The Myth of Odin's Eye: Privacy vs. Knowledge

Paolo Guarda*

I. Introduction

Knowledge has always represented a sort of human unrealizable dream and omniscience, that means the complete and unlimited knowledge, is considered by most of the cultures as a divine characteristic. To know means to have the power: in human history – and in recent philosophical reflections – it is a full stop!

But what are we willing to give to get access to knowledge? Myths and stories on knowledge research tell of long trips and of supreme sacrifices.

In the portrayals of Odin, the father of the gods in the Scandinavian tradition, he is often represented with a single eye[1]. One of the several changing Nordic legends tells that, in order to satisfy his thirst for knowledge, the god sacrificed his eye to Mimir, the Mimirsbrønd's keeper, the magic fountain of wisdom placed on Yggrdrasil's roots[2]. Since then, the divine eye is situated in the icy water of the fountain, as the price paid to acquire the eyes of the sage and to discern the essence of things, behind the appearance.

This is the price paid by Odin. But what are we willing to pay? Or rather: are we really aware of the cost of our access to knowledge, particularly in the digital context and despite the outward gratuitousness of the provided service? I will try to give an answer to these questions in the following pages.

In the first part I will focus on the issues linked to utilization of Digital Rights Management systems (DRM), suited to regulate the fruition of intellectual works protected by intellectual property rules, with specific reference to users' privacy in its different dimensions; then, always from a privacy oriented point of view, I will outline the problematic aspects that open access to knowledge presents; I will analyze a paradigmatic example of free of charge service that poses monstrous risks from the point of view of users profiling activity. At the end, in order to pull the strings of the matter together and to take back the cues offered by Odin's saga, I will develop some final reflections on the relationship between privacy and knowledge in the digital context.

---

* Ph.D. in Comparative Private Law, University of Trento

[1] For closer examination on Nordic saga, see **V. GRØNBECH**, *Miti e leggende del nord*, Torino, Einaudi, 1996 (Italian translation of the Danish edition *Nordiska Mita og Sagn*, Copenhagen, 1965).

[2] See *Völuspá* (*Prophecy of the Völva*), the first and the most famous of *Edda poems.*

II. Intellectual property, Digital Rights Management (DRM) and Privacy

The advent of digital technologies gave rise to two phenomena: on the one hand, we register a heightening of the rigid and concentrated control (i.e. the DRM systems), based on commercial rules; on the other hand, we have the coming of new ways to fruitions focused on open access rules.

I will concentrate on the hidden "price" that the user pays to get access to digital intellectual works, whether in the platforms that incorporate values and rules of the intellectual property, or in the architectures inspired by openness (I will analyse this issue in the next paragraph).

From a purely technological point of view, the information production model based on the information closure and the rigid and concentrated control of it is implemented through DRM systems[3]. This term singles out the most advanced anti-access and anti-copy protection system on the market. The heart of every DRM is constituted by two modules. On the one hand, the so called "content-module", that contains the digitalized data (text or audio files) "secured" through a cryptography process and ready to be distributed; before it happens, the name of the author, the copyright owner, the creation date, the title, the format, the dimension, and other technical information in order to identify the file (*i.e.* the ISBN) are incorporated into the content. On the other hand, the "licensing-module" generates the digital license that guarantees to the final user the access to the contents in the light of the usage rights of business rules. These determine the "what", that means the precise piece of content to be used, and the "when", that means to link particular features to every right: the type of user authorized to assert the "right", the extension of every right (duration or numbers of permitted utilizations), and the price in order to assert the right.

Let's see an example so as to understand how this mechanism works[4]. Suppose I wish to get access to the new publications of an on-line law journal. I registered myself giving my profile details to the portal that contains the journal: my data are saved in a database assigned to store information regarding the users' identity inside the licensing-module. At the same moment, the system generates the usage rules that sets the type of utilization, the cost, and the

---

[3] In-depth analysis, see **R. CASO**, *Digital Rights Management. Il commercio delle informazioni digitali tra contratto e diritto d'autore*, Padova, Cedam, 2004 (digital reprint, Trento, 2006, available at the Web-site: <http://eprints.biblio.unitn.it/archive/00001336/>); **A. PALMIERI**, "DRM e disciplina europea della protezione dei dati personali", in **R. CASO** (eds), *Digital rights management: problemi teorici e prospettive applicative: Proceedings of the Conference Trento Faculty of Law 21-22 March 2007*, Trento, Alcione, 2007, pp. 197 ff..

