EJLT European Journal of **Law and Technology**

Determinants in the online distribution of digital content: an exploratory analysis

M. Borghi, M. Maggiolino, M. L. Montagnani and M. Nuccio [1]

1. Introduction

This article shows the phases – and discusses the results – of an empirical analysis addressing the legal business models that are adopted online to distribute digital content. Although it premised upon the idea that these models would have been greatly affected by the strong protection that digital copyright law offers to rightholders, it concluded the opposite way. The 'all rights reserved model' that is tenaciously incentivized by the current legislative framework is, indeed, all but dominant: it is overcome by 'open' and 'interactive' models that take full advantage from the architecture of the online environment and from flexible forms of copyright licensing. More in general, the results of the empirical analysis show that the current legal framework does not considerably affect the models of distribution that are implemented in practice; rather, they are significantly influenced by many other factors, such as the technology used for making digital content available.

Therefore, the article acknowledges how the current legal framework is inadequate and inconsistent with the business tools that entrepreneurs are employing and testing to better satisfy the demand. In order to show how the market diverges from the on line distribution model of digital content envisaged at legislative level, in the first section we give an account of the digital age turmoil and expectations, so to show how evidence about on line distribution of digital content is needed. This need is further confirmed by the literature review about distribution models for digital content that we after carried out in section 2. In sections 3 and 4 we describe the methodology followed and present the initial results obtained, which will be further analyzed in section 5 via a cluster analysis. In section 6 we comment and discuss those last so to illustrate, in section 7, our conclusions. We affirm that while the current legal framework, though biased towards a pre-determined solution for online distribution of digital content, has not inhibited the development of a kaleidoscopic market for online distribution, an enabling legal framework could foster it further and, by doing this, could favour the match between the demand and the supply of digital content.

2. From the 'celestial jukebox' to the evidence of a 'patchwork quilt'

Digital technologies and the networked environment have dramatically changed the way by which copyright works are reproduced, distributed and used. This is because works have become as easily reproducible and exchangeable as never before in the history of media. It is no surprise that online distribution of copyright content has attracted much litigation, legislative response and public attention as one of the major issue of the so called digital age (Litman J, 2002; Ginsburg JC, 2002; Patry W, 2009; Giblin R, 2012).

At the beginnings of the digital age, the future of the online distribution of creative content was envisioned through the metaphor of a 'celestial jukebox' (Goldstein P, 1994 and 2003). By this, it was meant that the internet was about to become a vast storehouse of entertainment and information orbiting like a satellite above users' heads and awaiting for subscriber's orders. Content would have been permanently available on users' demand and technology would have enabled rightholders to track every single use of their works and

charge accordingly. The 'best prescription' to make online distribution fitting the needs of both rightholders and users was – as Paul Goldstein (1994) classically expressed – 'to extend rights into every corner where consumers derive value from literary and artistic works'. [2] This prescription did not remain on paper; it was promptly translated into internationally binding legislation with the so-called 'Internet treaties' of 1996 (Ficsor M, 2002). [3] As a matter of fact, the WIPO Treaties and their national implementations have secured rightholders with measures that enable property rights to expand into 'every corner' of the digital online environment. These include: extension of the reproduction right to the use of works in digital form, [4] introduction of the new exclusive right of communication on-demand, [5] limited availability of exceptions and limitations to exclusive rights [6] and, most importantly, legal protection against circumvention of technological measures [7] and against removal of rights management information. [8]

With this legislative background in mind, it was expected that the private sector would have taken full advantage from the extension of property rights and, accordingly, it would have developed models of distribution based on strong legal and technological control over the content. In other words, online distribution should have been largely dominated by players offering content protected by technological measures and licensed under the most possible restrictive terms ('all rights reserved').Instead, both literature and anecdotic evidence show that this is not the case, as a plethora of diverse distribution systems have flourished ever since on the internet.

Yet, no systematic empirical studies have been conducted on the determinants of online distribution. Most importantly, no evidence has been collected on the business and legal models that are adopted to distribute digital content online. What are the models that shape the current landscape of online distribution? Are there models that are more prominent than others? How much, and in which respect, does the current online distribution deviate from the image of the 'celestial jukebox' that was strongly seconded by the legislators at the beginnings of the digital age? Premising on the idea that the legal framework should not push a specific business model, but should enable the flourishing of diverse models so to satisfy at most the demand, only a systematic empirical study can provide evidence of the gap, if any, between the current legal framework and the market structure.

After giving an account of the literature about the evolving relation between copyright law and distribution models, this article discusses the results of an empirical survey, where a stratified sample of websites distributing music, video and videogames in digital format has been analysed. To compose a reliable sample of websites, two criteria of selection have been employed, namely distributed content and popularity. The resulting dataset is composed of the following five content-based sections: music services, web radios, web televisions, video services and videogames. To select the most popular websites used by digital media consumers, internet search engines and databases (in particular *Wikipedia* and *Google Search*) and *Google Ad Planner* have been the main reference. Industry Reports (e.g. IFPI Digital Music Report) together with Industry Associations Members Lists [9] have been fundamental to check the industry endorsement and verify the legality of online media services. Although the number of web contacts is a good proxy of the popularity of a websites distributing digital contents, innovative practices could be found not only among big players but also in fast emerging platforms. For this reason, our sample does include websites which pop up in industry-specific search engines, although not considered in the top rank of *Google Ad* Planner.
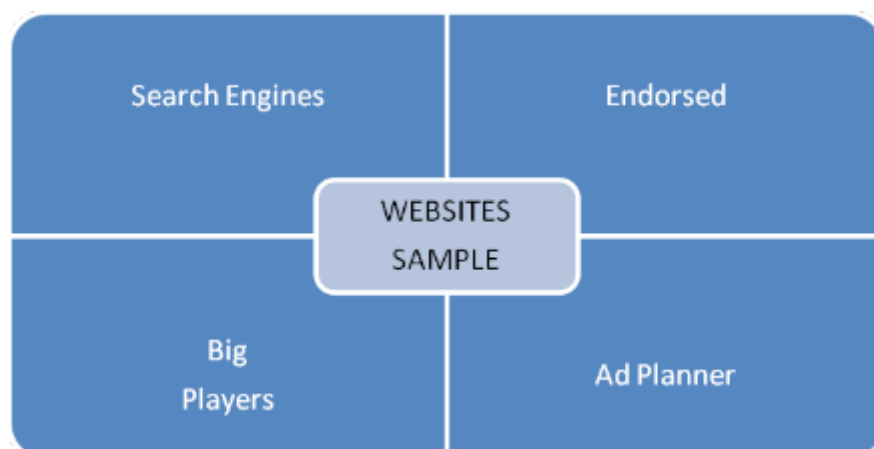

Figure 1: *A representation of the dataset composition*

For each website we have observed overall 76 variables, which provide information on a wide array of features, which include: the kind of content and services provided, the distribution methods, the technical restrictions imposed on content, the right management and licensing systems, the revenue models, the management of personal data and the level of interaction with users through social network facilities. The first aim of the analysis was to extract the specific economic and legal features which define the recurrent business models in the online distribution.

To identify the models of digital content distribution, we have applied to the resulting dataset an unsupervised neural network algorithm, namely Self-Organizing Maps (SOM), a technique of clustering and pattern recognition. The SOM, also called the Kohonen network or the topological map, is an unsupervised neural network algorithm developed by the Finish physicist Teuvo Kohonen (2001). It provides two useful operations in exploratory data analysis: reducing the amount of data into representative categories (clustering), and aiding in the exploration of proximity relations in the data thought a non-linear projection of the variables (feature mapping). The implementation of this technique is diffused in different fields of research, as image and sound detection, medicine and psychology, bibliometrics and in all sciences in which the pattern recognition is relevant to the investigation. In the socio-economical sciences the algorithm has been applied in research on financial indices and, more recently, on regional economy and local development (Carlei V, Nuccio M, Sacco PL, Buscema M,2008). The learning mechanism of the SOMs follows an intuitive logic: firstly, mapping the input data as accurately as possible, secondly, finding out its main hidden structural characteristics, and eventually being able to classify them appropriately. For this reason, SOMs are an effective tool for cluster analysis, even in cases in which more established statistical methods may not be sufficiently powerful.

Given the high number of variables monitored for each observation, the SOM has been able both to gather the websites which might be considered similar in respect to multiple economic and legal dimensions and to select the features which are particularly relevant to shape business archetypes competing in the distribution of digital content.

3. Copyright law and online distribution: an evolving relation

Literature on online distribution has approached the interaction between law and models of distribution from different angles. A first line of studies focuses on the relation between the law and the business model for distribution that the law incentivises (Berkman center, 2003, 2004, 2005). These studies were undertaken in the wake of the adoption of the WIPO Treaties in 1996 and their implementation. This took place in the US through the *Digital Millennium Copyright Act* in 1998 (DMCA) [10] and in the EU through the *Directive 2001/29/EC on the harmonization of certain aspects of copyright and related rights in the information society* ('Directive'). [11]

Back in the second half of the nineties, this line of studies illustrated how the model for distribution envisaged by these provisions (the so called 'digital copyright') revolved around the adoption of digital rights management systems (DRMs) that enabled and enhanced the control over the uses of the content distributed online. These systems, and the distribution model that they enabled, were foreseen as the solution to the lack of control over the uses of copyright works that right holders complained as a consequence of the digital technology and the internet. Moreover, the solution therein envisaged also aimed at extending to the online environment the distribution models already implemented off-line. In order to achieve such twofold purpose, the digital copyright provisions introduced a three-layered system of protection. Under such WIPO provisions and their implementations, digital works were – and still are – firstly protected under copyright law; secondly protected by technological measures of protection that are a key element of distribution via DRMs; and finally protected through the provisions that sanction the circumvention of technological measures of protection. As a matter of fact, the WIPO Treaties, though indirectly, do legitimize the recourse to systems of DRMs (Samuelson P, 1999), as they require the signatory states to introduce provisions that legalize the adoption of electronic information on the rights regime and TPMs (WCT, arts. 11 & 12; WPPT, arts. 18 & 19), both elements of the more complex DRM systems. Specifically, electronic information, i.e. information supplied by rights holders that identify the work or protected material, the author or any other owner of rights, or any information about the terms and conditions of use of the work or of other material, and to whatever number or code might represent such information, shall not be removed or altered, and materials from which information has been removed or altered shall not be disseminated.

