Privacy and Intellectual Property Rights

Edoardo Righi

December 6, 2019

1 Lecture 1

Attending students will have to submit a short paper for the exam. During the course, students will have to provide three short summaries of the course contents. Summary will be handled using the link provided in the main page of this course. The deadlines for handling the summaries are:

- 30/10/2019 (content of lectures until 24/10/2019)
- 5/12/2019 (content of lectures between 31/10/2019 and 28/11/2019)
- 31/12/2019 (content of lectures between 4/12/2019 and 19/12/2099)

2 Lecture 2

A formal definition of the Law:

> *"a body of rules to regulate the social behavior of a community of people in a given territory"*

But Law is also a formal definition of the values that define a society: in some cases it reflects the common sense of a population, in others it tries to shape it. Particularly in this last case, the lawmaking is inspired by the most prestigious ideologies; for example, the Napoleonic Code after the French Revolution, or the US Law in recent years, influenced the Law of a lot of other countries. This is more recurring for nations that have similarities or deals (like Italy and Germany before WW2).

A scientific approach to the Law consists in the categorization of legal rules according to their provenience and their content: Public versus Private Law.

- **Public Law**: the area of the law governing the relationship between individuals (citizens, companies) and the state. Constitutional law, administrative law and criminal law are thus all sub-divisions of public law.

  - Constitutional Law: *"Constitutional law is the study of foundational laws that govern the scope of powers and authority of various bodies in relation to the creation and execution of other laws by a government"*.
  - Criminal Law: also known as penal law, is the body of law that punishes criminals for committing offenses against the state or other individuals or organizations.
  - Administrative Law: is the body of law that arises from the activities of administrative agencies of government.

- **Private Law**: that part of a legal system that involves relationships between individuals such as the law of contract or torts as it is called in the common law and the law of obligations as it is called in civilian legal systems.

  - Civil Law: regulates economic relationships amongst persons and private organizations.
  - Family Law: regulates (economic and non-economic) relationships between married individuals and between them and their children.
  - Commercial Law: regulates the creation of firms, economic organizations (corporations), and their relationships with consumers.

3   Lecture 3

Civil Law can be divided into:

- Civil Liability

- Contract Law: formation and conclusion of contracts

- Property Law

Contract and **Property** are the core concepts of the Natural Law, in contrast to the Positive Law.

- **Natural Law**: based on the *rationality principle*, meaning that there can not be any contraddiction in the law. Often associated with the divine law first (natural, so it can not be changed by the humans), after the enlightment it was based on the words of notorious and influencial personalities, like the Greek philosophers.

- **Positive Law**: from Latin, meaning **placed**, **created**.

The property in law is also seen as a way to asses social stature, and it is linked to the concept of liberty. There are two types of property, in contrast:

- **Real property** (in rem right): property of things

- **Intellectual property**: property of ideas

4   Lecture 4

Legal issues in Computer Science:

- Copyright

- Patents

- Trademarks

The copyright was created to protect the publisher. After the invention of the press in the 15th century, it became relevant the importance of protecting the producers and the writers of books.

In the UK the **stationers' company**, a gild established by Queen Mary in 1557, composed of booksellers, publisers and more, worked as a monopoly; this ended with the enlightment, when scholars asked for the access to culture for all people, giving the momentum for the creation of a public domain.

The first statute dealing with the authors' exclusive right came in 1710, with the **Statute of Anne**. The *Statute of Anne* (1710) and the *US Constitution* both state that the exclusive right is given in order to make society better, by disseminating knowledge. But how to encourage learning giving to the author the power to exclude you from learning? The solution is to limit this control, by controlling the duration and the extensiveness of the right.

First of all the right extends only to original works, which can not be an *"idea, procedure, process, system, method of operation, concept, principle, or discovery"* ; they have to be *"original works of authorship fixed in any tangible*

medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device". Also, limitation on exclusive rights to allow fair use, where added. For example in the USC the factors to be considered in determining the fair use where:

- the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;

- the nature of the copyrighted work;

- the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and

- the effect of the use upon the potential market for or value of the copyrighted work.

The First Sale Doctrine was also implemented, to entitle the owner of a particular copy the possibility *"without the authority of the copyright owner, to sell or otherwise dispose of the possession of that copy"*.