[4] The example, modified in the lights of the considerations I am developing, is partially taken from **P. GANLEY**, "Access to the Individual: Digital Rights Management Systems and the Intersection of Informational and Decisional Privacy Interests", *International Journal of Law and Information Technology*, 2002, p. 241.

specific category of authorized users; then it puts all this information into the licensing-module. Whenever I desire to read the latest publication, the DocuReader device contacts the content-module and a decrypted copy is sent to my computer. Now I am able to display what I was looking for, but every future access attempts will be denied. In order to get other accesses, I will have to take out a subscription that will enable me to get more displays, or greater usage margins. This is a simple example of how a DRM system works.

The DRM system also presents functionalities that have a direct impact on user privacy[5]. It contains at least one of the following basic functions[6]:

a. content access control;
b. content usage control;
c. content identification, content owners, and general condition of usage;
d. authentication of identification data.

Moreover, for its own particular peculiarities, a DRM system is able to monitor the content fruition and, in case it has been set in that way, it can "sanction" the behaviours that are not compliant with its rules, for instance disconnecting the access.

Looking through what a DRM can do, a privacy scholar immediately thinks of "profiling"[7]. Personal data of Internet users, consumers of intellectual works, and, more generally, their commercial interests symbolise a real treasure for people working on the Net. The possibility of monitoring the user activities does not only represent the way to control-manage his digital content fruition in order to impose – and when it is supposed to be necessary to sanction – the behaviours allowed according to the license that regulates the usage rights; but it becomes also a way to commodify the profile itself: this profile can be used by the profiler to adopt a more effective marketing strategy (maybe through some advertising banners answering user interests) or becomes itself an exchanging good, when the information is sold to a third party.

This kind of considerations makes us easily understand the real impact that these technologies have with respect to users privacy. Now I need to clarify what I mean by the term "privacy". When you deal with the description of this concept, even more if you place yourself by the point of view of the relation between law and technology, you tell how from a

---

[5] In the digital age law, privacy and copyright often know clashing points. See, for example, the case *Peppermint* annotated in **R. CASO**, "Il conflitto tra copyright e privacy nelle reti peer to peer: in margine al caso Peppermint. Profili di diritto comparator", *Dir. dell'Internet*, 2007, p. 471 (available at the Web-site: <http://www.jus.unitn.it/users/caso/DRM /Libro/peppermint/home.asp>).

[6] See **R. CASO**, *Digital Rights Management*, *o.c.*, p. 100.

[7] On the profiling activity, see **B. CUSTER**, *The Power of Knowledge. Ethical, Legal, and Technological Aspects of Data Mining and Group Profiling Epidemiology*, Nijmegen, The Netherlands, 2004.

privacy intended as the "right to be let alone", originating from the paper of great-renown of Warren and Brandeis[8], we arrived to a meaning much more linked to the idea of control of personal data, thanks to digital technologies diffusion. And actually, when we say "privacy" referring to the provisions that regulate it, we immediately hasten to remember that we are talking of personal data, of data control, of information to data subject, etc.[9].

But now I would like to quote a more extensive concept of privacy. This kind of description of the concept allows us to better understand the real importance of privacy in the Net, its actual role, in a context in which more and more information, and the commodification of it, comes into the limelight. People personality itself is taken to small pieces that are personal data. These pieces, once you have reassembled them, sketch out the profile of a person in the digital context, his virtual identity. The control on these pieces of identity has more dangerous repercussions than we can imagine if we take into account a narrow interpretation of privacy, since it involves the life fundamental aspects of the person to whom the data are related.