TPMs, on the other hand, shall not be circumvented. Their protection extends to the prohibition of producing and commercializing devices that allow and favour such circumvention. Already in this first line of studies (iVir 2006, 2007, 2008, 2009), however, there is the awareness that online distribution does not consist in the mere transposition of off-line dynamics, but may be the result of the adoption of sui generis models of distribution. While a complete description of these models emerge solely in the second wave of studies on the topic, some specificities of the online distribution are already depicted: they regard the licensing regime applied to the content and the revenue systems used by the distributors.

As far as the licensing regime is concerned, the phenomenon of online distribution is characterized by the adoption, along with the traditional copyright 'all right reserved' licences, of 'open licenses' (Guibault L, Westkamp G, Rieber-Mohn T, Hugenholtz Bet al., 2007, 149-153), such as the creative commons (www.creativecommons.org). This stream of studies acknowledges that there are different dynamics within the online distribution market which are capable of amplifying the inclusive capacity of copyright. This is because the open licensing schemes facilitate the dissemination of works and the creation of derivative works. The spread of open licences is the signal of the need to adopt rules tailored to the Internet, namely rules that can facilitate a wider dissemination of works without infringing copyright and without depriving right holders from control over all possible uses of their works.

As far as the revenue system is concerned, it is suggested that revenues could be collected not only through licensing fees but also via advertisement (Slater D, Gasser U, Smith M, Derek EB, Palfrey JG, 2005, 11-16). As a matter of fact, ad-based distribution, although initially neglected, has soon established as a key driver, especially in the context of free distribution of content in 'user-generated content' platforms and social networks, which represents the further edge of digital distribution.

3.1 The digital media shops and the first phase of digital distribution

The legal framework that was adopted as a response to the challenges of the Internet envisaged a distribution model that corresponds to the one adopted off-line, i.e. it aimed at establishing digital media shops (hereafter DMSs) which would recreate the dynamics of offline distributors in the virtual environment. DRMs were the technological tools that, in accordance with the legislative framework, would enable this model by offering right holders a control over the use of the products on offer that was far superior to that available for tangible goods. The analysis of the typical DRM-supported distribution as depicted by the legal literature demonstrates their initial implementation through subscription and à-la-carte downloads, which were subsequently augmented with the introduction of music 'to rent' (Holden B et al, 2003, 9). All these services involved, and in most cases still involve, a proprietary management of copyright, i.e. an 'all rights reserved' regime which, when applied to technologically protected digital goods, can be even more strongly proprietary. By mostly belonging to the realm of proprietary distribution, these models embody the distribution system foreseen at legislative level.

Downloads on payment

The success of DMSs arose from the spread of à-la-carte download services. As is widely known, the success of iTunes is linked to the offer of pay-per-download music files, in comparison to previous DMSs (e.g., Pressplay and MusicNet) which made music files available only by subscription. Moreover, iTunes was the first to offer a wide selection of single musical works, including major popular hits, although the iTunes business revolved around, at least initially, the sale of iPod players (Gasser U, 2005c, 45). However, over the years iTunes has modified its offer by enlarging the uses that users can do of the files 'acquired' – for example it allows an unlimited number of copies on iPods, and an increasing number of copies on computers and within the compilations created by the purchaser of the file – and by offering the files without DRMs. Moreover, iTunes has been among the first distributors to offer content in streaming and to morph from distributor to integrated platform.

Subscription

Examples of subscription systems were Napster 2.0 and Rhapsody, which permitted access to the entire catalogue upon payment of a monthly sum. However, the use that could be made of the downloaded files was extremely limited, at least until further fees were paid to be able to transfer the file to a portable player, copy it onto a CD, or to save a copy on the hard disk where it was initially downloaded, after the subscription

expires. In the field of cinematographic works and video, Starz was prompt to offer the possibility to download recent films directly onto computers, with users paying a monthly subscription. The user was 'granted a personal, non-exclusive, non-assignable and non-transferable limited license to use the site and access content, that allows one, nevertheless, to download the film onto three different PCs'. [12] Starz also uses the proprietary DRM technology of Real Network, a company which occupies a leading role in the online distribution of films, whereas in the realm of music were the DRM systems of Microsoft and Apple to predominate.

Rented music

The first phase of online distributions, which is the one that the most mirrors what envisaged by the legal framework and depicted in the sector studies, was enriched by a further way of distributing files made possible by the rented music service. It was Napster that launched one of the first rented music services, 'Napster To Go', based on a subscription (i.e., payment of a monthly sum against which the user can download as many pieces of music as desired), but defined it as 'music for rent' as it provides users with the possibility to use the files downloaded and transfer them to a compatible player as long as the subscription is valid. [13] What made possible this service was the implementation of a specific DRM named 'Janus' – at that time a cutting-edge software – which checked the validity of the subscription in relation to the files downloaded and, when the subscription expired, cancelled the files.

3.2 The utopia of superdistribution

The initial resistance of legislators to exploring new ways of managing copyright online and to rely on proprietary copyright regimes was quickly overcome by market dynamics. Models for online distribution have developed in a direction different from that originally foreseen at the legislative level, resulting in a mix of distribution models for copyright works in digital format that only partially coincides with what envisaged by the implemented regulatory system.

These market dynamics have been acknowledged by the legal literature that developed in the field. 'Super-distribution' (Rosenblatt B, 2004) and 'peer-to-peer shops' constitute one of the first alternatives to DMSs envisaged at legislative level. They have been identified as tools for enriching the online distribution models as well as for exploiting peer-to-peer networks for the legitimate exchange of authorised content. This would enable the same networks that allow the illegal file-sharing to distribute encrypted content in which sharing can take place under conditions stipulated by legitimate copyright owners. The incentive provided for users of P2P networks to exchange authorised content differs depending on the 'P2P store' chosen, but it is generally a form of recompense (either monetary or often in the form of a point-collection system) whereby credit is earned whenever authorised, rather than infringing, files are shared. P2P stores are the typical expression of this form of distribution and make use of the advantages and the capillary nature of the P2P networks, but also limit digital piracy by the adoption of DRM systems.

One of the first examples of 'super distribution' was initially developed to permit the embedding of encrypted files in the KaZaA and Grokster networks. These files were identified by an Altnet icon which certified the legal circulation of content, allowing use of these files on payment of a sum determined by the content holder (Fisher W, Gasser U, Slater D, Smith M, Palfrey J, 2005). In substance, Altnet incentivised the exchange of legal files on P2P networks (no longer in Grokster) at the price and according to the license methods decided by the content holder. Users who shared these files accumulated points that could then be used to claim prizes. Another example of 'super distribution' was Weed [14] which, at least initially, had a diffusion that was even more widespread than Altnet. Indeed, while Altnet circulated only on the KaZaA and Grokster networks, files from Weed could circulate freely in almost any file-sharing network. In this system, encrypted files could only be decrypted following payment, and could only be used under the conditions imposed by the content owner. On the wake of the superdistribution systems, also some DMSs started to offer the sharing of acquired files with other subscribers of the same service alongside their traditional services ('Napster Share'), just as sites dedicated to 'super distribution' added services for downloads on payment of subscription (http://www.wippit.com/default.aspx).

3.3 The Boost of advertisement-based distribution

Whilst rightholders have progressively abandoned the use of DRM and 'super distribution' has not delivered the expected market results, distribution systems based on income generated from advertisement rather than economic exchange have gained ground so to provide a real alternative to DMSs. Although ad-based models were initially overlooked as it was believed that the return generated by the sale of advertising could never equal what generated by the sale of music files using DRM systems to control, at least from a theoretical point of view, their access and use. However, market figures overcame this believe and these models have increasingly been adopted (Peterson L, 2007; http://www.grabstats.com/statmain.asp?StatID=1454).

Besides an initial attempt to legalize peer-to-peer networks and make them distributors of music on a free basis supported via advertisement (Fisher K, 2008), the real success of ad-based distribution is to trace back to commercial distribution. Among the first websites to adopt an ad-based revenue model was Mjuice which sold banner and virtual room for commercials on its webpages. [15] Internet Underground Music Archive tried to include commercials in the files offered for streaming. [16] Neither of these businesses, however, matched the apparent success of We7 (http://www.we7.com). The service was launched in May 2007, with initially only 30 tracks available to users which the year after had expanded to over 3 million downloads and over 100,000 subscribers and, in the upstream market, had managed to license music from some of the world's leading artists, record labels and distributors.

The website feature that made it so popular was that it was the first to permit the choice between streaming and downloading of music for free with advertisements or the purchase of music files without advertisements. This distribution model was – and still is – is enabled by the MediaGraft technology which places advertisements at the beginning of each track streamed or downloaded for free. The advertisements target users by a system of off-web advertising. This new media broadcasting method targets customers and grafts the brand audio message onto music tracks that people choose, and is presented each time the tracks are played. The We7 service allows users to remove the advertisements from downloads after four weeks. Moreover, We7 philosophy of sharing the revenue generated from these advertisements with the artists, the labels, and the other rights owners also contributes to its popularity. It is worthy noticing that the distribution system implemented by We7 is DRM-free and relies on the traditional copyright licensing regime 'all rights reserved'. In the terms and conditions for the service, users acknowledge and agree that they have no right to provide any files obtained through the We7 service to any other party or through any other means, save as otherwise provided through the We7 sharing service, and that they may only make copies of any file obtained through the service for their own personal use.

