument that RSS feed carries an implied licence to redistribute the content on the Internet, since this is the underlying rationale of the protocol, is dubious. Internet is not a 'law-less' environment. Relevant copyright rules apply adapted to the web's specific features. Assuming that any content available on the Internet comes with an implicit copyright licence of usage is not accurate.

The concern of the application of copyright on digital curators, such as Google News which makes available its content through RSS, has come up on an international level. In the United States (U.S.) case of *Agence France Presse v. Google, Inc.* [No. 1:05CV00546 (GK) (D.D.C. Apr. 29, 2005)], AFP alleged, *inter alia*, that Google infringed AFP's copyright by providing the headline, lede and photo of AFP's articles on the Google News website. Google News claimed that, under U.S. law headlines do not fall within the copyright protected subject matters. Eventually, the parties reached an agreement outside the court [Kimberly Isbell & the Citizen Media Law Project, 2010].

In a somewhat similar case, Associated Press (AP) accused All Headline News, a news aggregator, *inter alia*, for copyright infringement on the basis that the latter copied stories found on the internet or rewrote them and sold the content to its subscribers [*The Associated Press v. All Headline News Corp.*, No. 1:08-cv-00323-PKC (S.D.N.Y. Jan. 14, 2008)]. In the end, a settlement was reached where All Headline News was obligated to stop using AP's material and paid an amount of money to reconcile the claims for previous unauthorized uses of expression and content [Kimberly Isbell & the Citizen Media Law Project, 2010].

In 2008, another similar complaint was filed this time by GateHouse Media against The New York Times Co. [*GateHouse Media, Inc. v. The New York Times Co.*, No. 0-12114-WGY (D. Mass. filed Dec. 22, 2008)]. GateHouse maintained that The New York Times unlawfully reproduced the headlines and ledes from its Wicked Local websites on the latter's Boston.com website. Likewise, the parties settled before the trial. In all these cases, a settlement was reached, so there was not any final court ruling to enlighten the relevant legal merits [Kimberly Isbell & the Citizen Media Law Project, 2010].

According to a 2013 article in *The Jerusalem Post* [Yonah, 2013], in an Israeli case of RSS feed copyright infringement, News1 website was accused of making available on the Internet visual and textual materials from Tomer Ofaldorf's RSS feed. The plaintiff argued that although RSS aims to enhance the circulation of content, yet RSS feed users have no specific right to do anything else with the content but to use it for themselves. The defendant claimed that by making available not only headlines but also full content via RSS feed, the RSS feed provider seemed interested in promoting his content also through its redistribution on other websites. After all, the defendant's argument goes, RSS is all about upholding the free flow of information. The Court took into account that the defendant provided for acknowledgment of Ofaldorf's content and that it took down the posts once Ofaldorf complained for the lack of authorization. The Court did not assert that syndicating content from a third party's RSS feed is not a copyright infringement, but in this specific case News1's postings

were legal. Most importantly, the Court indicated that, by choosing an automatic and aggressive means of sharing content such as the RSS protocol, the plaintiff gave the reasonable impression that he did not oppose to such a use of his content. Eventually the Court ruled in favour of the defendant News1 website. This case illustrates aptly the uncertainty that exists concerning the application of copyright law on exchanging information via RSS feeds. Unfortunately, the authors of this paper have no any additional details on this court ruling which could further illuminate its legal merits.

## **5.2. RSS feed and Greek copyright law**

A user that obtains the RSS feed of a third party on his computer in order to use it for himself, is more likely to fall under the aforementioned private use exception (see section 3.1 above), assuming that the content of the RSS feed is copyrighted material such as a newspaper article. For instance, the terms and conditions of the Greek website *www.lawnet.gr* specifically mention that its RSS feed is available for private uses and not for commercial purposes. Likewise, the New York Times provide for quite lengthy and detailed terms and conditions for the usage of its RSS feed, stating, *inter alia*, that it is indented for personal use or as a part of a non-commercial blog [www.nytimes.com]. The tricky part comes with the redistribution of the obtained RSS feed content.

Article 3 par.1(h) of the Greek Copyright Law has adopted the ‘right of communication to the public’ indicated by Article 8 of the WIPO Copyright Treaty. Accordingly, the author of a work has the explicit right to authorize or prohibit the communication to the public of his works, by wire or wireless means or by any other means, including the making available to the public of his works. In other words, the copyright owner of a work has the right to control the distribution of his work on the Internet, which takes place when uploading his work on the web [Greek Court Judgment ΜΠρΑθ 3431/2002].