After all these expectations, I will describe which are these theories I was referring to. The consumption of intellectual works, intended as the fruition of work of the intellects and of other information, is directly linked to three fundamental dimensions of the concept of privacy: "spatial", "informational", and "decisional"[10].

The first dimension, the "spatial" one, concerns the physical space, and in particular the size of the area in which the solitude of a person is protected from external invasion: this kind of privacy corresponds to a particular idea that sociologists connect to the relation between the private and the public space. The dimension I am talking about is related to freedom from external nuances or, more generally, from every kind of source of disturbances coming from the outside. In the matter I am dealing with now, spatial privacy involves those spaces in which a person is at liberty to behave in such a way that would be considered

---

[8] **D. WARREN, L. D. BRANDEIS**, "The Right to Privacy", *Harvard L. Rev.* , 1890, vol. 4, p. 193.

[9] For closer examination on privacy regulation, see *ex plurimis* **L. A. BYGRAVE**, *Data Protection Law. Approaching Its Rationale, Logic and Limits*, The Hague – London - New York, 2002; **P. GUARDA**, "Data Protection, Information Privacy, and Security Measures: an Essay on the European and the Italian Legal Frameworks", *Ciberspazio e dir.*, 2008, p. 65.

[10] See **J. KANG**, "Information Privacy in Cyberspace Transactions", *Stan. L. Rev.*, 1998, vol. 50, p. 1193, at 1202-1205; **P. GANLEY**, "Access to the Individual", *o.c.*, at 251 ff.. Professor Caso draws from an American commentator a double dimensions of the concept of privacy (informational and spatial): **R. CASO**, *Digital Rights Management, o.c.*, p. 103 ff.; **J. COHEN**, "DRM and Privacy", *Berkeley Tech. L.J.*, 2003, vol. 18, p. 575. More generally, on the implications of DRM systems with respect to privacy and freedom of speech, see **I. KERR, J. BAILEY**, "The Implications of Digital Rights Management for Privacy and Freedom of Expression", *Info., Comm. & Ethics in Society*, 2004, vol. 2, p. 87.

aberrant by the point of view of the dominant social norms or, more plainly, not normally used in public.

The second dimension, "informational" privacy, regards the flow of personal information. More accurately, it relates to the individual control with respect to processing of his personal data. Therefore, it is directly linked to that conceptualization of privacy focused on the control and the regulation of personal data treatment. From the point of view that focuses on the impact of DRM systems, processed data are information regarding the intellectual consumption, acquired by means of the functionalities we saw typifying the technological protection measures.

The third, and certainly most important, dimension of privacy is the "decisional" one: it concerns the choice, the freedom that must be recognized to every person in order to be able to take a decision without any kind of external conditioning. I wish to focus my attention on this dimension.

The "decisional" privacy involves the essence itself of human being. The free will, the freedom of self-determination, ultimately the freedom to be a man. There is not law, if there is not free will, if there is not the possibility to choose, even if to be wrong. Therefore, the violation of this dimension straightforwardly affects the "capacity" to be a man. The monitoring activity, the awareness of being spied on, the consciousness that the context around us is unceasingly changing in the light of the profile that other people are designing to us, modify person's behaviour. The famous Panopticon of Bentham was based on this idea[11]: the radiocentric form of the building and the appropriate architectural and technological contrivances, that enabled a single warden to watch all the prisoners at every moment, without the possibility for these to establish if they were actually controlled or not (this point reminds us dreadfully the monitoring on the Net), gave the inmates the sentiment of an invisible omniscience, influencing their behaviour and persuading them to not violate the rules.

Therefore, the perception of being watched influences, and will influence, our intellectual consumption, in such a way that we probably are not yet able to clearly determine.

Let's go back to the example I was using to better understand the potential consequences of DRM systems with respect to privacy.