Ad-based distribution is however something more than a further model for digital distribution. It represents the point of contact between traditional distribution and the phenomenon of user-generated content (UGC) which are hosted on websites – or better platforms – devoted to the creation of a space where digital content can be posted and shared freely since the incomes derive from the advertisement to which users are exposed (Boyd D, Ellison NB, 2007). From a different perspective, UGC platforms can be considered paradigmatic of the limits encountered by the distribution model envisaged within the legal framework. Furthermore, they not only constitute a further method of distribution, but also contribute to enlarge the amount of content available online.

Facebook (http://www.facebook.com), MySpace (http://www.myspace.com), Friendster (http://www.friendster.com), and Xanga (http://www.xanga.com) are among the first social network platforms that rely on UGC to attract and obtain revenue primarily from the sale of online display advertising. It is worth noting that UGC platform incomes seem to be generated also through the accumulation of user information, which use for targeted marketing purposes is highly controversial (Trusov M, Bucklin RE, Pauwels K, 2008). Both UGC and social networks present elements that are highly debated and require analysis in terms of both the business models implemented and the content posted and disseminated.

As far as the entrepreneurial activity and the business models are concerned, it is worth noting that social network members are both 'content providers' and 'customers' of the website as their exposure to advertising while using the platform generates revenue. In 2007 MySpace was the fifth most popular web domain (behind Google, Yahoo, MSN, and AOL) in total number of individual pages and now serves 8% of all advertisements on the Internet (Adegoke Y, 2006; Smith D, 2006).At the moment it is Facebook to present the business model most relying on advertising. Because of its deep penetration within a series of micro communities (e.g., college campuses), this platform enables local advertisers to use Facebook to target

specific colleges or other audiences (Gabbay N, 2006). Facebook 'social ads', for example, are not only a very specific tool to target customers, but can also leverage the power of 'Facebook News Feed' by updating all the contacts of a user that shows interest in a specific product.

However, the economic sustainability of the UGC platforms is not only derived from the sale of advertising. This is combined with income from other sources such as subscriptions, donations licensing and e-commerce (Helberger N., Guibault L., Janssen E. H., Van Eijk N., Angelopoulos C. J., Van Hoboken J., Swart E. et al., 2008, 115). The use of a combination of revenue generation mechanisms reflects the fact that online advertising has not evolved at the same rate as the platforms themselves, and still tend to replicate the same methods used offline in the digital environment (e.g. banners and sponsored links – the traditional forms of online publicity). However, developments in online advertising such as profiles within social networks, virtual communities, the sponsorship of widgets and web 2.0 applications etc. The lack of development of systems of online advertising specifically adapted to the Internet appears to explain the reason why some services offered by UCC platforms have become 'on payment' systems (adopting such devices as annual subscriptions, a-la-carte or pay-as-you-go methods of payment).

All Internet users are aware of the amount of content created and posted by amateurs online every day. A recent survey of 2000 Internet users aged 13-75 confirmed that more than half of them create content by editing their own photos, videos or music, and nearly half create music, videos, photos, blogs and websites for other users (Deloitte, 2007). Since distributing content online has become increasingly easy, one-third of those interviewed considered themselves as 'broadcasters'. Although not posting content, almost 70% of those interviewed said that they 'access' content created by other Internet users to the extent that computers are more of an entertainment device than television for many people. However, amateur content is not the only one to be available on social networks which have recently started to be exploited as new communication channels. This is the case, for example, of traditional broadcasters that open their own channel on YouTube.

4. Models of online distribution: empirical evidence

From the literature review it emerges that, although the legal framework encourages the adoption of an online version of the brick-and-mortar shops, markets move in another direction. To have a in-depth understanding of this different trend, we have developed an empirical study whose methodology (§§ 4.1 and 4.2) and initial results (§ 4.3) are described in what follows.

4.1 Characteristics of the sample and method of data collection

A stratified sampling technique has been applied to have a representation of the current models of digital media distribution available on the web. As mentioned in the introduction, relevant part of the sample is represented by the so called 'big players': these include websites considered to be the major players on the market by the industry reports and by Google Ad Planner. In particular, Google Ad Planner enabled the identification of services that are classified as very popular (among the top 150) by using the 'Unique visitors' [17] index:

For music services, web radio, web tv and videogames, a website has been considered to be a big player if its Unique Visitors are more than 1 Million.

> For the video services, a website has been considered to be a big player if its Unique Visitors are more than 5 Million.

A random selection of the industry endorsed digital services has been added to the sample. The industry lists to which the researcher has been referring are different for every section of the dataset, in particular:

For the music services and the web radios: International Federation of Phonographic Industry – Digital Music Report 2010/2011 and Pro-music, the coalition of music record companies and retailer.

> For the video services and the web televisions: Motion Picture Association of America;

For the videogames: Entertainment Software Rating Board, Academy of Interactive Arts and Sciences, Entertainment Software Association, International Game Developers Association, Associazione Editori Software Videoludico Italiana.

Some follower players and innovative services have been included into the dataset because of their service typology, even though they do not belong to the top 150 most popular online services. For the music services, websites have been selected from the list of case studies listed on Creative Commons website (http://wiki.creativecommons.org/Musician). Websites selling *only* physical products (CDs, DVDs), or which distribute content without the authorization of right owners, have been excluded as not being part of the project. Moreover, websites providing only links to other websites, but which do not make content directly available to users, have been excluded.

Concerning the music services, only websites providing digital music on demand have been taken into consideration; in particular, online music stores; free music distribution legal websites, creative commons or open source online music websites (commercial and non commercial) and digital music hosting services (e.g. content uploaded). [18] Among the web radios, all the websites included in the sample are radios available online, providing the so called 'webcasting' or 'live streaming' service. [19]

With respect to the video services, *only* websites providing digital video content 'on demand' have been considered. This include, in particular: online video/movie stores, free distributors of legally acquired videos, creative commons or open source online video websites (commercial and non commercial) and video hosting services. [20]. Web-tv includes websites providing the so called 'webcasting' or 'live streaming' video services. [21]Finally, the videogame section gathers videogames stores and websites distributing videogames, either to download or to play on line. This includes free videogame services and open source projects. [22]

On the whole, as shown in Table 1, the sample includes about 30per cent of Video Services, 23 per cent of web radios and 21 per cent of music services. Nine web radios also have music services. There are only seven websites belonging to both Music and Video Service (Shop2Download, Umusic, ARTIST direct, Zune (Microsoft), Bearshare, Los 40 Principales and CoolRoom), while five websites include music, video and videogames (Free record Shop, DigitalHmv, Virgin Media, Amazon and Apple – iTunes).

	Count	%
Music on Demand	125	20,90%
Video Service	180	30,20%
Videogames	97	16,20%
Web Radio	139	23,30%
Web TV	94	15,70%

Table 1: *Distribution of industries (597 valid)*

It is worth mentioning that the number of websites that result from the sum of the different categories does not account to 597, but to 635. This is due to the fact that, given the hybridization of the services offered, some websites belong to more than one category. The same occurs with the variables related to each of the websites analysed: they are observed following a binary method and they are not mutually excluding.

As shown in Figure 2, the sample consists of 30 per cent of major players, 48 per cent of followers, 13 per cent of innovators and 8 per cent of indies.
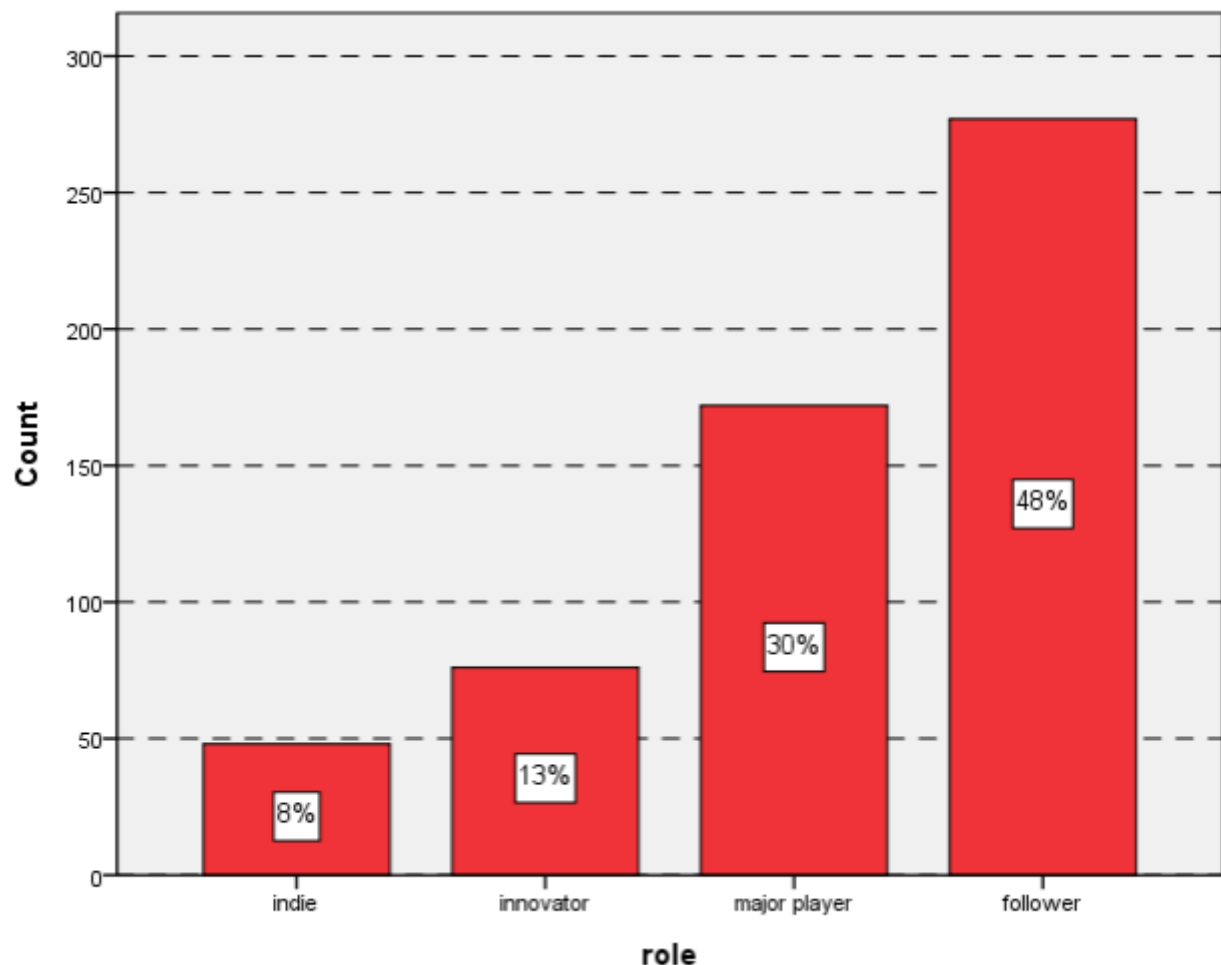
Figure 2: *Dataset composition as to market roles*

