

CONTENT PRODUCTION AND PERCEPTION OF COPYRIGHT: AN ANALYSIS OF HABITS AND BELIEFS OF INTERNET USERS

A paper by Simone Aliprandi and Andrea Mangiatordi¹, published in [European Journal of Law and Technology](#), Vol. 4, No. 3 (2013) and released under a [Creative Commons Attribution - ShareAlike 3.0 unported license](#).



Please cite as: Aliprandi, S. & Mangiatordi A., "*Content production and perception of copyright: an analysis of habits and beliefs of internet users*", *European Journal of Law and Technology*, Vol. 4, No. 3, 2013.

More details about the project: www.aliprandi.org/copyrightsurvey. Other related articles: <http://copyrightsurvey.blogspot.it/p/related-articles.html>.

1 Simone Aliprandi (PhD): www.aliprandi.org; Andrea Mangiatordi (PhD): www.mangiatordi.net.

Table of content

1. Introduction – 2. State of the art – 2.1. "Behaviors of consumers of digital content in Italy. The case of file sharing" (Fondazione Luigi Einaudi 2007) – 2.2. "Discovering behaviors and attitudes related to pirating content" (PwC 2010) – 2.3. The report on innovation culture in Italy (Wired-Cotec 2009) – 2.4. The Global software piracy study (BSA 2010) – 3. The research – 3.1. Questionnaire's structure – 3.2. Dataset description – 4. Method and results – 4.1. Principal Components Analysis – 4.2. Cluster Analysis – 5. Discussion and Conclusions – Bibliography.

Abstract

This article presents an analysis based on data from an open dataset made available by an independent empirical research on the topic of "Copyright in the digital age". The original data includes information about attitudes, social perceptions and levels of awareness.

After describing the initial objectives of the research and presenting the set of variables considered, an in-depth analysis is here discussed, focusing on the similarities and differences encountered in distinct categories of respondents.

The main objective of this analysis is to test the soundness and the descriptive power of the classification proposed in the original study, which divided respondents into the four categories of "Generic", "Active", "Creative" and "Professional creative" users. Other possible clusters are identified here, and their possible associations with those groups is discussed.

Results indicate that the views of authors of digital content about the practice of illegal download tend to differ from those of non producers. Non publishers of digital content seem to be almost equally distributed between those who are against download, and those who tolerate it.

Introduction

This article describes a second and more specific work of analysis and interpretation on a dataset collected through an independent empirical research which was conducted between February and June 2011 as a Ph.D. final project. The title of the original study was "Copyright in the digital age: attitudes, social perception and level of awareness". Discussion about the main results of this initial survey are available both on the web² and in articles appeared on open access journals (Aliprandi 2012a, Aliprandi 2012b, Aliprandi 2013).

The initial research study aimed at providing an impartial approach to the analysis of intellectual property issues, taking into direct account the viewpoints of "average citizens", as opposed to those of stakeholders and experts. As a matter of fact, very few studies address such framework so far. Most of them have been produced by major entertainment or software companies in order to monitor and anticipate market trends, thus considering common individuals as consumers and potential buyers, instead of users in more general terms. This strategy gives way to a distorted perception of collected data, or at least provides a very limited perspective on their outcome and analysis.

Therefore, this study focuses on a different perspective and will address three broader research themes:

- the most common behaviors of Internet users when they get, distribute, or otherwise deal with online content under copyright;
- the average perception of copyright itself, that is, whether users see it as a primary or minor problem, a useful tool or a useless burden, etc.;
- the level of awareness of Internet users about mechanisms and principles currently governing copyright law, in order to expose their

2 <http://www.aliprandi.org/copyrightsurvey>

actual level of knowledge on related issues.

In this work, the descriptive information is not going to be further commented: the intent is rather to learn more about who answered the original questionnaire, by tracing user profiles adopting a bottom-up approach based on correspondence in answers.