The duration of the copyright changed over the years:

- Statute of Anne − 1710: 14 years (+ 14 if author still alive)

- U.S. Copyright Act 1790: 14 years

- Berne Convention art. 7: author's life + 50 years

- Council Directive 93/98/EEC of 29 October 1993 harmonizing the term of protection of copyright and certain related rights: author's life + 70 years

- Sonny Bono Copyright Extension Act of 1998: author's life + 70 years

5 Lecture 5

A copyright has value where it is accepted. What if I have a copyright in a nation and I cross the border? Obviously if there are not regulations, the copyright is useless. That is why an important point was the *globalization* of copyright law. The agreements between different nations were formalised during the **Berne Convention** (Berne Convention for the Protection of Literary and Artistic Works), in 1886. US joined it only in 1988, when they started to become exporters of patents.

What about the copyright on code? At first it was possible only on source code, because the compiled code needed the aid of machines to be understood.

In Berne Convention there is no:

- authorship

- originability

EU specific law for database copyright is called *Sui Generis Right*.

6 Lecture 6

The US Constitution Copyright Law embrace the possibility for the owner of a copy of copyrighted material to make everything with it: reselling, destroying, etc. By selling it the former owner loose the distribution right to the new owner. This is a concept difficult to translate to the digital intellectual works, which were not taken into account, as well as internet. The council made a directive in 1991 to fix this.

The Gutenberg Project was created to make a public domain for books. The concept of books without legal restraints caused some issues, like the copy of Alice in Wonderland, part of the public domain, made by Adobe that contained a copyright. It was obviuosly motivated by the fact that the original copy was packaged with something new, but this fact caused a significant buzz around the legal regulations over the public domain.

Domain that was also affected by the change of laws over time: extension of term could make a material not part of the public domain, retroactively. Internet publisher Eric Eldred sued the US Congress after one of these changes because Sonny Bono's law was in his opinion uncostitusional. The Supreme Court decided to review the case but Elder lost.

Also the European Union messed up recently with the copyright law, with the update on legislation about phonorecords in 2015, which could result in the loss of old recordings that were before part of the public domain. The change was made because the previous time limit was of 50 years, and 1965 matches with the boom of pop music (beetles, etc), works that still generate profit to authors and publishers.

Proposal were made to change the law, like one to extend the copyright with a tax, so that only valuable and profitable works are stopped from going public.

7 Lecture 7

Some development in the Computer Industry:

- 1952: First computer sold by IBM

- 1955: IBM creates the SHARE user group for its 704 Data Processing System. Members of the SHARE user group were competitor in their respective markets (for instance Lockeed and Boing)

- 1957: FORTRAN programming language

- 1959: Cobol programming language

- 60s:

  - First operating system: FORTRAN Monitor System
  - Multiprogramming
  - Time-sharing

- 1969-1973: Unix Operating System at the Bell Laboratories (AT&T) and development of the C programming language, a high level programming language with the use of libraries, making it the first (easily) portable programming language

- 1974: Publication of D. M. Ritchie & K. Thompson, The UNIX Time-Sharing System

The Unix copyright belongs to AT&T, owner of the Bell Laboratories: in 1950 Consent Decree between AT&T and the U.S. Department of Justice stated that AT&T could not enter the computer and software industries. Unix was "Freely" distributed to Universities and research center in source code format. But in 1979, with Unix Version 7, AT&T adopts a new license agreement that prohibits the use of the Unix source code as a teaching material:

- Need to keep to source code secret for the application of the trade secret rules

- Tanenbaum (Professor of Operating System, Vrije Universiteit of Amsterdam) writes MINIX as a teaching tool, without AT&T code.

In 1982 a New Consent Decree with the Department Of Justice: AT&T looses its monopoly over telephone services and can now enter the computer industry: the Unix System Laboratories were created. This is the beginning of the commercial Unix.

There were different flavours of Unix: AIX, SOLARIS, ULTRIX, etc. The amount of versions created pushed for the creation of a standard of Unix OSs, POSIX.

Back to Copyright Law and Software, before 1976:

- requires the protected work to be registered at the U.S. Copyright Office: *"The copies deposited for registration [must] consist of or include reproductions in a language intelligible to human beings."*, so no machine code allowed.