4.2 The variables observed on the websites

Each website has been analyzed by monitoring 76 variables. These include both nominal and binary variables grouped into eight categories:

1. Content and services
2. Distribution methods
3. Technical restrictions
4. Rights management
5. Uploading regime
6. Revenue model (which includes payment methods)
7. Privacy regime
8. Social networking

Content and Services variables cover the digital products and services offered on the website. These include music, video, videogames, ebooks, playlists, mobile applications, ringtones, pictures and news. Some variables included in this set might seem to be not strictly related to the distribution of digital media content, but in fact they add value to the core offer of major market players. The variable 'news', in particular, has been considered to be present on websites only if they presented a section dedicated to generalist news; consequently those websites providing only music news have been considered as not presenting this feature. Similarly, pictures have been considered to be available on a website only if this had a dedicated section specific for photos and pictures and not, for instance, if photos are simply present on the web pages.

Distribution Methods are the ways in which content is made available to users. In the distribution methods are included: download, streaming, live streaming, hosting, podcast, mobile, embedding, syndication. Download means the possibility to get a permanent copy of the file into the user's hard disk or other devices. Streaming happens when the website allows users to access the entire content on demand without need of downloading it. In 'live streaming' the content is streamed in real time and not on demand, meaning that users can only access the content when it is transmitted live by the website. Playlists created by users with the possibility to be accessed at a later stage are sometimes called 'radios' on some websites and they were considered as part of the streaming service. 'Hosting' refers to websites which allow users to upload content to be made available to other users. Podcast is the service that allows users to download pieces of previously live streamed content. In the mobile service are included all websites which provide facilities to use the service also from a mobile phone. Embedding is the possibility to share a webpage link to another webpage, and it is usually identified by the presence of a string of html code which can be copied and pasted on another page. Often, embedding is possible just through some social networks icons on the webpage or some banners with the clear sign 'share'. Syndication is the possibility to get constant updates on that webpage, usually identified by the symbol of 'RSS Feed' or just 'Feed'. Among the distribution methods, it is also included the presence of links to other music retailers such as Amazon and iTunes.

Technical Restrictions include all the limitations to the services provided, such as geographical limits for service and technical requirements to use the service. These include hardware and software restrictions, such as specific platform, operating systems or media player required. Moreover, this category includes restrictions on downloading, format shifting and interoperability. This set of variables is very important, since it permits to distinguish between 'DRMed' and 'DRM-free' content.

Rights Management include variables on the management of intellectual property rights, such as the provisions 'All rights reserved on website' and 'All rights reserved on content', or, alternatively, the Creative Common License signs, the GPL (General Public License) indication or other provisions. This classification allows to understand under which terms of protection is the content distributed by rights holders; for instance: if modifications are allowed, if the purposes of use will be commercial or not, for personal use or not. This category regards specifically the license regime, which provides information about download of or access to content.

The Uploading Regime is referred to how content submitted by users is licensed and protected by the website. In particular, the variables introduced are about the ownership of the content submitted by users, the nature of the license that the user is requested to grant to the website (non exclusive, irrevocable, etc.), the type of remuneration (if any) for the uploader and the regime on derivative works and copyright infringement. This category is of particular importance in the light of the massive development of 'user generated content' (UGC) platforms. Revenue Model includes information on the sources of income for the web services analyzed. The open source model includes free with AD or free without AD business models. The availability of content for free (with or without AD) regards downloads, streaming and podcasts. Some websites use also a form of revenue coming from donations made by users to artists. In the revenue model set are included also variables about payment methods, such as credit card, PayPal, mobile phone billing payment and other prepaid forms.

The Privacy Regime includes information about the privacy policy adopted by the website, including the use made of personal information collected and if users have to express their consent for such uses. Finally with the Social Network it is possible to verify whether the websites provides a link to the main social networks or if it just signalizes its presence on them. It was not part of the analysis to verify if a website is actually on such social networks, but the intent was only to evaluate whether they signal it or not.

The following table summarizes the major features of the collected sample of websites.

	Count	%
Books	28	4,7%
Mobile apps	84	14,1%
Music	302	50,6%
News	91	15,2%
Pictures	137	22,9%
Playlist	138	23,1%
Ringtones	40	6,7%
Video	385	64,5%
Videogame	158	26,5%

Table 2: *Content distributed*

4.3 Distributing content in web 2.0

The analysis of the features monitored in sample provides a picture of the status of the online distribution. Digital content can be made available to users either via downloading or streaming, or on both modalities. In turn, streaming can be 'on-demand' or 'live'. The other significant feature in the distribution methods is the possibility of posting content on websites so that other users can access them, mostly via on-demand streaming. This feature is the 'hosting' facility, which has been made popular by UGC platforms such YouTube [23] and many others.

Data in our sample (Table 3) show that, as far as sites combine different modes of distribution, downloading remains relatively more prevalent in the music sector (95 per cent), while streaming dominates in the video sector (92 per cent). Live streaming characterizes web-radio and web-television, but it is used also in other sectors. Although there is wide diffusion of streaming across the sectors, it is only in the video sector that, coupled with the hosting facility, streaming becomes a specific feature of the business model.

	Count	%
Downloading	313	52,4%
Streaming	464	77,7%
Embedding	230	38,5%
Hosting	197	33,0%
Link retailer	91	15,2%
Live streaming	211	35,3%
Mobile	171	28,6%
Podcast	100	16,8%
Syndication	221	37,0%

	Total	Downloading		Streaming		Embedding		Hosting		Live Streaming		Mobile	
		Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %	Count	Row N %
Music Service	125	119	95,2%	74	59,2%	37	29,80%	10	8,10%	12	9,7%	50	40%
Video Service	180	80	44,40%	166	92,20%	110	61,10%	110	61,10%	31	17,2%	49	27,2%
Videogames	97	85	87,6%	60	61,9%	5	5,20%	33	34%	0	0%	5	5,2%
Web Radio	139	25	18%	128	92,1%	53	38,1%	16	11,5%	131	94,2%	58	42%
Web Tv	94	19	20,2%	90	95,7%	50	53,2%	22	23,4%	68	72,3%	35	37,2%

Table 3: *Distribution methods (valid 597)*

Sites which offer hosting service, i.e. the possibility for users to upload content to be subsequently made available to other users, represents one of the new developments of online distribution and is consistent with the way in which the web 2.0 blurs the boundaries between commercial and amateur distribution. They represent about the 29 per cent of the sample. In this respect, it is worth observing (Table 4) that uploading has become a common feature also in websites which do not characterize as UGC platforms and are, rather, online content distributors based on streaming and downloading.

Additional elements exist with services that enable uploading which reflect and the relationship between the uploader and the site, as well as the licensing regime for uploaded content. In the majority of cases, sites tend not to acquire ownership of the content which is posted and it is, therefore, the uploader who decides the conditions under which other users can use it (unless the platform reserves the right to decide the licensing conditions for hosted content).

	Count	*%*
Downloading	86	49,70%
Streaming	150	86,70%
Embedding	110	63,60%
Hosting	146	84,40%
Link retailer	34	19,70%
Live streaming	45	26,20%
Mobile	52	30,20%
Podcast	34	19,80%
Syndication	89	51,70%

Table 4: *Distribution methods of websites allowing upload (valid 173)*

Copyright regimes applied to the distributed content are mainly divided into 'all rights reserved' and 'some rights reserved' regimes. While the first indicates that content is subject to the full extent of protection offered by the applicable copyright law, the latter designates so called 'open' licensing regime such as the Creative Commons licenses (from which the term 'some rights reserved' is borrowed) or the General Public Licence (GPL) (http://www.gnu.org/licenses/gpl.html). Also, it must be recognized that sites often present content subject to diverse copyright regimes, and that locating this type of information may be difficult.

Although websites can adopt different regimes for different content, a predominance of 'all rights reserved' regimes emerges in the distribution of music and video, whilst a coexistence of the two main regimes is observed in the videogames sector.

	Count	%
All rights reserved on content	477	79,9%
CC or GPL on content	64	10,7%
All rights reserved on website	517	86,6%
CC or GPL on website	42	7,0%

Table 5: *Licensing regime (valid 597)*

Within the wide category of technological restrictions that are adopted by online distributors, restrictions have been classified into three general categories, namely hardware, software and DRM-related restrictions (Table 6). Software restrictions concern more than 50 per cent of the sample.