1. State of the art

Barring a few exceptions, empirical studies currently available on the subject of copyright take only superficially into account aspects such as the perception of behaviors, the guilt that users can feel, the level of awareness of the rules. They mainly dwell (perhaps excessively) on the quantitative aspects relating to the time spent online, the type of activities carried out online, the method of acquisition and use of the content. This implies that we have a lot of data (well sampled and very specific) from which it is possible to draw several conclusions and on which it is possible to make comparisons and various analyzes.

For example, in Italy, the National Institute of Statistics (ISTAT) carries out periodic surveys of a nationally representative sample and periodically releases statistics on the use of technologies and consumption of entertainment products (ISTAT 2010). The same applies to a number of studies promoted by the EU institutions.

Despite these quantitative studies do not relate specifically to the subject of the acquisition and use of copyrighted content, it is still possible to find some useful data about this specific topic, which can be isolated, extracted and reprocessed in order to reach further conclusions.

Among this high number of studies there are some that are closer to the spirit of the research presented here: they are described briefly in the following paragraphs.

1.1. "Behaviors of consumers of digital content in Italy. The case of file sharing" (Fondazione Luigi Einaudi 2007)

The study that comes closest to the spirit and purpose of this paper is definitely the one conducted in 2007 by the Fondazione Luigi Einaudi, a research institution operating in Italy.

The study was divided in three parts: in the first one, referred to as "Desk Research", the authors analyzed on the one hand the available literature on the phenomenon from 2001 to 2005 and on the other the previous research that could provide useful information. A second phase, referred to as "Survey Research", was based on a questionnaire that, after the identification of a representative sample of the Italian Internet population, was administered via telephone to a total of 1600 respondents. Finally, a third phase referred to as "Websurvey Research" was based on a questionnaire published on the website of the University "La Sapienza" of Rome; this phase had the final outcome of 388 useful answers collected from the web.

One of the most interesting points in this research is probably the creation of specific categories of users based on the behaviors they declared in the survey. Researchers identified four main categories: the "downloaders" (those who said they downloaded music or movies from the Internet in the previous year), which were in turn divided into "downloaders pay" (those who downloaded from sites or online paying a proper fee) and "downloaders free" (those who obtained content for free from other users). On the other hand there were the "non-downloaders", which were in turn divided into "aware non-downloaders" (who know how to download content from the Internet) and "unaware non-downloader" (who are not aware of this possibility).

1.2. "Discovering behaviors and attitudes related to pirating content" (PwC 2010)

The research conducted by PricewaterhouseCoopers (PwC) consisted of a questionnaire administered online in September of 2010 to a sample of 202 people, aged between 18 and 59, who admitted to have performed online piracy activities in the previous six months.

The collected data was published in January 2011 in a report addressing eight key research themes. They mainly referred to the relationship between the so-called piracy and the "temptation of gratuity", the willingness or unwillingness of users to pay for the use of online content, the correlation between downloading and streaming, and finally, the impact of the new possibilities of "piracy via mobile".

This research is interesting primarily because it comes from one of the most influential individuals specialized in the online administration of surveys on a global scale and denotes a solid methodological approach. Although the sample is quite small, it was established in full compliance with the standards of statistical representativeness.

1.3. The report on innovation culture in Italy (Wired-Cotec 2009)

In the spring of 2009 the Italian edition of Wired magazine published a study containing some infographics taken from a 2009 report titled "The culture of innovation in Italy", written in collaboration with Cotec foundation.

The report is mainly focused on issues related to the concept of innovation in a broader and more general sense (biotechnology, genetic engineering, energy and environmental policies, telecommunications, power supply) and reserves only a smaller amount of space to the subject of copyright. However, the data provided in that part is particularly interesting, as it represents one of the first cases in Italy (along with the research by Fondazione Luigi Einaudi) in which the research

investigates not only behavior, but also users' opinions.

1.4. The Global software piracy study (BSA 2010)

The Business Software Alliance (BSA) is a multi-national organization that brings together all the major realities of production technologies and acts as a unique spokesman for these stakeholders, especially on issues related to intellectual property enforcement.

Every year BSA commissions, supports and disseminates studies and statistics relating to major issues affecting the software market.