- notion of copy: *"A copy requires the possibility to see and read with the naked eye"*

After the 1976 U.S. Copyright Act changes were made to include software: *"'Literary works' are works, other than audiovisual works, expressed in words, numbers, or other verbal or numerical symbols or indicia, regardless of the nature of the material objects, such as books, periodicals, manuscripts, phonorecords, film, tapes, disks, or cards, in which they are embodied."*. The Congress created the Commission on New Technological Uses of Copyrighted Works ("CONTU"), for further amendments to the Copyright Law.

At first there were some missing explanations in the use of the new provisions:

- is a user allowed to make changes to software to fit personal needs?

- ????????????????????????????????? (missing)

- is a user allowed to make backup copies in case of failure?

- can the program be translated?

The CONTU and Computer Software Copyright Act of 1980 added the possibility of a backup copy and the adaptability of the software for personal needs.

Europe came late, because the software was not used as much as in the US. There were some national regulations, but the first European directive on the legal protection of computer programs came in 1991. This directive was a copy of the US one but it included the possibility of **decompilation**:

> "The authorization of the rightholder shall not be required where reproduction of the code and translation of its form within the meaning of Article 4 (a) and (b) are indispensable to obtain the information necessary to achieve the interoperability of an independently created computer program with other programs"

This allowed european hackers to decompile data from MS word's documents to make their content usable in other programs.

Tech development and legal development shaped the software industry. There was also social pressure because PCs were becoming cheaper. For example, the **Altair 8080**, a basic PC that costed less than 500 bucks. Coding requested machine code, but two students, Bill Gates and Paul Allen, made a BASIC interpreter which was used freely. Gates wrote an open letter to hobbyists, who used his interpreter without having paid for it, complaining about the piracy on software:

> *As the majority of hobbyists must be aware, most of you steal your software. Hardware must be paid for, but software is something to share. Who cares if the people who worked on it get paid?*
> *...*
> *One thing you do do is prevent good software from being written.*

8 Lecture 8

The end of the 70s marks the beginning of the **open source** movement, in response to the computer industry and the changes in law. After 1976, the communications via the soon to become internet included copyright messages (?give things away by prohibiting to make private copies?).
Founder of the open source movement was Richard Stallman. He was a programmer at the MIT AI laboratory who made a driver for the lab's printer and wanted to port it to the new laser one; he asked XEROX, producer of the printer, the code to include his changes, but they prohibited it. Also the guy who made the driver could not give it to him because of a non-disclosure agreement. Stallman was frightened by the idea of people in the research community doing this. So he left his job (1984) in order to be the only owner of his work: because of the *work of hire*, the code written by an employee belongs to the employer. In the EU there is a moral right included: an employee has the right to be recognized as the author, while the economic right belongs to the employer (example: in the US it is possible to be a ghostwriter, in the EU it is not).
Stallman started working in 1985 on a new OS: the GNU project (GNU's Not Unix). The legal structure to preserve it was defined in 1989 with the GNU General Public License. The copyright of every addiction to it goes public, it is owned by the GNU project. The GPL was created when EMACS, the editor made by Stallman, received criticism because the copyright forced every change to be accepted by the original creator (central hub). So changes were made

to this, that later contributed to the GPL: the license used copyright to avoid privatizing. License is a defense: writing that it is a contract causes problems in the US legislation, while writing that it is not causes problems in the EU legislation.

9   Lecture 9

There is a variance of the GNU GPL, the Library GPL: the LGPL allows the work to be linked with (used by) a program outside of the GPL, regardless of whether it is free software or proprietary software. Without it the linking of the GNU library would cause a copyright infringment.

GNU was still not enough to make an OS, a kernel was needed. In 1991 Linus Torvald released the first version of the Linux kernel, without any license (just a few words, you kept the copyright after an edit). In 1992 he adopted the GPL for it. The joining of GNU and Linux made possible the OS "GNU/Linux".

