Cloud Computing: Legal Issues in Centralized Architectures

Primavera de Filippi, Smari Mccarthy

Cloud Computing: Legal Issues in Centralized Architectures

Primavera DE FILIPPI[1], Smari McCARTHY[2]
[1]*CERSA, 10 rue Thénard, Paris, 75005, France -, Email: pdefilippi@gmail.com*
[2]*IMMI, Laugavegur 3, Reykjavik, 101, Iceland - Email: smari@gmail.com*

Abstract: Cloud computing can be defined as the provision of computing resources on-demand over the Internet. Although this might bring a number of advantages to end-users in terms of accessibility and elasticity of costs, problems arise concerning the collection of personal information in the Cloud and the legitimate exploitation thereof. To the extent most of the content and software application are only accessible online, users have no longer control over the manner in which they can access their data and the extent to which parties can exploit it.

Keywords: Cloud Computing, Centralization, Privacy, Confidentiality

1. Introduction

The advent of "cloud computing" has created an imbalance in authority structures that is very similar to the structural changes witnessed during the Industrial revolution. Just as the industrial revolution has progressively alienated workers from the means of production, today, most of the means of production (in terms of hardware, software, content or data) are concentrated within the hands of large Internet service providers.

Although the Internet constitutes a great opportunity for users to express themselves and to engage in collaborative production, many modern web applications are decreasing the capacity (or willingness) of people to produce content by their own means. The problem has been exacerbated by the deployment of Cloud Computing. Given that everything can be stored, processed, or executed on any computer system regardless of its whereabouts, most of the means of production, as well as the output of production (user-generated content), are increasingly owned or at least controlled de facto by large companies.

The trend is clear. Resources are moving away from end-users, towards centralized systems that possess huge processing power and storage capacities. Users' devices are devolving from personal computers to laptops, smart phones or integrated devices whose main function is to access particular sections of the Cloud through browsers or mostly dumb applications. While front-end processing is perhaps becoming slightly more common in the form of in-browser application, data storage is heavily biased towards centralized back-ends.

The implications are numerous: users are giving away their labor under an expectation of reciprocity; they are giving away their privacy for the sake of a more personalized service; they are giving away their rights in the name of comfort and accessibility; but, most importantly, they are giving away their freedoms and very frequently they do not even realize it.

The paper will analyze the impact of Cloud Computing on society. By analyzing the way the Internet has developed over time, it will draw attention to the fact that the Internet has been and is evolving in a way that might strongly impair the right to privacy of end-users and endanger the confidentiality of information stored into the Cloud.

2. The Emergence of Cloud Computing

2.1 – Definition of Cloud Computing

Given its recent and very fast adoption in everyday language, the actual definition and scope of Cloud Computing are still under debate. In part, this stems from the fact that Cloud Computing does not actually provide much in terms of new technology, but rather an alteration of the use of older technology to serve new types of business structures. For the purposes of this paper, we consider Cloud Computing to represent the sharing or storage by users of their infrastructure or content on remote servers which are accessible online. Although this can be achieved at many levels - i.e. at the level of the infrastructure (IaaS), platform (PaaS), or software (SaaS) - this paper is only intended to analyze the consequences of Cloud computing on the privacy of end-users. The focus will therefore be set on the concept of public Clouds, intended as a variety of applications that users can access and use through their browsers as if they were installed on their own computers or devices [1]. Although not all public clouds are browser-based, this focus does not come out of thin air, as the browser is increasingly used as a catch-all approach for user applications.

Although this is generally seen as an advantage by end-users, in terms of flexibility of access and scalability of costs, these benefits necessarily come with a cost. Indeed, while the Internet might have been regarded early in its existence as a possible implementation of a decentralized market economy [2], we are moving toward a thoroughly centralized market where the power of the service providers increases as the power of end-user terminals decreases. Since the heavy processing is performed in the Cloud and only the results are displayed to the users, neither high processing power, large amounts of RAM, nor even hard drives are nowadays required on the user-side to perform most everyday operations. A smart phone connected to the Internet can be more powerful than an actual computer because it can borrow storage capacity and computational resources from the thousands of machines that constitute the Cloud; any complex processing is done remotely while the front end simply deals with presentation. The technical characteristics of the terminal are no longer relevant as (a) software is for the most part executed through online servers, and (b) data no longer resides on end-user devices, but is instead stored in the Cloud.

This trend suggests that most of the computing activity that is today performed locally on end-user computers will eventually shift into the Cloud. Whether or not this is desirable, from the perspective of end-users, depends on the way the Cloud is implemented and on the policy of the Cloud provider, in particular, in terms of privacy and data protection. The problem is, however, that policy is inherently malleable. In practice, there is not any privacy policy, uptime assurance or data protection mechanism that can eliminate the added operational risk created by shifting to a third party infrastructure. At best, the risk can be minimized by not storing sensitive data and mitigated by not relying on one single cloud platform.