### **Redistribution of whole content**

In case a curator engages in the redistribution of the whole content of a third party’s RSS feed, then generally speaking, it is highly likely that there will be a finding of copyright infringement of the reproduction and communication to the public right of the copyright owner, provided that the content is a copyrightable subject matter. Although there is no specific provision in the Greek copyright law pertaining to RSS feed, it does not necessarily mean that copyright law does not apply on this occasion since, as reiterated, the online world is subject to copyright rules as well. To the knowledge of this paper’s authors, there have not been any Greek court rulings on this specific matter. There has been one involving an RSS feed aggregator but it rather focused on criminal offenses not related to copyright issues [Γιαννόπουλος, 2013]. It remains to be seen the direction that Greek jurisprudence will point at in this blurry legal area.

### **Redistribution of parts**

Now, in case a curator engages in the redistribution of a very small part of a third party’s RSS feed, such as the headline of an article, its lede or a thumbnail photo, and then places a link to the original source, it is less likely to find copyright infringement

comparing to the previous situation. Yet, there is no definite answer and each case is decided upon its particular circumstances. Some considerations on this topic are unfolded right below.

**Titles & excerpts** According to Greek legislation and jurisprudence, titles of newspapers or short extracts can be protected under copyright, provided they are original works of authorship [Despotidou, 2011; Τσίγκου, 2008; Greek Court Judgments ΠΠρΑθ 7147/2000, ΜΠρΑθ 4526/1988]. The fact that these are not explicitly included in Article 2 GCL, does not make them unfit for copyright protection. As aforementioned, the list of copyrightable subject matters is not an exhaustive one, and since these are not specifically excluded from the list, titles, books and newspapers headlines, or short excerpts can be copyright protected. It is reasonable to assume that one-word titles or slogans hardly ever meet the requisite level of originality but this does not entail their exclusion from copyright protection, if they are original works.

The principle stating that ‘the law does not concern itself with trifles’ (*de minimis* principle) does not seem to have been upheld by Greek courts [Despotidou, 2011; U.K. cases: *Sinanide v. La Maison Kosmeo*, 139 LT 365 (1928); *Francis Day & Hunter Ltd v. Twentieth Century Fox Corporation Ltd*, AC 112 (1940); *News Group Newspapers Ltd v. Mirror Group Newspapers Ltd*, (1986); *IPC Magazines Ltd v. Mgn Ltd*, FSR 431 (1998); U.S. case: *Southco Inc v. Kanebridge Corp*, 324 F 3d 190 (2003)]. In other words, under Greek copyright law, the size of the work does not matter when determining its originality [Despotidou, 2011].

Hence, RSS feed users, such as curators, may find themselves troubled with copyright offenses when they redistribute online the title of a text or a brief excerpt, if these are deemed original. In the Belgian case of *Copiepresse v. Google Inc.* [*Copiepresse v. Google Inc.*, 2007], the Court found, *inter alia*, that the titles and three lines of the linked articles provided by Google in its news website were deemed copyrightable materials and Google committed copyright infringement by using them without the rightholders’ permission [Xalabarder, 2012].

**Photographs** It may happen that the RSS feed includes a photo that is also syndicated when using the content of the RSS feed provider. Photos, if original works of authorship, are also protected under Article 2 par. 1 GCL [Κοριατοπούλου-Αγγέλη, 2008; Greek Court Judgment ΜΠρΘεσ 40026/2006]. Thus, copyright infringement may take place with respect to the redistribution of photographs as well.

It is worth noting that using thumbnail photos may also give rise to copyright infringement issues. For instance, in the U.S. case *Leslie A Kelly v. Arribasoft Corp* [280 F.3d 934 (9th Cir., 2002)] the defendant was accused of illegally reproducing, by using in his visual search engine, many images in a thumbnail format. The Court found that there was infringement of the copyright owner’s right to display his photos publicly, but this was excused under the ‘fair use’ doctrine [Flint at al., 2006; WIPO, 2002].