A 2010 report was based on a survey involving a sample of about 15000 pc users (both business and private) stratified into the levels of geographical origin (representing 32 countries), technological expertise and differences in social status.

The research also focused on some socio-cultural aspects linked to the method of acquisition of software. Among the questions there were some aiming at highlighting the perception of legitimacy of some behaviors related to the acquisition of software and the level of awareness by users. Some of the questions were about the perception that users have of the more general phenomenon of "intellectual property."

2. The research

In this section, some background information will be provided about how the research was designed and conducted. The dataset was collected through a self-administered Computer Assisted Web Interview (CAWI), and the set of questions is still available online³.

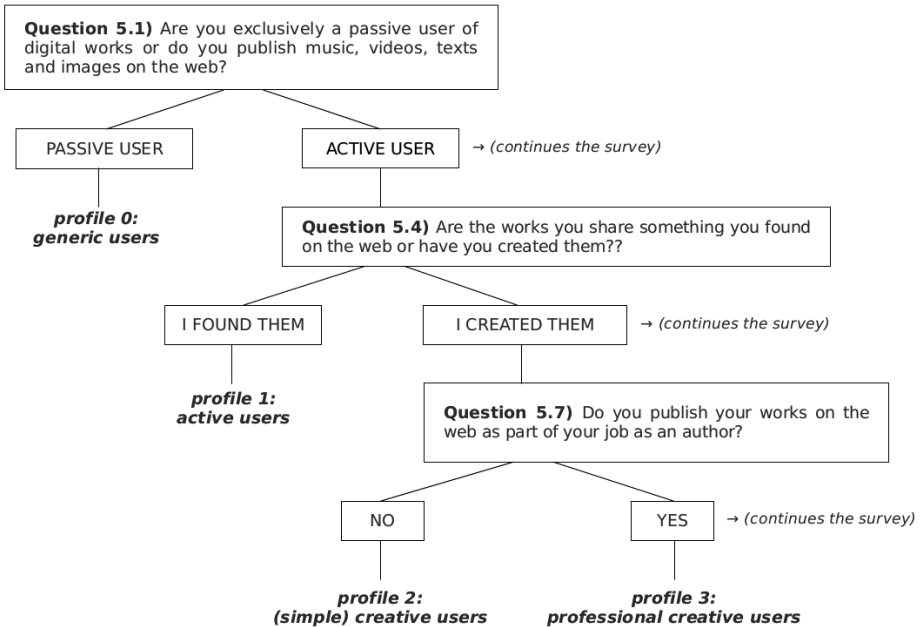
2.1. Questionnaire's structure

The questionnaire included:

- 9 general questions about demographic data;
- 35 questions requiring an answer from all respondents, which included:
 - 15 questions about behaviours;
 - 14 questions about opinions and perceptions;
 - 6 questions about the level of awareness about copyright-related issues;
- 10 questions addressing specific sub-categories of respondents: access to these questions was filtered on the basis of the answers given to specific questions about use and production of content.

Section 5 included in-depth questions about specific online activities based on previous filter questions, in order to group respondents in specific sub-categories, as shown in the following chart:

³ <http://copyrightsurvey.blogspot.it/p/the-questionnaire.html>



The questionnaire was localized in Italian and English. It was accessible for 120 days (from February 1st to June 1st, 2011). Given its target and content, the questionnaire was promoted mostly over the Internet. A press release was issued on February 1st, 2011 announcing the research project and the publication of the questionnaire. Initially this press release – both in English and in Italian – was posted on several websites about digital activism and IT topics, with a clear invitation to re-distribute it on any place and channel deemed appropriate. Mainstream social networks such as Facebook and Twitter also proved to be effective distribution channels. Frequent posting of news items, events notices, status updates, and short messages – all including a direct link to the research description page – was complemented by additional promotion at conferences and university lectures.

2.2. Dataset description

The results of the survey are divided in two different studies: Study 1 (Italy) and Study 2 (rest of the world). This selection is based on the answers to question n° 1.3 ("country") and not on the language of the questionnaire.