Before the published content in the on-line law journal has been encrypted, a metadata, called "Digital Object Identifier" (DOI), is inserted into every section of all the

[11] **J. BENTHAM**, *The Panopticon Writings*, edited by **M. BOZOVIC**, London, Verso, 1995. Bentham himself described the Panopticon as "a new mode of obtaining power of mind over mind, in a quantity hitherto without example".

articles. The granularity level of this DOI (every article, every section, or every paragraph), that are unique for every type of content, is decided by the editor. The outcome of this process is the following: all crypted data in the content-module are directly linked to some "pieces" of identifier metadata, that will always remain in the content when this is subsequently distributed. Then, every time I wish to modify the content of the document I have asked to the system (for instance to open, display, print a specific section of the publication), the "ContentControl" application sends an information package to the licensing-module, inclusive of the DOI and of the details of the action carried on; the package will be stored in the "logging program". Simply accessing to the licensing-module, the editor can compile aggregate statistics of the utilization of the contents or display the profile of a specific user combining the information stored in the logging program with data saved in the database suited to collate users personal information.

III. "Free of charge" on-line information access services and privacy

I will deal in this section with the other moon's face: the Open Access world or, more in general, the open – and seemingly free of charge – access to the fruition of on-line information[12]. I will hold the "Google" search engine as a paradigmatic example, also in its versions of "Google Scholar" and "Google Book Search"[13]. I made this choice since it represents a model, among the most invasive ones in the Net, of personal data gathering and of users profiling. Obviously, many of the issues I am going to describe can be met in other portals that provide access to scientific knowledge, especially when they are provided by an authentication system.

From a purely economic point of view, this scenario could seem as an ideal world, that is the world of the free of charge and completely open access to knowledge; unfortunately, it is actually hiding a dark side. Notwithstanding the outward idyll, also in this context we are paying, often completely unawares, a very expensive price.

Google is one of the most evident examples of personal data collection and, more generally, of information related to users. Every day millions of people use this search engine, giving to it information regarding their interests, needs, desires, fears, etc..

---

[12] For a closer examination on Open Access issues, see **R. CASO**, *Ricerca scientifica pubblica, trasferimento tecnologico e proprietà intellettuale*, Bologna, Il Mulino, 2005, p. 312; **ID.**, "Proprietà intellettuale, tecnologie digitali ed accesso alla conoscenza scientifica: Digital Rights Management vs. Open Access", available at the Web-site: <http://eprints.biblio.unitn.it/archive/00001407/>.

[13] I took as a point of reference of this part: **O. TENE**, "What Google Knows: Privacy and Internet Search Engines", (October 2007), available on the Web-site: <http://ssrn.com>.

I want to begin with a starting remark: Google is recording every query and is connecting them to a given Internet Protocol address (IP). This sentence probably frightens, astounds, immediately makes us calling to mind all the queries we have done, fearful of "what did we ask to Google?".

Let's proceed in order. From this point of view, privacy concerns personal data, the information that European Directive 95/46/EC ("on the protection of individuals with regard to the processing of personal data and on the free movement of such data") at article 2, lett. a, defines as relating to "an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity". Hence, in reverse order, information that cannot be connected with a specific person is not a privacy issue. Then, the question lies in the possibility to establish this connection.

In the following part I will explain how the log files relating to our queries can be associated to our person (*rectius*, identity).

First of all, it happens through the IP collection. These are numbers that identify unequivocally the devices connected to a single network; IPs can be statics, that means fixed, or dynamics, that means assigned each time by the Internet Service Provider (ISP) to his subscribers depending on necessity. The IP address, that can be banally considered the equivalent of a street address or of a telephone number, is regarded a personal data according to art. 2 Directive 95/46/EC (and art. 4, Italian Data Protection Code)[14]. Even if Google is not able to match the IP address linked to the log file that contains the queries, to a certain user, the fact that the ISP has this kind of data and that the public authority could force him to communicate data relating to the subscribers makes the log files close to personal data.