Under Greek copyright law, the reproduction right afforded to the authors refers to any reproduction of the work no matter the size or dimensions of the copy [Μαρίνος, 2000]. Considering that thumbnail photos are photos in smaller size, providing for thumbnail photos on the Internet, without proper permission, may constitute infringement of the reproduction and making available to the public right of the author. Additionally, infringement of the moral rights of paternity and integrity could

be maintained. The paternity right would be infringed in case there is no proper attribution of the photo to its creator. The integrity right because, as reiterated, the author has the right to prohibit any distortion, mutilation or other modification of his work, and altering the dimensions of his work, as in a thumbnail photo, could be included in this right [Μαρίνος, 2000].

**Links** As regards to linking, normally this creates no problem when the link provides access to the home page of a third party's website. Arguably this resembles to the traditional method of referencing [Καλλινίκου, 2005; WIPO, 2002]. Nevertheless, problems arise when using deep linking; meaning links that bypass the third party's website homepage, directing the user to secondary material of the website. The user of the links may assume that the new information provided by the deep linking is part of the initial website he was surfing on. In this way, there may be infringement of the public display right or the communication to the public right of the copyright owner of the second website [WIPO, 2002]. The paternity right of the creator of the second website is potentially infringed as well [Τσίγκου, 2008]. In addition, the moral right of the author to prohibit any distortion, mutilation or other modification of his work and any offence due to the circumstances of the presentation of the work in public could be violated by the method of deep linking [Καλλινίκου, 2005].

In other European jurisdictions, infringement in the database right (*sui generis* right) of the owner has also been invoked with respect to the use of deep linking [*Danish Newspaper Publishers' Association v. Newsbooster.com ApS*, Denmark Bailiff's Court, 2002]. According to the Database Directive (see section 2.4), which was incorporated in Greek law [Article 7, Law 2819/2000], database makers are protected, *inter alia*, from 'repeated and systematic extraction and/or reutilization of insubstantial parts of the contents of the database implying acts which conflict with a normal exploitation of that database or which unreasonably prejudice the legitimate interests of the maker of the database' [Article 7(5) Database Directive; Article 45A par.4, GCL]. In the German case *Stepstone v. Ofir* [Case No Az 208 0692/00 Landrericht Koln, 2001] the Court held that the defendant's usage of deep-linking to the plaintiff's website via repeated and systematic access violated the latter's database rights [Flint at al., 2006].

### 5.3 Recourse to exceptions/limitations?

In the aforementioned case of *Copiepresse v. Google Inc.*, the defendant claimed, *inter alia*, that its activities could be excused under the Belgian law's exceptions/limitations of 'quotation for the purpose of critique, argument, review or teaching' and 'reporting news' [Murray, 2013]. The first claim was overruled on the basis that Google's indexing activities are automatically executed by machines, so no human critique or review is performed which could be warranted under the relevant exception/limitation. The second claim of Google was also rejected by the Court which indicated Google's contradictory stated status, meaning on one hand that it is a news portal and on the other hand that it is a specialized search engine and not an 'informational portal' [Murray, 2013].

As reiterated, the Greek copyright legislation has adopted certain exceptions/limitations to the authors' rights. These are exhaustively prescribed in the law and no expansive interpretation can apply lest the copyright system could be undermined [Μαρίνος, 2000]. It is likely that a curator raising the 'quotation of extracts' defense [Article 19, Greek Copyright Law] to support its activities, likewise

Google in *Copiepresse v. Google Inc.*, would be rejected under Greek law as well. The rationale of this exception/limitation is the usage of others' quotations within a work in order to support or reinforce the comments or opinions in this work for scientific, research or ideological purposes [Μαρίνος, 2000; Greek Court Judgment ΜπρΑθ 15951/1989]. This entails that a new work should be available for the quotations to be embodied in and also that the specific purposes should be pursued. Apparently, it does not seem that a curator's activities, using others RRS feed, fit in with the previous explanation.