The questionnaire was completed by 1735 people, with 1289 for Study 1 (Italy) and 446 for Study 2 (rest of the world). This amount includes only the valid answers, meaning that only the questionnaires correctly filled at least till question 4.5 included were considered. Only study 1 will be taken into consideration in this paper.

Male respondents in Study 1 were 739 (57.3%), while 550 were female (42.7%). Over 60% of all respondents is below the age of 34. Therefore, many qualified themselves as students (40%), with 51% of them holding a high school diploma and only 36.8% of them saying to hold a basic University degree. The majority of respondents (993, 72.4%) lived in Northern Italy, while 16.5% were from Central Italy and 11.1% from Southern Italy and Isles. The most common kind of Internet access was broadband (95% of the whole group).⁴

3. Method and results

The analysis performed on the dataset followed two steps: first, data from different variables was aggregated, operating a dimensional reduction of the dataset itself by the means of Principal Component Analysis. Second, the principal components were used to create newer variables and the sample was classified by those variables, applying clustering analysis. This allowed to define tendencies in the way respondents agreed or disagreed with different kind of assertions. We then compared the clusters with demographic data and with the original four categories of "Generic", "Active", "Creative" and "Professional creative" users, in order to

4 A complete and detailed report (with more than 500 charts) is available online at <http://copyrightsurvey.blogspot.it/p/complete-report.html>.

describe them using deeper indicators.

3.1. Principal Components Analysis

The principal component analysis of the dataset variables (rotated using the Varimax method), calculated using the “psych” package within the R software package, allowed to define 8 factors with eigenvalue above 1. For each factor only the variables with loadings above 0.35 were considered.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
download crime	0.79							
download damages economy	0.75							
download stealing	0.73							
copyright should be strongly enforced	0.71							
download damages authors	0.69							
download needs to be circumscribed	0.68							
download improper	0.67							
itunes or similar store		0.8						
get contents from online store		0.79						
get software from official website		0.7						
buying after download		0.51						
buying software from store		0.43						
get contents from filesharing			0.79					
search in p2p networks			0.79					
get cracked from p2p			0.61					
have all music on harddisk			0.56					
get contents from friends				0.83				
get software from friends				0.7				
afraid of beeing caught					0.87			
feeling guilty					0.83			
search on youtube						0.7		
get contents full streaming						0.65		

get cracked from p2p	0.37
get software from internet	0.71
rip after buying	0.39
download damages companies	0.77
download common behaviour	0.59

The eight factors were then described and given a label, in order to use them to classify respondents in the following cluster analysis. Some of them were quite predictable, while some showed interesting connections.

PC1 - **download_criminal**: this factor is composed by all the variables related to bad feelings generated by the act of illegally downloading digital content; there is substantial coherence between opinions about illegal download (considering it as a crime is the variable with the highest loading here) and the idea that it should be regulated, or that copyright should be strongly enforced;

PC2 - **paid_download**: whether it is from a specific market (e.g. iTunes) or from a website, people who declared to be used to buying online from official sources seemed to do it using multiple channels;

PC3 - **p2p_download**: this factor connects the use of p2p networks to the fact that many respondents already own a huge amount of contents and sometimes they simply need that in order to be satisfied; this relationship suggests p2p use to be linked to a sort of “collecting practice”, as people download content and accumulate it;

PC4 - **friends**: the two variables composing this factor are easily associated by the fact that both are related to getting contents from friends, unifying the aspects of a real life sharing activity;

PC5 - **bad_feelings**: also in this case the association was quite

predictable, as the sense of guilt and the fear of being prosecuted for copyright violations pertain to the same domain of “having bad feelings” about something;

PC6 - **streaming_piracy**: streaming contents, be it from sites like YouTube or from less copyright compliant sources correlates with the download of cracked software; those acts have a lower respect of copyright in common;

PC7 - **digital_support**: a connection between software download and the habit of ripping content suggests that users have now the possibility of choosing the best support for the content they consume, and that material goods are being sent to the background as long as digital storage is more convenient;

PC8 - **download_damage**: this last factor was probably the hardest to interpret, as it links the idea that download can damage companies with the belief it is unacceptable, yet tolerated; we interpreted it as an indicator of the idea that damaging companies is not as bad as damagin authors, therefore it can be tolerated.