	Count	%
DRMed content	102	17,1%
Geographical restrictions	139	23,4%
Platform or hardware	147	24,7%
Software	315	52,8%
Restrictions on downloading	72	23,0%

Table 6: *Technical restriction (valid 597)*

Information regarding technological restrictions is particularly interesting if correlated with the pre-chosen copyright regime and with uploading opportunities (Table 7). Sites which apply technological restrictions tend to adopt a regime of protection for their proprietary content (all rights reserved). Nevertheless, many website are increasingly adopting a double regime which allows the coexistence of technologically protected content and DRM-free content, catering for different targets of consumers

	All Rights Reserved On Content (valid: 477)		All Rights Reserved On Website (valid: 517)		CC or GPL On Content (valid: 64)		CC or GPL on Website (valid: 42)	
	Count	%	Count	%	Count	%	Count	N %
Uploading allowed	130	27,3%	153	29,6%	37	57,8%	18	42,9%
DRMed Content	94	19,7%	97	18,8%	3	4,7%	0	0,0%

Table 7: *DRM and upload vs. licensing regimes (per cent row/column)*

An important element in the identification of business models for the online distribution is the system by which websites generate revenues. Traditional online advertising (e.g., banners and sponsored links) still plays a central role in the system of free use of content and represents about the 48% of the sample (Table 8). However, it can be observed an increasing use of systems which include promotional messages inside the files which are distributed. This appears to be frequently used for videos (e.g., 'prerolls', 'postrolls' and 'overlays').

In general, websites seem to adopt different strategies to revenue their activities on the web. Among the revenue systems, the subscription model of payment seems to effectively suit to streaming as the emerging method of online distribution for digital content

	Count	Column N %
Donation	49	8,5%
Pay Per Download	179	30,2%
Subscription	84	14,1%
Free No Advertisement	157	26,3%
Free With Advertisement	291	48,7%

Table 8: *Distribution methods (valid 597)*

An important dimension of the analysis of the website privacy policy is the ease with which users are informed of the conditions under which their personal data and other information gathered by the site are used. Distributors of digital content are aware of the economic value of personal data and many of their business models rely on the exchange of this data. Nevertheless, some companies are still reluctant to apply laws on the privacy and websites which do not provide any kind of information about the processing of personal data amount to almost 1/3 of the sample (Table 9). When websites include a privacy policy document, the information recovered tends to be homogeneous across sectors. In most cases users are required to register to access (at least some) content or services, and this requires the compulsory disclosure of personal data which varies from nickname, email address and password to gender, age,

profession and postal code. Information on hobbies, attitudes and interests are usually non-compulsory fields to be completed.

	Count	%%
Registration Required	376	63,0%
Profiling Marketing	350	58,6%
No Privacy Policy	178	29,8%

Table 9: *Privacy policies (valid 597)*

5. Cluster analysis

The descriptive presented above shows that online distribution of digital goods is a kaleidoscopic phenomenon that could be interpreted via a cluster analysis. This aims at making stand out divergences and similarities among the different business solutions adopted to distribute digital content. Technically, through the cluster analysis we pursue the task of assigning a set of objects into groups (called clusters) so that the objects in the same cluster are more similar (according to some variables) to each other than to those in the other clusters. Such a task can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them.

In our analysis, we use an unsupervised neural network, the Self-Organizing Maps (SOM) algorithm, which provides two useful operations in exploratory data analysis: first, it allows a simultaneous classification of the data; second, it permits to depict within a two-dimension grid the multi-features that characterize our sample. The SOM is a multidimensional scaling method, which projects data from input space to a lower dimensional output space. Some resemblance can be found between the projection properties of SOM algorithm and of the principal component analysis (PCA). The main difference is the nonlinear projection of the SOM that can be a great advantage in many cases (Bodt, E, Verleysen M, Cottrell M, 1997), like revealing local or fine details. In contrast to PCA, in SOM local factors impact locally and, there, they will have greater importance in the projection (Kohonen, 2001).

In SOM networks, the key element is a layer, called Kohonen layer (KL), which is made up of spatially ordered Processing Elements (PEs). The global state of the layer evolves during the learning process, specializing each PE unit as an indicator of useful statistical features of the input data. This representation of relevant statistical features through an idiosyncratic spatial organization of PEs is known as Feature Mapping. SOMs elaborate their feature mappings through an unsupervised learning technique. A key feature of SOM mapping is, indeed, the logical splitting of the input space into clusters. For each given PE in the KL, there is a set of input data which, if submitted to the SOM, make the given PE the winning one. Such input points define a region. As the learning algorithm is based on the distance between the input and weights vectors, such regions in the input space are made up of contiguous points, namely, clusters. In other words, close-by input data tend to map onto the same PE, or more generally within the same neighbourhood of PEs.

5.1 The algorithm: SOM and K-means clusters

The SOM algorithm has some resemblance with vector quantization algorithms, like k-means, and is closely related to non-linear Principal Component Analysis. The great distinction from other vector quantization techniques is that the units are organized on a regular grid and -along with the selected unit- also its neighbours are updated. The topological preservation is accomplished through an ordered mapping in such a way that nearby patterns in the input space are mapped also to nearby units in output space.

The SOM consists of an array or lattice of elements called neurons or units, usually arranged in a low dimensionality grid (2-D) and typically shaped in a hexagonal form. Associated with each unit, there is a pattern called*codebook*, having the same dimensionality of the input patterns: in the actual analysis 76 binary variables have been processed per each website in the sample. Plotting this data on a 2-D grid and using a colour scheme, a landscape with walls (mountains in warm colours) and valleys (plains in cold colour) is visualized: the walls separate different classes, and patterns in the same valleys are similar. The unified distance matrix (U-Matrix) is a graphical representation of the SOM array structure: in Figure 3 it is composed by 30 (6x5) cells and each cell gathers a certain number of similar records (websites).
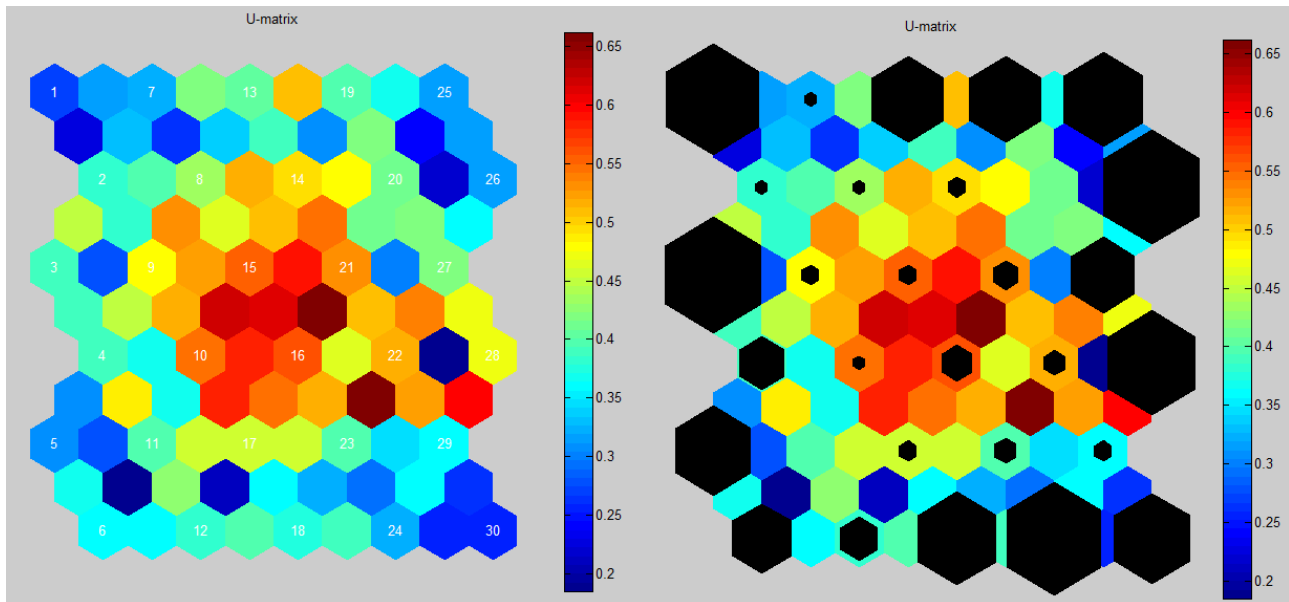

Figure 3: *Unified distance matrix (U-matrix a) and density of clusters (U-matrix b)*

The SOM portrays a faded landscape where the boundaries between clusters could be theoretically drawn where the differences among clusters are higher, i.e., in the U-matrix above, in proximity of the red and yellow hexagons.

For this reason, further processing of the SOM is needed. Several methods can be applied: in this study k-means clustering is applied to the SOM, consistently with the procedure included in SOM Toolbox for Matlab (Kuo RJ, Ho LM, Hu C.M, 2002. The *k-means clustering* has enabled us to obtain four main clusters, which can be easily used to describe the sample, and whose dimensions of the clusters are available in the Table 10.