As for the 'reporting news' claim, it is likely that this could not be upheld by Greek law. According to article 25 par.1(a) GCL, the reproduction and communication to the public of works seen or heard in the course of the event are permitted without the consent of the author and without payment for the purpose of reporting current events by the mass media. The ultimate goal of this exception/limitation is for the mass media to keep the public easily and quickly informed on the current events [Μελάς, 1965]. However, a news aggregator or a curator engaging in somewhat similar functionalities could not fall under the notion of mass media within Greek law. These are usually computer-generated news websites that re-transmit the news that have been written/edited by others [News Google, 2014]. For a website to fall within the mass media term, firstly it has to aim at the massive diffusion of information and secondly its informational content needs to be shaped by its owner and journalists [Greek Court Judgment ΠΠρΘεσσαλ 22228/2011]. Thus, it is possible that curators' creations utilizing others RSS feed, like Google News, could not fulfill the definition of mass media and therefore they could not raise the 'reporting news' claim to defend themselves from copyright infringement.

Finally, the exception of reproduction for teaching purposes, meaning that it is permitted, without prior consent or remuneration, to copy articles lawfully published in a newspaper or periodical, short extracts of a work or parts of a short work or a lawfully published work of fine art work exclusively for teaching or examination purposes at an educational establishment. This provision refers strictly to educational purposes and it does not apply to scientific research purposes [Μαρίνος, 1995]. Thus, a curator-researcher does not seem to be able to have recourse to this exception for his activities.

## **6. Conclusions**

In this work we investigate the legal implications of information flow curation, i.e. aggregating, validating, enhancing and distributing digital content. Following the current trends as to the volume of information created and the ease of its dissemination, herein information is considered of variable flow rate and transient in nature. The legal implications of information flow curation are examined under the Greek legislation framework.

It is the opinion of the authors, and a common sense practice, that all curators of digital information, when in doubt about the legitimacy of any activity they engage in, with regard to copyrighted materials, should bear in mind that the fundamental principle points at the need to obtain the copyright owner's authorisation.

As a final remark, it should be noted that the examination of the application of the Greek copyright law legislations on information flow curation conducted herein is just

a passive approach that brings interested parties up-to-date on legal matters. Future work could include the extended examination of the matter on both U.S. and U.K. copyright law legislations as well as the introduction of amelioration proposals to all copyright law legislations examined, in order to promote the exchange of curated digital information, of course, without affecting the incentive of the rights' owner to create.

## 7. References

- Abbott, D. (2008), What is Digital Curation?, online at <http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/what-digital-curation> accessed on 01 April 2014.
- Aggarwal, C. C. (2007), Data streams: models and algorithms, 31, Springer.
- Aplin, T. & Davis, J. (2009) Intellectual Property Law : Text, Cases, And Materials, Oxford University Press.
- Atom, online at [http://en.wikipedia.org/wiki/Atom\\_\(standard\)](http://en.wikipedia.org/wiki/Atom_(standard)) accessed on 01 April 2014.
- Beagrie, N. (2006), Digital Curation for Science, Digital Libraries, and Individuals, The International Journal of Digital Curation, 1, 1, 3-16.
- Council Directive 91/250/ECC on the legal protection of computer programs.
- Despotidou A. (2011), The subject matter of protection - Works, in Kallinikou, D. & Stavridou S. et al. Boundaries and Interfaces, Report of Hellenic ALAI group, ALAI Dublin Congress.
- Dworkin, G. (1995), The moral right of the author: Moral rights and the common law countries, 19 Columbia-VLA Journal of Law and the Arts .
- Flint, M. F., Fitzpatrick, N. and Thorne, C. D. (2006), A User's Guide to Copyright, Tottel Publishing.
- Gervais D. (2002), Feist Goes Global: A Comparative Analysis Of The Notion Of Originality In Copyright Law, Journal of the Copyright Society of the U.S.A. 949, 968.
- Golab L., and Ozsu, M. T. (2003) Issues in data stream management. SIGMOD Rec. 32, 2, 5-14.
- Golab, L. and Ozsu, M. T. (2003), Data Stream Management Issues - A Survey, Technical report, University of Waterloo.
- Greek Copyright Law, online at <http://www.opi.gr/> accessed on 01 April 2014.
- Greek Law 3057/2002 (Article 81), FEK A' 239/10-10-2002 incorporating the EU Directive 2001/29 on the harmonization of certain aspects of copyright and related rights in the information society.
- Greek Law 3183/2003, FEK A' 227/26-9-2003 ratified the WIPO Performances and Phonograms Treaty see WIPO Performances and Phonograms Treaty, Geneva, 20 December 1996, online at <http://www.wipo.int/> accessed on 01 April 2014.
- Greek Law 3184/2003, FEK A' 228/26-9-2003 ratified the WIPO Copyright Treaty see WIPO Copyright Treaty, Geneva, 20 December 1996, online at <http://www.wipo.int/> accessed on 01 April 2014.
- Higgins, S. (2008), The DCC Curation Lifecycle Model. The International Journal of Digital Curation, 1, 3, 134-140.
- Intellectual Property Office, online at <http://www.ipo.gov.uk> accessed on 01 April 2014.