3.2. Cluster Analysis

The above described variable groupings were used to build indexes corresponding to each factor. Every respondent got a score in the eight newly generated variables and this data was used in the following analysis. The following table shows the centroids of four clusters identified by the K-means algorithm, calculated using the “cluster” package within the R software package. Values above the row mean are highlighted using bold-face.

	uncomfortable users	online/offline consumers	collectors	detached users
download_crime	3.13	2.11	1.72	2.82
paid_download	2.38	1.83	2.24	2.07
p2p_download	2.18	3.29	2.94	1.68
friends	2.39	3.34	1.92	2.28
bad_feelings	4.00	2.51	2.59	1.72
streaming_piracy	2.72	3.58	2.99	2.30
digital_support	3.38	3.63	4.07	2.68
download_damage	3.64	3.85	3.81	3.65

Four groups were hence individuated, which could be described as follows:

1. **Uncomfortable users (234 respondents):** they tend to think that downloading copyrighted content is a bad thing, and they feel guilty for it more than all the other groups, which is probably why they avoid it;
2. **Online/offline consumers (418 respondents):** they get content from various channels, including p2p and streaming, and they also tend to receive it from their networks of friends;
3. **Collectors (348 respondents):** they get content from any channel, also from the paid ones, and tend to strongly disagree with the idea that downloading is bad or criminal; they also have the highest rate of preference for digital supports; they are somehow complementary to the previous group;
4. **Detached users (289 respondents):** they also condemn download as a criminal activity, and generally do not practice it; as a result, they are those with the lowest guilt and fear feelings about such a behaviour.

In order to better understand the nature of the above described four groups, an analysis of their relationship with demographic variables was performed.

Gender: male respondents were mainly classified as *collectors* (37.2%) or as *online/offline consumers* (26.8%). In this last group there was high presence of female respondents (40% of them). Women had as a second most frequent outcome that of *detached users*.

	uncomfortable users	online/offline consumers	collectors	detached users	total
Male	18.3%	26.8%	37.2%	17.7%	100.0%
Female	18.0%	40.0%	13.3%	28.7%	100.0%

Age group: younger respondents tended to fall into the *online/offline consumers* category, with no substantial differences between those under the age of 25 and those in the 25-34 range. Almost half of older respondents (54+ years old) were classified as *detached users*. The two intermediate age groups (35-44 and 45-54) had similar distributions in both the *uncomfortable* and *consumers* groups, but had inverse distributions in the two categories of *collectors* and *detached users*. Again, older respondents were more frequently classified as *detached*.

	uncomfortable users	online/offline consumers	collectors	detached users	total
<25	15.2%	43.4%	22.0%	19.4%	100.0%
25-34	17.6%	43.8%	26.7%	11.8%	100.0%
35-44	25.4%	14.7%	35.7%	24.2%	100.0%
45-54	22.4%	14.7%	26.3%	36.5%	100.0%
54>	8.8%	8.8%	33.3%	49.1%	100.0%

Respondent categories: the original dataset included a classification of respondents based on specific filtering questions. Comparing this top-down classification with the bottom-up approach proposed in this paper it is possible to visualize some specific correspondence patterns.

	uncomfortable users	online/offline consumers	collectors	detached users	total
Generic user	18.68%	31.92%	18.00%	31.41%	100%
Active user	15.13%	46.22%	26.89%	11.76%	100%
Creative user	19.39%	27.21%	37.76%	15.65%	100%
Professional creative user	18.45%	23.81%	39.88%	17.86%	100%

Generic users have a strong presence in both the *consumers* and *detached* groups. Summing up *online/offline consumers* and *collectors* (two groups with very similar centroids, that could be seen as complementary) on one side, *detached* and *uncomfortable users* (again, groups with lots of similarities) on the other, it is clear that generic users are almost perfectly distributed between two fronts: people who see download of content as something negative, and people who don't.

A strong majority of Active users falls into the *online/offline consumers* group, with a strong presence among the *collectors*, too.