In the second place, to overtake the difficulty of profiling users caused by dynamic IP addresses, an open access service portal can use "cookies": these mark the user browser (the application used for surfing in the Net) with some "unique identifying numbers"[15]. Actually these files, that are supposed to facilitate users in remembering their login data and queries at every connection to the same Web page, permit the search engine to recognize the user as a

[14] See Opinion of Article 29 Data Protection Working Party 2/2002 on the use of unique identifiers in telecommunication terminal equipments: the example of IPv6, available at the Web site: <www.ec.europe.eu>.
[15] In-depth analysis, see **J.J. THILL**, "The Cookie Monster: From Sesame Street to Your Hard Drive", *S.C. L. Rev*., 2001, vol. 52, p. 921.

"recurrent visitor" of the site and to collate the history list of his queries, even if he connects to the Net using different IPs[16].

Cookies in themselves cannot be considered personal data, since they identify a particular browser (*rectius*, a computer), rather than a user. The cookie can be imagined as a sort of label, as "7583co3948le24", without any visible repercussion on privacy, but as it were stick on a personal data box of an anonymous person. If you could match a given cookie, linked to a log file, with the queries on the search engine, to a specific person, then again you would relapse into the personal data uncomfortable – at least for the profilers – hat. But also this obstacle can be bypassed: let's see how. As everybody knows, in addition to the search engine, Google provides several additional services (from the e-mail account to the investment portfolio) that are subject to registration through credentials, as the real name or the e-mail address (the most obvious example is Gmail). At this point, Bob's your uncle: the user logs into the Internet, loads the search engine portal, in the meanwhile he introduces a query. Google uses the same cookie to identify the list of queries and the e-mail account: the cookie anonymity is once and for all lost, the missing name on the users' interests log files box has now a name and a surname!

In the end, a final possibility to collate personal data is left over: this time the only guilty party is the user himself, apart from every kind of tactics he could adopt to reduce the use and the identification of his cookies. We all have tried in our life to look for our name on Google or, in our specific case, to search for our essays on a scientific database, maybe repeatedly in the same year[17]. These queries are called "ego searches" or "vanity searches". They allow the profiler to link the information relating to users interests to a certain identity.

Now let's put another question to ourselves: where do the log files containing all these information on our queries go to?

In the Google privacy policy we find innocently stated: "We use cookies to improve the quality of our service by storing user preferences and tracking user trends, such as how people search. Most browsers are initially set up to accept cookies, but you can reset your browser to refuse all cookies or to indicate when a cookie is being sent. However, some

[16] Google has lately announced the intention to reduce the cookies retention period, initially stated until 2008, to "only" two years starting from the latest user search in its search engine: see "Cookies Expiring Sooner to Improbe Privacy", Official Google Blog, 16 July 2007, available at the Web site: <http://googleblog.blogspot.com/2007/07/cookies-expiring-sooner-to-improve.html>. If we think that we all use it almost every day, we immediately realize that this undertaking has very little impact on privacy.

[17] See **C. SOGHOIAN**, "The Problem of Anonymous Vanity Searches", Winter 2007, available on the Web-site: <http//ssrn.com>.

Google features and services may not function properly if your cookies are disabled"[18]. Thus, the declaimed reason lies in the wish to improve the quality of the service! What's a pity if we understand that in the meanwhile Google is starting the most extended and unprecedented in human history "Database of Intentions": "the aggregate results of every search ever entered, every result list ever tendered, and every path taken as a result"[19]. Incentives to maintain this huge amount of information are represented by: the relatively low gathering and maintaining costs, the lack of a unique and clear regulation on this issues (this even more in the US legal system), and the very remunerative potential utilization of information.

The Database of Intentions symbolizes also a very valuable good, a sort of goldmine if we take into account the economic value of information contained, that are a great temptation for many people: national security agencies, hackers, identity thieves, etc.. Just now, the search engine provider are not allowed (or should not be allowed) to sell personal data to third person. Hereafter it does not mean that it is not going to be allowed in the future, they exchange user data with subsidiary companies or other trusted business partners, in order to process this information and to provide services[20].