Kimberly Isbell & the Citizen Media Law Project, (2010), The Rise of the News Aggregator: Legal Implications and Best Practices, The Berkman Center for Internet & Society Research Publication Series, Research Publication No. 2010-10.

King, A. (2003) The Evolution of RSS, online at <http://www.webreference.com/authoring/languages/xml/rss/1/index.html> accessed on 01 April 2014.

Murray, A. (2010) Information Technology Law - The law and society, Oxford University Press.

Muthukrishnan, S. (2005), Data streams: Algorithms and applications. Now Publishers Inc.

News Google, online at <https://news.google.com> accessed on 01 April 2014.

Pedley, P. (2003), Essential Law for information professionals, London: Facet Publishing.

RSS Usage Statistics, online at <http://trends.builtwith.com/feeds/RSS> accessed on 01 April 2014.

RSS, online at <http://en.wikipedia.org/wiki/RSS> accessed on 01 April 2014.

Sikos, L. (2011) Web Standards: Mastering HTML5, CSS3, and XML, Apress Media, USA.

Stamatoudi, I. A. (1997), Moral rights of authors in England: the missing emphasis on the role of creators', 4 Intellectual Property Quarterly, 478-513.

Swack, C. (1998), Safeguarding artistic creation and the cultural heritage: A comparison of droit moral between France and United States', 22 Journal of Law and the Arts, 361-401.

The Open Graph protocol, online at <http://ogp.me/> accessed on 01 April 2014.

WIPO, (2012), Intellectual Property On The Internet: A Survey Of Issues, December 2002, online at <http://www.wipo.int/> accessed on 01 April 2014.

Wolff, A., and Mulholland, P. (2013), Curation, curation, curation. ACM Conference on Hypertext and Social Media.

Xalabarder, R. (2012), Google News and Copyright, in Aurelio Lopez-Tarruella ed., Google and the Law-Empirical Approaches to Legal Aspects of Knowledge-Economy Business Models, Springer.

Yakel, E. (2007) Digital curation, OCLC Systems & Services, 23:4, 335 - 340.

Yonah, J. B. (2013), Posting RSS feed items not copyright infringement, The Jerusalem Post, online at <http://www.jpost.com/National-News/Posting-RSS-feed-items-not-copyright-infringement> accessed on 01 April 2014.

Youtube statistics, online at <http://www.youtube.com/yt/press/statistics.html> accessed on 01 April 2014.

Youtube, online at <https://www.youtube.com> accessed on 01 April 2014.

Γιαννόπουλος, Γ. Ν. (2013) Η ευθύνη των παρόχων υπηρεσιών στο Internet, Νομική Βιβλιοθήκη.

Καλλινίκου, Δ. (2005), Η προστασία της πνευματικής ιδιοκτησίας στο διαδίκτυο.

Κοριατοπούλου-Αγγέλη Π. & Τσίγκου, Χ. Σ. (2008) Πνευματική Ιδιοκτησία - Ληματογραφημένη Ερμηνεία, Νομική Βιβλιοθήκη.

Κοτσίρης, Λ. Ε. (2005), Δίκαιο πνευματικής ιδιοκτησίας, Εκδόσεις Σάκκουλα.

Μαρίνος Μ.-Θ. Δ. (2000), Πνευματική Ιδιοκτησία, Εκδόσεις Αντ. Ν. Σάκκουλα.

Μαρίνος, Μ.-Θ. Δ. (1995), Η αναπαραγωγή έργων για εκπαιδευτικούς σκοπούς- Συμβολή στην ερμηνεία του άρθρου 21 Ν. 2121/1993 για την Πνευματική Ιδιοκτησία, ΕλΔ 36, 19-26.

Μελάς, Β. Θ. (1965), Οι περιορισμοί του δικαιώματος του πνευματικού δημιουργού, Αρμενόπουλος.