Meanwhile in Berkley, San Francisco, another OS was being created. In 1978 Bill Joy created **vi** editor and a PASCAL compiler and packaged together in the BSD. In 1979 DARPA asked the Berkley University to develop a TCP/IP (internet protocol) for UNIX. This work was made under the BSD license; it allowed to take the program, change it and make it private or change even the license. In 1991 an UNIX-like OS was made from BSD, the same period of GNU/Linux. The following year BSDI (private company from BSD) started selling commercial version of BSD code, calling it UNIX OS; they were obviously sued by AT&T. When later AT&T sold all the UNIX's rights to Novell, GNU/Linux was already dominating the scene; so Novell and BSD reached an agreement for the publication.

10   Lecture 10

Software as a service:

- **Cloud computing:** someone offers computing on different machines
- **Financial model:**
    - subscription
    - freemium
    - free

Example: gmail is halfway between free and freemium: you have a maximum space, to have more you must pay.

These services have to guarantee:

- availability, collaboration, maintenance
- data protection
- reliability (of the network): depends on the end of the service provider and of the user

There are problems with SaaS (Software as a Service) substitutes:

- online/offline = if you are offline you can't access your data

- endpoint security

- control of data, data integration = where is my data? The location of it (country) can cause legal problems

- control of software version/features = if provider changes a functionality you can't refuse and keep using it

- exit cost/lock-in = complicated to transfer online data offline

- SLA obscurity = maybe changes can be made

There is also a specific GPL version, AGPL, with the addition of considering "distribution" running a program on a server; this is included to avoid gaining from GPL programs using the SaaS.

"Free" services are tricky. They may be aggregations of other services, dynamically chosen at run-time; is this trustable? Also user data stored can be used; google terms of services include a licence on whatever is uploaded on their services.

11   Lecture 11

What is the meaning of data? There are different answers to this question. A possible one is data are the attributes collected in digital format and that are not ambiguous.

$$DATA \rightarrow INFORMATION \rightarrow KNOWLEDGE \rightarrow ACTION$$

Open data is an argument that has been going on for many years, as the 1898 Cesare Battisti's quote states.

Data can be divided in:

- **Personal data**: informations about individuals.

- **Small data**: data easy to understand.

- **Big data**: huge amount of data $\rightarrow$ a lot of problems related to technology and management of it.

- **Open data**: is just a policy (each of the previous data types can be open).

Personal data is usually private. If someone asks you to give your data to him, you will refuse to do it or ask for money. But you give them to companies like Facebook and Google because they offer you something for it, something you would pay for otherwise.

Open data means freely accessed, used, modified and shared for any purpose by anyone. Anything with NON COMMERCIAL restrictions is NOT open data.

Creative commons, an American non-profit organization devoted to expanding the range of creative works available for others to build upon legally and to share, uses different layers to add copyright:

- PUBLIC DOMAIN

    - PERMISSION = use, modify, share
    - RESTRICTIONS = no one
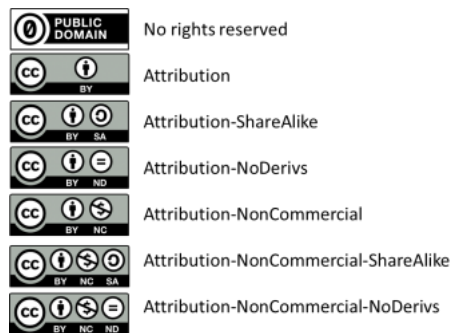
- ATTRIBUTION

    - PERMISSION = use, modify, share
    - RESTRICTIONS = a reference to the original has to remain

    This creates problems when joining more data with different attributions.

- SHARE-ALIKE

    - PERMISSION = use, modify, share
    - RESTRICTIONS = reference + give back to the community + can not be close (private).

There also the 'Non Commercial', to prohibit profit, and 'No derivs', to prohibit the creation of any type of derivative work.

PUBLIC DOMAIN	No rights reserved
cc BY	Attribution
cc BY SA	Attribution-ShareAlike
cc BY ND	Attribution-NoDerivs
cc BY NC	Attribution-NonCommercial
cc BY NC SA	Attribution-NonCommercial-ShareAlike
cc BY NC ND	Attribution-NonCommercial-NoDerivs

Open data needs interoperability, (?), and documentation, important to understand and improve data.