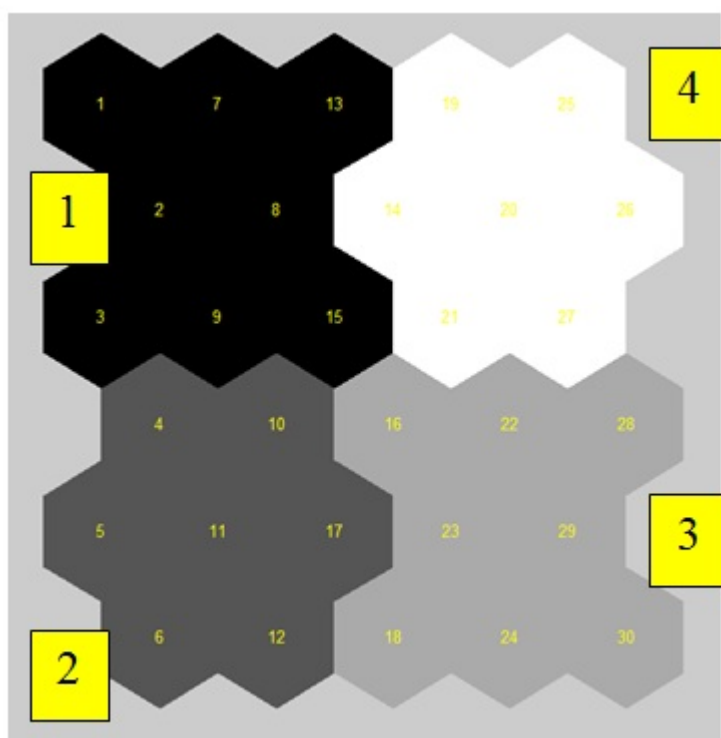|	n.	%
Cluster 1	155	26
Cluster 2	83	13,9
Cluster 3	196	32,8
Cluster 4	163	27,3
Tot	597	100,0

Table 10: *Cluster dimensions.*

Figure 4: *K-means cluster on the U-matrix*

Before getting into the description of the features of each cluster, so to elicit the archetypes for the online distribution of digital content that it depicts, two preliminary remarks are necessary.

First, the clustering analysis with SOM has been carried out, on the one hand, without ascribing each website to a given category of distributors (such as music distributors, video distributors and the like) and, on the other hand, without predetermining the numbers of clusters available. Indeed, this bottom-up approach has been considered the most capable of letting the existing models of distribution emerge without any interference from the prior knowledge of the sector.

Second, the following description of the archetypes does not focus on the nature of content distributed within each cluster because, as showed in Figure 5, a similar strategy of diversification is followed by the most websites of our sample, with the only difference of audio-visual content (i.e. Video) which predominate within Cluster 3. This difference of content distributed in each cluster is not sufficient, though, to justify the emergence of four diverse archetypes (see table 11), which – we believe – are the results of this minor feature *combined with* the major differences in terms of distribution methods, technological restriction, IPR management and the like that exist among the websites belonging to the clusters.
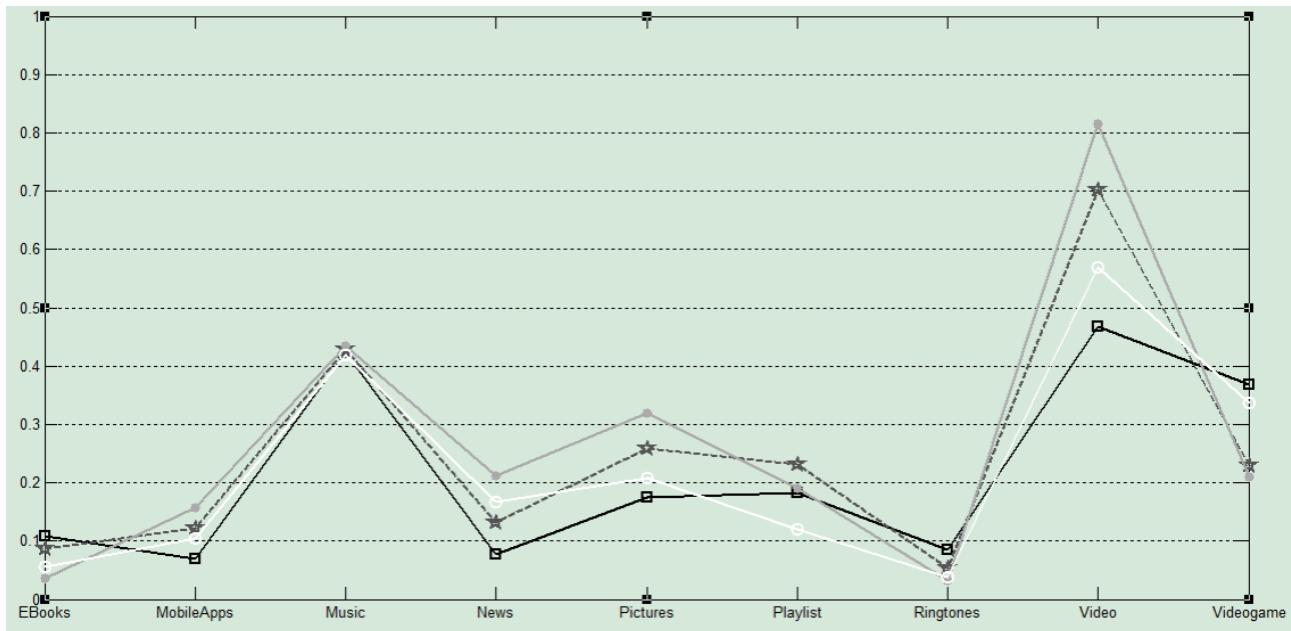

Figure 5: *Normalized values of the variables 'content' in the SOM codebooks*

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total	n.
Ebooks	68%	21%	0%	11%	100%	28
Mobile Apps	11%	17%	48%	25%	100%	84
Music	28%	13%	30%	28%	100%	302
News	3%	7%	59%	31%	100%	91
Pictures	12%	11%	56%	20%	100%	137
Playlist	29%	23%	35%	13%	100%	138
Ringtones	55%	13%	15%	18%	100%	40
Video	13%	15%	48%	24%	100%	385
Videogames	40%	11%	20%	28%	40%	158

Table 11: *Frequency distribution of variables 'content' across the four cluster*

6. The models of online distribution

The four clusters resulting from the analysis are characterized by specific individualizing features. The following table 12 shows the distribution of the main features across the four clusters.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	
Downloading	48%	21%	16%	15%	100%
Streaming	16%	14%	41%	29%	100%
Hosting	26%	14%	33%	27%	100%
All rights reserved on content	26%	14%	33%	27%	100%
Free with advertisement	5%	10%	58%	26%	100%
Registration required	38%	21%	36%	5%	100%
Geographical restrictions	26%	14%	33%	27%	100%
Software restrictions	33%	16%	23%	28%	100%
Platform / hardware restrictions	26%	14%	33%	27%	100%
No privacy policy	26%	1%	13%	60%	100%
Pay-per-download	26%	14%	33%	27%	100%
Subscription	26%	14%	33%	27%	100%
DRMed content	52%	25%	19%	4%	100%
Presence on Facebook	26%	14%	33%	27%	100%
Uploading allowed	10%	12%	71%	7%	100%

Table 12: *Frequency distribution of main features across the four clusters*

Cluster 1 groups 155 websites, i.e. the 26 per cent of the whole sample. In comparison to the other clusters it presents specific features in terms of distribution methods and revenue sources: it collects websites whose business strategies pivot almost exclusively around 'downloading' and 'pay per downloads'. In addition,

cluster 1 differentiates strongly from the other clusters as far as technological restrictions are concerned. Indeed, although DRMs are equally used by all the websites of our sample, the websites belonging to cluster 1 adopt other technological restrictions more frequently than those belonging to the other clusters. As to the IPR management adopted, however, cluster 1 does not stand out: similarly to the websites of cluster 3, those belonging to cluster 1 opt for the so-called 'all rights reserved regime' with respect to both the content available on the website and the content downloadable by users.

Cluster 2 is the smallest one: it collects 83 websites, corresponding to the 13.9 per cent of the sample. These websites share many features with those of cluster 1, such as the frequent adoption of technological restrictions (which are implemented by more than a half of the websites), and the chosen IPR regime. Indeed, also cluster 2 websites impose an 'all rights reserved regime' on website pages and content, although – and differently from cluster 1 websites – they are less likely to claim ownership over the content. Indeed, the most outstanding individualizing feature of cluster 2 websites is their focus on interactivity. First, at least a quarter of cluster 2 websites allows users to upload their content. Second, these websites offer, among the many services, *fora*, chat, reviews, recommendations, i.e. services that support interaction between users and 'community' activities. Third, many of cluster 2 websites are present on the main social network platforms. Other significant individualizing features of cluster 2 websites relate to distribution methods and sources of income. Indeed, websites belonging to this cluster implement distribution methods other than downloading, such as streaming as well as hosting, embedding, and mobile distribution. Furthermore, while these websites broadly use the pay-per-download method (adopted by more than a half of the websites belonging to this cluster), they also implement other methods, such as advertisement-based distribution (preferred, indeed, by more than one third of websites).

Cluster 3 is the largest one: it groups 196 websites, equal to the 32.8 per cent of the whole sample. It achieves the highest degree of interactivity. Indeed, two third of cluster 3 websites offer the possibility of uploading content, an increased number of *fora*, chats, and other interaction tools, as well as an enlarged presence on social networks and an increased number of links to the major platforms. In addition, the uploading regime has different characteristics in comparison to that adopted by websites of other clusters. In fact, in the majority of cases where upload is possible, cluster 3 websites allow users to keep the ownership on the uploaded content (95 websites out of 123), license their content under a non exclusive license (106 websites out of 123) and authorize the creation of derivative works by third parties (85 websites out of 123). However, even in this context, the most frequent IPR regime chosen on the content distributed is the 'all rights reserved'. With respect to the distribution methods and the sources of revenue, websites belonging to cluster 3 modulate their offer among all the available systems of distribution – with a prevalence of streaming and live streaming – and implement various systems of revenue, with a prevalence of ad-based systems. In contrast, the recourse to technological restrictions is significantly lower than in clusters 1 and 2 (less than 20 websites adopt DRMs), with the only exception of software restrictions. This is because streaming and live-streaming may require the installation of specific rendering application and its update.