The relative majority of both kinds of creative users can be classified as *collectors*. In both cases the second most likely group for them is that of *online/offline consumers*, which differs for the “friends” factor.

Discussion and Conclusions

The analysis performed here allowed to build a bottom-up classification of the respondents of a previous questionnaire based on the analysis of patterns in declared behaviour. The profiling system of the respondents already present in the dataset, realized with filter questions, was rediscussed in the light of this new classification. The main finding is that categories such as “active” or “inactive” only partially describe the positions that can be assumed by internet users towards copyright and download of digital creative content.

As already highlighted in the previous section, inactive users (those who only consume digital content, but do not upload content as a habit) split almost perfectly among positive and negative views of download. The fact of being active, on the contrary, seems to determine a shift towards a more favorable view. Being creative, in the sense of being producers of original contents, seems related with the view that an internet user can form about copyright. Interestingly enough, the data shows an higher tendency towards the download of digital content by creative users. Finally, there is no great difference among “creative” and “professional creative” users (the latter ones being those who gain money from their publishing content online).

The future perspective is to use this pilot research to build a better designed survey for monitoring the perception and attitudes of users towards the issue of copyright, in different countries and contexts. The analysis carried out here will allow to redesign the original questionnaire in a more effective way, by reducing the number of redundant questions and by refining the classification criteria of respondents.

Bibliography

- Aliprandi, S. (2012a). Measuring the so called “piracy”: a commented review of the most important empirical studies. *SCIENTIFIC RESEARCH AND INFORMATION TECHNOLOGY*, 2(1), 59–82.
- Aliprandi, S. (2012b). Copyright, from criminalisation to normative efficacy. *CIBERSPAZIO E DIRITTO*, 13(45), 147–166.
- Aliprandi, S. (2013). Copyright in the digital era: a pilot on behaviours, social perception and consciousness. *JOURNAL OF LIBRARY AND INFORMATION SCIENCE*, 4(2), 45–83. Retrieved from <http://leo.cilea.it/index.php/jlis/article/view/8802> – last visited on 9th July 2013
- BSA (2011). *Eight annual BSA Global Software Piracy Study* (pp. 1–18). Retrieved from http://globalstudy.bsa.org/2010/downloads/study_pdf/2010_BSA_Piracy_Study-Standard.pdf – last visited on 9th July 2013
- Customer Insight Group - New York Times (2011). *The psychology of sharing: why do people share online?* (pp. 1–46). Retrieved from http://nytmarketing.whsites.net/mediakit/pos/POS_PUBLIC0819.php – last visited on 9th July 2013
- Fondazione Luigi Einaudi (2007). *I comportamenti di consumo di contenuti digitali in Italia. Il caso del file sharing.* (pp. 1–68). Retrieved from http://www.anica.it/online/attachments/033_filessharing_report.pdf – last visited on 9th July 2013
- ISTAT. (2010). *Cittadini e nuove tecnologie (anno 2010)* (pp. 1–25). Retrieved from <http://www.istat.it/it/files/2011/01/testointegrale201012232.pdf?>

[title=Cittadini++e+nuove+tecnologie+-+23%2Fdic%2F2010+-+testointegrale20101223.pdf](#) – last visited on 9th July 2013

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2013). *cluster: Cluster Analysis Basics and Extensions. R package version 1.14.4.*

PricewaterhouseCoopers (2010). *Discovering behaviors and attitudes related to pirating content* (pp. 1–7). Retrieved from <http://www.pwc.com/us/en/industry/entertainment-media/assets/piracy-survey-summary-report-0111.pdf> – last visited on 9th July 2013

R Core Team (2013). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. URL <http://www.R-project.org/>

Revelle, W. (2013) *psych: Procedures for Personality and Psychological Research. R package version 1.3.2*

WIRED/COTEC (2009). *La cultura dell'innovazione in Italia* (pp. 1–71). Retrieved from http://www.cotec.it/it/wp-content/uploads/2009/06/cultura_innovazione_italia_rapporto2009.pdf – last visited on 9th July 2013