Example of *OpenStreetMap*: community of mappers, provides maps and geographic datas that are open. They use the ODbL licence, that require the derivative data to be open and the attribution when using it (maps, graphs, etc). A problem they have is related to the fact that people collect data without thinking about privacy (private swimming pools added in the maps).

12   Lecture 12

A **patent** is a legal tool granting exclusive rights to an inventor. It is aimed at promoting innovation and inventions, it has territorial range (granted by a state), a time range (limited number of years) and requires disclosure of the invention. Patents are normally managed by Patent Offices, the main ones being the American (Title 35), the European and the Japanese; there are international agreements like the TRIPS (Agreement on Trade Related Aspects of Intellectual Property Rights). To enforce the patents, yearly payment of a fee are required.

The European Patent Convention is not entirely overlapping with the European Union; for example, Switzerland, which is not a member, is included in the EPC. It is defined by articles, the most relevant being:

- Article 52 (**Patentable inventions**): states that patents shall be granted for "any inventions which are susceptible of industrial application, which are new and which involve an inventive step", while they shall not be in case of:

  - discoveries, scientific theories and mathematical methods;
  - aesthetic creations;
  - schemes, rules and methods for performing mental acts, playing with games or doing business, and **programs for computers**;
  - presentations of information.

- Article 54 (**Novelty**): an invention is considered new "if it does not form part of the state of the art", which includes everything made available to the public before the filing of that patent.

- Article 56 (**Inventive step**): an invention is considered as involving an inventive step "if, having regard to the state of the art, it is not obvious to a person skilled in the art".

- Article 57 (**Industrial application**): an invention is considered as susceptible of industrial application "if it can be made or used in any kind of industry, including agriculture."

- Article 83 (**Disclosure of the invention**): states that the patent must describe the invention in a clear and complete way, "for it to be carried out by a person skilled in the art" (making it replicable).

The US Patent Law is defined, in sections (§), in Title 35 of the U.S. code:

- §101 (**Inventions patentable**): a patent is obtainable by "whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof". Excluded from patent protection are laws of nature, physical phenomena and abstract ideas.

- §102 (**Novelty**): conditions for patentability are novelty and loss of right to patent.

- §103 (**Originality, non-obviousness**): conditions for patentability is non-obvious subject matter.

- §112 (**Specification**): the specification (description) of the invention (included manner and process of making it) have to be written in "full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention".

So what about software? In the EU regulation it is precisely stated that computer programs can not be patentable, while the US regulation is more open to it. Technically, software may be included in laws of nature, physical phenomena and abstract ideas, but this vision changed over time:

- '70s: there was no protection if an invention used a calculation made by a computer;

- '80s: supreme court said some computerized inventions are patentable;

- '90s: the federal circuit says almost all software is patentable.

The USPTO maintained the position that software was in effect a mathematical algorithm, and therefore not patentable, into the 1980s. This position of the USPTO was challenged with a landmark 1981 Supreme Court case, Diamond v. Diehr. The court essentially ruled that while algorithms themselves could not be patented, devices that utilized them could. In 1996 the USPTO issued Final Computer Related Examination Guidelines, paving the way for software patents.
Also Europe started to ease the interpretation of Article 52; this was also made because it was not in Patent Offices' interest to reject patents, which they get paid for. The European Patent Office started to accept patents for any invention making a non-obvious technical contribution, also in computer programs. (?????) However this creates big problems when patents are given away easily: companies started to ask more and more of them, to create a portfolio of patents. For example, IBM is applying for over 3000 patents each year. This portfolio can be helpful to exchange patents with other firms in legal disputes, but can also be used by the so-called Patent Trolls (individuals or companies that collect a lot of patents) to legally blackmail companies that use the most simple feature; fight them in the court of justice could be detrimental, especially in the US where litigations work with a public jury so the outcome is unpredictable (Microsoft example for embedding programs). Settlement is always cheaper.

To sum it up, the problems in the software area are caused by the abstractness of the level of description (if it is not clearly explained it should not be patentable) and by the market (patent trolls). Software patents can be used to stop free software development:

- no reverse engineering allowed;

- long protection for software industry;

- no idea/expression dichotomy.