However that may be, there is always a third part with respect to data processing that could try to obtain personal data communication from the search engines, through a completely lawful procedure. I am referring to public authorities that could be interested in using log files for national security purposes. Especially after the tragic event of September 11th, an in progress trend is modifying the way in which citizens consider the State and its tasks, and is gaining force, remarking the traditional role of public security guardian. At the beginning the State was remaining in the background in Internet regulation, but now it is more and more starting again to play a protagonist role, making strategic alliances with ISPs: these person are in possession of very precious information for prevention and contrasting activities against criminality. Then, we talk about "Invisible Handshake"[21].

---

[18] See "Google Privacy Policy", available on the Web-site: < http://www.google.com/intl/en/privacypolicy.html>.

[19] Term taken by **J. BATTELLE**, "The Database of Intentions", *John Battelle's Searchblog*, 13 November 2003, available at the Web site: <http://batellemedia.com>; see also **ID.**, *The Search: How Google and its Rivals Rewrote the Rules of Business and Transformed our Culture*, Boston, MA - London, Nicholas Brearley Publishing, 2005.

[20] Again in the Google Privacy Police we read: "We offer some of our services in connection with other web sites. Personal information that you provide to those sites may be sent to Google in order to deliver the service. We process such information in accordance with this Policy. The affiliated sites may have different privacy practices and we encourage you to read their privacy policies".

[21] See **M.D. BIRNHACK, N. ELKIN-KOREN**, "The Invisibile Handshake: The Reemergence fo the State in the Digital Environment", *Va. J. L.* Draft, 2003, vol. 3., available at Web-site: <http://ssrn.com>.

I took into account the Google example, but the same considerations could be developed for the scientific specialized portals that offer services of the same kind. Actually most of them require a prior necessary authentication in exchange for the download of full text materials proposed by the Web site (otherwise you are allowed to get only the abstract of the paper: see, as example, the Social Science Research Network (http://ssrn.com).

IV. Conclusion: toward the overcoming of privacy vs. knowledge conflict

At the end of the way I lead, let's pull together the strings of what I said, trying to unveil, if it is not already clear, what the Odin saga has to do with our issue.

Whether we are dealing with fruition of scientific knowledge under the proprietary schemes transposed in the digital context by DRM systems, or with free access to contents following the trend that brings to Open Access, we are losing (exchanging, if we were aware of that) our identity, our profile, our data in return for intellectual works we wish to enjoy. As it happened for Odin, we are handing over an eye, that means a very important part of our body, to Mimir, that is the holders of the knowledge we are so fervently seeking. Remaining on the same mythological example, other interesting – and at the same time appalling – similitudes occur to us. To drink by the magic fountain a gulp of knowledge we hand over an eye, and with that we are giving to other persons the possibility to see inside us through the given eye, to know what we think, what we desire, briefly speaking to appropriate our most hidden, and simultaneously most precious, part: our thoughts.

Privacy is not only characterized by the tridimensional concept I described before (spatial, informational, decisional). It represents also a cultural phenomenon: everybody must be fully aware of the risks connected to an unlawful data processing and becoming themselves the first "security measures", proper to avoid illegal utilization of data. Until Internet users understand the risks connected to the processing of their personal data and the real cost they pay to get access to the several services offered by the Net, they will always be dreadfully exposed to the monitoring of all those who, on the contrary, commodified exactly this awareness.

Furthermore, we may had to say: scientific knowledge vs. privacy. And actually this conflict is manifest enough, I hope even more. But, if you really want to solve the contrast, you have to get over the traditional way to conceive the opposite poles.

Privacy does not consist only in making some rules, some restrictions, after all some obstacles to data circulation; it must be declined in the correct management of the flows of

data that marks out the information age in an ineliminable way. Knowledge, on the other hand, knows the already famous phenomenon of dematerialization that changes books, poems, music, drawings, into files expressed in binary code. Maybe it does not anymore make sense to conceive both terms of our subject in a physical way, carrying on a contrast that does not exist anymore in the digital context. It would be better to start reasoning on the fact that they can be considered as they are in the digital world, that is information, and thus reconsidering rules, technologies, and customs in the light of this new uniforming category that moves our attention to the management aspect of the exchanging flows, rather than to their inherent diversity.