Cluster 4 collects 164 websites, i.e. the 27.3 per cent of the sample. Websites belonging to this cluster share some similarities with those belonging to other clusters. For instance, similarly to websites of cluster 2 and 3, these websites prefer streaming and live streaming to the other distribution methods. Similarly to those of cluster 1, they offer a limited possibility of uploading, do not recognize users' ownership on the uploaded content, do not provide a high degree of interactivity and implement software and hardware restrictions. However, two specific features distinguish websites belonging to Cluster 4, namely the frequency of free distribution and the adoption of open licenses. This preference for openness is confirmed by a low implementation of technological restrictions, such as DRMs and geographical restrictions.

Interactivity		
	Low	*High*
Low	**Cluster 1 – N IC** • 'All rights reserved' • Downloading • DRMs • Pay-per-download • Limited UGC • Ownership on UGC	**Cluster 2 – IC** • 'All rights reserved' • Downloading *and* streaming • DRMs • Pay-per-download *and* ad-based • UGC allowed • No ownership on UGC
High	**Cluster 4 – NIO** • Open content licensing • Streaming / live streaming • No DRMs • Free distribution • Limited UGC • No ownership on UGC	**Cluster 3 – IO** • 'All rights riserved' • Streaming / live streaming • Limited DRMs • Ad-based • UGC allowed • No ownership on UGC

Openness (left side label)

Figure 6: *Models of online distribution*

In sum, the four clusters above identify four archetypes of the current on-line distribution, meaning four paradigms summarizing the changing ways in which nowadays on-line distribution is carried and developed. As shown in Figure 6, these archetypes can be described in terms of interactivity and openness and, accordingly, placed along two scales: one moving from a high level of interactivity down to a minimal degree of interactivity; the other one running from openness down to closeness. The word 'interactivity' stands for the attitude of websites to include users in their business model, via uploading and communities; the term 'openness' not only addresses the IPR regime adopted, but also focuses on the ability of websites to extract value from the whole relationship that they have with their users rather than from each single transaction that take place between websites and users.

Because of its closeness to the off-line commercial distribution, cluster 1 identifies the archetype of *non-interactive and close distribution* (say, 'NIC'). As to the absence of interactivity, consider that the real innovative features of on-line distribution, i.e. the website-to-user and user-to-user types of interactions, are almost inexistent within the websites belonging to cluster NIC. And, even in those few cases where a specific kind of interaction such as uploading is allowed, cluster NIC websites put a significant barrier to interaction

by acquiring full ownership on the uploaded content. As to the lack of openness, this derives both from the existence of a proprietary regime upon all the kinds of content available on websites – including uploaded content – and from the imposition of technological and geographical restrictions, coupled with the substantial lack of sources of revenue different from the pay-per-download system.

At the opposite end of the spectrum we can place cluster 3, which lets emerge the archetype of *interactive and open distribution* (say, 'IO'). Indeed, on the interactivity side, in cluster IO websites favour the involvement of users to the extreme that users are not only treated as the final points of the distribution activity, but they are also spurred to become producers (or 'pro-users'). Therefore, it should not come as a surprise that many of the websites belonging to cluster IO can be labelled UGC platforms, distributing far and foremost audio-visual content. On the openness side, while free distribution and open licenses are not represented significantly in cluster IO, and the 'all rights reserved' clause is likely to be imposed on the distributed content, these websites tend nonetheless to allow uploading and to leave ownership to the users over the uploaded content. This confirms what has been observed above as to fact that 'openness' do not derive only from the IPR regime.

The remaining clusters highlight two intermediate archetypes of on-line distribution, respectively the archetype of *interactive and close on-line distribution* (cluster 2) (say, 'IC'), and the archetype of non- *interactive and open on line distribution* (cluster 4) (say, 'NIO'). The former is characterized by a significant degree of interactivity, although its websites move towards the model of traditional distribution when it comes to the IPR regime and restrictions adopted. The latter counterbalances its preference for a very open IPR regime with a low level of interaction, as the flow of content is unidirectional from the website to the users.

7. Conclusions

The empirical analysis above illustrated gives evidence to the idea that the online distribution of digital content has not taken the shape of the 'celestial jukebox', rather that of a 'patchwork quilt'. In other words, while the current copyright regime strongly incentivizes proprietary and unidirectional distribution of digital content, diverse models of digital distribution have emerged that together cover the spectrum of the business means currently available to entrepreneurs so to match the demand. In particular, the cluster analysis conducted over a representative sample of websites distributing digital content online has allowed the identification of the key features of these models so to show that they all share, to different extent, the new features of interactivity and openness (as previously defined). Whilst this result confirms what the literature and case studies have already suggested as to the gap between the legal framework and the market reality, our empirical analysis allows us to make a step forward. It appreciates how legal, technological and economic features operate as 'differentiating factors' among the models of digital distribution. In other words, it answers the question on what are the key determinants of the models of online distribution and, to further narrow the question, to what extend does the choice of the copyright regime influence the model of distribution.

Consistently with the gap theorized between the legal framework and the market, the legal features do not represent the major differentiating factor among the identified archetypes. On the one side, 'openness' and 'closeness' are determined by other factors than the copyright regime. This means that 'open' models of distribution can coexist with a restrictive copyright regime ('all rights reserved') on the ground that they show a higher degree of interactivity, and similarly 'close' distribution models do not necessarily rely on proprietary regime. On the other hand, 'interactivity' does not depend much on legal factors, but rather on technological ones. Technology and business, not law, are the key discriminating factor between models of distribution. In our analysis, while the first two clusters share a similar proprietary legal regime, they diverge as to the main technology used to distribute the content (downloading in one case, streaming in the other) and the system to collect revenues (pay-per-download vs. ad-based distribution).

Finally, while technology and business are powerful differentiating factors, the models for online distribution seem not to be significantly influenced by the digital content which is distributed. To a certain extent, the same or similar model could apply to the distribution of music, videogames – and, conversely, the same digital content could be distributed according to different legal and business models. Although from the empirical analysis a predominance of certain models arise in relation to video, this is less likely in reason of the trend which has been observed since the beginnings of the digital age, namely the fact that the 'disembodiment' of information enhances the convergence of content by digital distribution. More than 'what'

it is distributed, it is and will be important 'how' it is distributed. And – contrary to what legislators perhaps expected – the legal framework is a less and less significant part of this 'how'.

In conclusion, we have seen that the current legislative framework strongly suggests a model that, when initially adopted, was supposed to be very appealing for the business, as it enabled an enhanced control over the content distributed. Yet, we have seen that entrepreneurs have developed a variety of different business models, which less and less coinciding with what suggested at the legislative level. In fact, we showed that the legal variables do not affect their strategic choices, as evidence that what in this context shapes distribution models is not the full control over the content distributed.

What is the cause of this phenomenon? As economic theory teaches, the supply follows the demand. Therefore, if the business moved toward models others from that suggested at the legislative level, it did it so to satisfy the demand. Hence, the main conclusion of our analysis is that the current legal framework does not answer to the needs and wants of consumers of digital content and that, on the contrary, they should be encompassed within a new and revised copyright law.

Bibliography

Adegoke Y (2006) 'MySpace to Sell Music from Nearly 3 Million Bands, Boston.com Business', http://www.boston.com/business/technology/articles/2006/09/02/myspace_to_sell_music_from_nearly_3_million_bands/

Benkler Y (2001) 'A Political Economy of the Public Domain: Markets in Information Goods vs. The Marketplace of Ideas', in Dreyfuss R, Zimmerman DL and First H (eds), Expanding the Boundaries of Intellectual Property (New York: Oxford University Press)

Bodt E, Verleysen M, Cottrell M (1997) 'Kohonen Maps versus Vector Quantization for Data Analysis' (Bruges: ESANN).

Boyd D, Ellison NB (2007) 'Social Network Sites: Definition, History, and Scholarship', 13 Journal of Computer-Mediated Communication 211.

Carlei V, Nuccio M, Sacco PL, Buscema M (2008) 'La complessità strutturale dei distretti industriali: un approccio basato sulle similarità multi-dimensionali', Scienze Regionali 2/08.

Deloitte (2007) 'The State of the Media Democracy, 'Second Edition'. Deloitte's Reality Check on the Future of Media', http://www.deloitte.com/assets/Dcom-UnitedStates/Local%20Assets/Documents/TMT_us_tmt/us_tmt_SOMD2_121411.pdf

Ficsor, M (2002), The Law of Copyright and the Internet (New York: Oxford university Press).

Fisher K (2008), 'Qtrax's free, ad-based P2P: Gnutella meets Zune-esque DRM', Jan 28 2008, http://arstechnica.com/tech-policy/2008/01/qtraxs-free-ad-based-p2p-gnutella-meets-zune-esque-drm/ .

Fisher W, Gasser U, Slater D, Smith M, Palfrey J (2005), 'Comments on the OECD Working Party on Information Economy Draft Report 'Digital Broadband Content: Music'', Berkman Center for Internet & Society at Harvard Law School 6, http://cyber.law.harvard.edu/digitalmedia/oecd_comments.pdf.

Gabbay N (2006) 'Facebook Case Study: Offline Behavior Drives Online Usage, Startup Review', http://www.startup-review.com/blog/facebook-case-study-offline-behavior-drives-online-usage .

Gasser U (2004) 'iTunes How Copyright, Contract, and Technology Shape the Business of Digital Media . A Case Study', Berkman Center for Internet & Society at Harvard Law School Research Publication No. 2004-07, http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/GreenPaperiTunes041004.pdf .

Gasser U (2005a) 'Copyright and Digital Media in a Post-Napster World: 2005 Update (January 2005)', Berkman Publication Series No. 2005-03, http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/wp2005.pdf.

Gasser U (2005b) 'Copyright and Digital Media in a Post-Napster World: International Supplement', Berkman center, http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/wpsupplement2005_0.pdf .

Gasser U (2005c) 'iTunes: How Copyright, Contract, and Technology Shape the Business of Digital Media - A Case Study', Berkman Center for Internet & Society at Harvard Law School Research Publication No. 2004-07, http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/iTunes_August_update_final[1].pdf ;

Gasser U, Slater D, Smith M, Palfrey J, McGuire M (2003) 'Five Scenarios for Digital Media in a Post-Napster World', Berkman center, http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/2003-07.pdf .