13 Lecture 13

The technical definition of File Sharing is quite broad and can be supported in a number of ways. We focus on allowing public access to copyrighted material in digital form, without consent of the copyright holder (the music case). There are 2 type of actors:

- The end-users: "direct" infringement

– accessing material

 – sharing material with others

- The mediator: secondary infringement

  – can be P2P provider, index, or other service

  – promotion of infringing activities

  – aiding or contributing to activities

  – lack of supervision of activities (vicarious liability)

  – less clarified in law

Is it really bad to share? Downloading certainly infringes copyright and copyright infringement is a crime, but it usually does not affect sales (victimless crime) and, if it does, it affects more the publisher than the author. Also copyright is no longer an "ecouragment for learning" when the file is not made available. Anyway it doesn't make sense to punish the downloaders.

The most known debate on file sharing is the one on music; it has always been "shared" and it is a social experience and has a relevant role in human culture. The music industry started with the possibility to create physical artifacts containing music, so live performances moved from the only way to experience music to promotion for artifact sale. Traditionally, Music Industry focuses on production, manufacturing, promoting, distributing and selling artifacts. The unauthorized use of it has a long history: in the analog age it was possible but long and slow, plus it had small impact on business, while now it is easier and faster thanks to the digital format, the compression algorithms (like mp3) and the online sharing. Internet allowed also services like Napster to make sharing even easier. Napster was a centralised index and discovery service for P2P (Peer to Peer) distribution, commonly identified with the start of "illegal file sharing", with up to 70M registered users (2.5M simultaneous). It created supposed damage for some artists (pre-release distribution of songs of Metallica and Madonna) and supposed benefit for others (Radiohead). It lost a trial in 2001 and declined, today is (part of) Rhapsody, a streaming music company. As a tool to fight it, the DMCA (Digital Millennium Copyright Act) of 1998 was implemented, targeting the technology for circumventing DRM (Digital Rights Management) that are protecting copyrighted material plus just the information to do so (not explicitly targeting linking to material infringing copyright). After Napster P2P started to use fully decentralised file sharing networks (Gnutella, Torrent), where the communication is encrypted and the material is replicated and distributed to a number of peers. Often an artifact is splitted in smaller parts. To find this material sites like Pirate Bay were founded: web sites providing a directory of links to (torrent) files provided by third parties. Pirate Bay had several legal issues:

- founders sentenced 4m to 10m plus 5 million euros;

- blocked by ISP in several countries;

- blocked by applications and services (Facebook).

Its business model was based on donations and advertisement.

In Italy the Urbano decree in 2004 had broader aim (funds for cinema and sport) but also "to contrast the unauthorised telematic distribution of audiovisual material". Then the Italian law on authors' right was modified (diritto d'autore − law 633, issued in 1941) and originated a debate on terminology ("lucro" versus "profitto"). It stated strong punishment for "non personal use" (1y to 3y, 2.5K euros to 15K euros).

Another example is MegaUpload, a File sharing website that allowed free upload and basic download services, with the possibility of premium accounts for faster download. It was closed in 2012, after a controversial international case on founder extradition to US (still not closed). Claims are that the service costed more than $500M to copyright owners and that it produced more than $175M in illegal revenues. The main issue was that, to reduce the amount of data stored in their database, MegaUpload used de-deduplication: with this technology, a file storage system scans files as they come in and if they are recognized as something that's been uploaded previously, the system will not store the new files, but instead it will reference back to the version already on the servers. In addition to being a great space-saver, this can be an easy way to wipe out all versions of a copyright-infringing file in one swoop. But when MegaUpload would receive a takedown request for copyright violations, it would only disable the one reported link, instead of every link associated with that file, and the file itself.

Spotify is a music steaming service , with a focus on track (not album) and social net integration. It was one of the first service to sign agreements with major record labels. Their content is protected with DRM and it used over time various kind of limitations. Their business model focuses on advertisement for free use and premium accounts. Several artists criticise Spotify for not producing enough revenues for them. The relationship between an artist and a music label rely on contracts, licensing, copyright transfer, share of revenues, ... The relationship between a music label and a distributor instead can include share of revenues, upfront payment, exchange of advertisement, purchase of equity, ... Some part of cash flow from distributor contributes to artist's revenues, some is not considered, some can have a negative impact.

14   Lecture 14

Difference in law between privacy and secrecy.