Giblin, R (2012), Code Wars: 10 years of P2P Software Litigation (Cheltenham: Edward Elgar).

Ginsburg JC (2002) 'Essay – How Copyright Got a Bad Name For Itself', 26 Columbia Journal of Law and the Arts 1.

Goldstein, P (1994), Copyright's Highway. From Gutenberg to the Celestial Jukebox (New York: Hill & Wang) (second revised edition, Stanford, CA: Stanford University Press: 2003).

Guibault L, Westkamp G, Rieber-Mohn T, Hugenholtz B et al (2007)'Study on the implementation and effect in member states' laws', Final Report, Institute for Information Law, University of Amsterdam, The Netherlands, 2007, http://www.ivir.nl/publications/guibault/Infosoc_report_2007.pdf.

Helberger N, Guibault L, Janssen E H, Van Eijk N, Angelopoulos C J, Van Hoboken J, Swart E et al (2008) 'User-Created-Content: Supporting a participative Information Society', Final Report, Idate, Tno and IViR,http://www.ivir.nl/publications/helberger/User_created_content.pdf.

Holden B, Hwang R, Lazebnik R, Palfrey J, Rigamonti C, Slater D, Wentworth D, McGuire M, Baker V, McNealy PJ (2003) Copyright and Digital Media in a Post-Napster World, 2003, Berkman center, Research Publication No. 2003-05 8/2003, http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/2003-05.pdf .

Hugenholtz B, Van Eechoud M, Van Gompel S et al (2006) 'The Recasting of Copyright & Related Rights for the Knowledge Economy', Final report, Institute for Information Law, University of Amsterdam, The Netherlands, 2006, http://www.ivir.nl/publications/other/IViR_Recast_Final_Report_2006.pdf .

Huygen A, Rutten P, Huveneers S, Limonard S, Poort J, Leenheer J, Janssen K, Van Eijk N, Helberger N (2009) 'Ups and downs. Economic and cultural effects of file sharing on music, film and games', TNO-iVir,http://www.ivir.nl/publicaties/vaneijk/Ups_And_Downs_authorised_translation.pdf .

Kohonen, T (2001), Self-organizing maps, 3rd ed. (Berlin: Springer-Verlag).

Kuo RJ, Ho LM, Hu CM (2002) 'Integration of self organising feature map and K-means algorithm for market segmentation', 29 Computers & Op. Res, 1475.

Litman J (2002) 'War Stories', 20 Cardozo Arts & Ent. LJ 337.

McGuire M, Slater D (2005) 'Consumer Taste Sharing Is Driving the Online Music Business and Democratizing Culture', Berkman center, http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/11-ConsumerTasteSharing.pdf .

Patry, W (2009), Moral Panics and the Copyright Wars (New York: Oxford University Press).

Peterson L (2007) 'Internet Ad Spending Set to Overtake All Other Media by 2011: VSS', Media Post Publications, http://publications.mediapost.com/index.cfm?fuseaction=Articles.san&s=65282&Nid=33013&p=352811 .

Rosenblatt B (2004), 'Learning from P2P: Evolution of Business Models for Online Content', INDICARE, http://www.indicare.org/tiki-read_article.php?articleId=61.

Samuelson P (1999), 'Intellectual Property and the Digital Economy: Why the Anti-Circumvention Regulations Need to Be Revised', 14 Berkeley Tech. L.J. 519.

Slater D, Gasser U, Smith M, Derek EB, Palfrey JG (2005) 'Content and Control: Assessing the Impact of Policy Choices on Potential Online Business Models in the Music and Film Industries', Berkman Publication Series Paper No. 2005-01, http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/content_control.pdf .

Smith D (2006) 'Online Social Networks & Communities are Here to Stay, Know More Media, http://www.knowmoremedia.com/2006/09/online_social_networks_communi.html .

Trusov M, Bucklin RE and Pauwels K (2008) 'Effects of Word-of-Mouth versus Traditional Marketing: Findings from an Internet Social Networking Site' (Working Paper Series), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1129351.

---

[1] Maurizio Borghi is Senior Lecturer at Brunel University Law School, London; Maria Lillà Montagnani is Associate Professor of Commercial Law at Bocconi University, Milan, Department of Law and faculty member of ASK research centre; Mariateresa Maggiolino is Assistant Professor of Commercial Law at Bocconi University, Milan, Department of Law; Massimiliano Nuccio is visiting Professor for Culture and Region at the Leuphana University of Lüneburg. The article is based on research conducted within Counter – Counterfeiting and Piracy Research, a project funded by the European Commission's Framework 7 programme, 2007-2010. Authors wish to thank research assistants that helped to collect the data: Aura Bertoni, Elisa Brodi, Indranath Gupta and Ana Haiman.

[2] For a critique of this statement see Benkler Y 2001.

[3] WIPO Copyright Treaty (WCT) and WIPO Performances and Phonograms Treaty (WPPT) of 20 December 1996.

[4] Agreement Statements concerning the WCT, Article 1(4)

[5] So-called 'making available right' (WCT, art. 8 and WPPT, arts. 10 and 14)

[6] WCT, art. 10 (extending the Berne Convention's 'three step test' to the limitations or exception that the Contracting Parties 'may' provide for with respect to the rights introduced by the Treaty).

[7] WCT art. 11, WPPT art. 18.

[8] WCT art. 12, WPPT art. 19.

[9] The main Industry Association Members Lists used to compose the dataset are: Motion Picture Association of America, Entertainment Software Rating Board, Academy of Interactive Arts and Sciences, Entertainment Software Association, International Game Developers Association, Associazione Editori Software Videoludico Italiana, International Federation of Phonographic Industry.

[10] Digital Millennium Copyright Act, Pub. L. No. 105-304, 112 Stat. 2860 (Oct. 28, 1998) (hereinafter DMCA).

[11] European Parliament and Council Directive 2001/29, 2001 O.J. (L 167) 10-19.

[12] Even in the case of Starz the service offered is now much more articulated: http://www.starz.com/channels.

[13] The service is not available any more, yet the description of the terms of use of 'Napster To Go' at June 30, 2006 is still available under 'Napster, Terms and Conditions', http://www.napster.com/terms.html.

[14] The Weedshare service has been suspended. See http://www.weedshare.com/.

[15] The service offered by Mjuice has been suspended and afterword the website has been closed.

[16] The Internet Underground Music Archive is no longer available. For an archived version of the site: http://web.archive.org/web/*/http://www.iuma.com.

[17] The index used to identify the popularity of a website on Google Ad Planner has been the 'unique visitors' index, which was set on different levels for each category of digital service (e.g. music, video, etc.): the higher the 'unique visitors' index was, the more the site was considered to be massively used and, consequently, popular. The levels used to classify a big player (e.g. 1 Million unique visitors or 5 Million unique visitors) have been set up taking as reference the unique visitors index of those services considered to be the most popular digital services by the industry reports.

[18] Music websites have been randomly collected using the following sources: IFPI Digital Music Report 2010 and IFPI Digital Music Report 2011; Pro Music (pro-music.org), the coalition of music record companies and retailers; list of case studies available at the Creative Common Licence website; list of top websites resulting from Google AD Planner, 'search by audience'. The resulting dataset has been further reduced and

modified in time, maintaining only the working websites, since many of them have been closing or changing the service or changing their position in the Google Ad Planner charts during the period of analysis.

[19] Webradios have been randomly collected using the following sources: IFPI Digital Music Report 2010 and IFPI Digital Music Report 2011; lists of web radio sites resulting from a Google search entering the keywords 'web radio list', in particular: onlineradiostations.com, web-radio.com, radiodirectory.com; list of popular web radios available on Wikipedia, entering the key words 'List of Internet stations'; list of top websites resulting from Google AD Planner, 'search by audience'. In Google Ad Planner search, the following settings have been selected: geography: all countries; ranking method: best match (results balanced between big and small sites); category: radio.

[20] Video websites have been randomly collected using the following sources: Motion Picture Association of America members list; Federazione Antipirateria Audio Visiva; list of top websites resulting from Google AD Planner, 'search by audience'; list of video service on Wikipedia, entering the keywords 'comparison of video services' and 'comparison of video hosting services'; lists of video services resulting from a Google search entering the keyword 'online video services', in particular: golem.it. The resulting dataset has been modified in time, since some websites have been changing the service or changing their position in the Google Ad Planner charts during the period of analysis.

[21] Web tv services have been randomly collected using the following sources: lists of web radio sites resulting from a Google search entering the keyword 'web tv list', in particular: golem.it, internettvlist.com, wwitv.com; and the list of top websites resulting from Google AD Planner, 'search by audience .

[22] Videogame online distribution websites have been randomly collected using the following sources: members lists of Entertainment Software Association, International Game Developers Association, Associazione Editori Software Videoludico Italiana, Entertainment Software Rating Board, and Academy of Interactive Arts and Sciences; list of top websites resulting from Google AD Planner, 'search by audience', composed by matching two different results of 'search by audience' as resulting from different ranking methods: 'best match' and 'audience reach' (category selected: video games / online games'); lists of videogames services resulting from a Google search entering the keywords 'online games list'. The peculiarity of the videogames section is that a part of it is represented by websites which are not available on Google AD Planner search: the reason for it is that many services of this section are 'one game sites', which have too low traffic rates to be recorded by Google AD Planner. Consequently, no information about them is available on such source and they are not included in the official lists, but they are however included in the dataset because of the characteristics of their service, which are coherent with the rest of the videogames section.

[23] According to Alexa, about one fourth of global internet users visit YouTube daily (http://www.alexa.com/siteinfo/youtube.com; last visited November 2010).