2.2 – The changing face of Networked Services

On the early Internet, centralized services were uncommon. Service providers were of small enough scale that utilizing the distributed nature of the network was a necessity. To wit, most early websites were developed to cater local communities, competing head to head with older peer-to-peer (P2P) systems and protocols, such as e-mail and Usenet, and working from a very limited set of use-cases and metaphors. There was a strong initial momentum towards community-based websites and user-driven journalism, with lengthy articles and feedback. Before the advent of blogging platforms such as Wordpress.com, Livejournal.com or Blogger.com, it was not uncommon for small groups of people to set up a web server to host personal home pages, frequently running on custom made software managed by somebody in the group. Likewise, instant messaging and interactive discussions were generally done through direct communication between peers and on decentralized platforms, such as the Internet Relay Chat (IRC), as opposed to the more centralized systems which have emerged today, such as ICQ, Microsoft Messenger or Skype.

The case of social networks is particularly interesting given their manifest evolution from a local and community-centric to a global and extremely centralized architecture. Prior to the globalization of social networking sites such as MySpace and Facebook, smaller scale social networking sites were common within local communities, such as hugi.is, an interest-based social network in Iceland, irc-galleria.net, a Finnish website providing social networking and photo gallery services to IRC users, and cu2.nl, a Dutch social network offering forums and photo galleries, amongst other things. Most early social networks did not manage pair-wise relationships between users. User relations were typically flat and unrestricted, with all users of the system seeing each other profiles and general information. Initially introduced in such systems as MySpace, Orkut and Bebo, pair-wise relationships have since then become part and parcel of any system intending to provide social networking, although symmetric relationships are not always necessarily the desired format. Twitter was the first major social network to demonstrate the value of asymmetric relations. Today, the majority of social networking websites provide similar features and characteristics. All provide public and private messaging systems, albeit with variable levels of service and emphasis [3]. Some systems allow photographs or other media to be added, such as Facebook and MySpace in particular, which allow photo albums, videos and other rich media.

Accepting these variations on the theme and acknowledging the untold other differences, we will focus the remainder of this study on two social networking sites; one local, the Icelandic site Hugi [4], and one global, the infamous Facebook.

Technologically, Hugi is very similar to the early Facebook [5]. Even today, apart from the improved friendship management, the technology behind Facebook is not far removed from that of Hugi. Facebook most certainly has a far more polished user interface and a much more weighted approach to features such as internal chat, external chat through XMPP, statuses and other aspects of messaging, but most features are primarily user experience tweaks which have come along over various iterations of the Facebook user interface [6].

Until 2003, a large portion of Icelandic people aged from 16 to 24 were actively contributing on Hugi. Today, however, most of the user-base has nowadays shifted to Facebook. As of 2011, it is estimated that over 65% of people in Iceland have accounts on Facebook [7].

While there are certainly many elements of user interface which influence people towards using Facebook, as the various interface changes to Facebook have shown, it is hard to believe that the trigger is merely a technical one. Rather, we claim that the key factor for the shift from Hugi to Facebook was essentially due to the more integrated and international nature of the latter, as opposed to the local character of the former.

In order to back up this claim, an online questionnaire was sent to some former users of Hugi and current users of Facebook. The results reveal that the scope of the service (i.e. its extension in the Internet landscape) weights very strongly in the mind of end-users. Despite a general inclination towards the private management of personal data, all users have declared to value the size of the community and the worldwide scope of the platform above other factors [8].

As a result of their difference in scope, the two services are not even considered to serve the same function by many users [9]. Hugi is little more than a communal sounding board which maintains a local culture fitted to meet the needs of its original operator, Síminn, a telecommunications company. Facebook, on the other hand, is both an agora and a marketplace. Like Hugi, it is controlled by a single company, but, unlike Hugi, it has reached global significance. As a commercial start-up, the goal of its operator is to increase the number of users on the network, as well as their dependency upon it, so as to lock a maximum number of users into the system [10].

Network effects are such that the more users are on a platform, the more valuable the platform is to each user. In spite of their significance in the context of social networks, network effects are not, as such, a sufficient justification for there to be only one centralized social networking platform [11]. The network is fully capable of allowing for decentralized systems, as various peer-to-peer protocols have demonstrated [12]. It is possible to devise a peer-to-peer infrastructure based on an open protocol, which would allow users to keep control over their own data, and, theoretically, even to use the social network locally on their computer, without the need for any Internet connection. This is, for instance, the ultimate goal of Diaspora, a distributed and open-source social network that purports to enable users to control their respective nodes in the network [13]. The problem is that social networking platforms were primarily developed by companies with vested interests in holding as much mind-share as possible. Thus far, with the exception of Diaspora, no peer-to-peer system has emerged to compete with such centralized systems.

Although the analysis has focused exclusively on social networks, the intention was to provide an example to illustrate a general trend: the increasing concentration of the market and the consequent concentration of power in the hands of a few enterprises. We believe the conclusions of this analysis to apply, by and large, to the majority of applications provided by large centralized companies over the Internet.

Because of their dominant position, large service providers can exert a degree of subjugation never conceived of by smaller and more local services, and a degree of control that would be impossible in a peer-to-peer network. This creates a series of legal issues in terms of control, privacy, and confidentiality of information that will be specifically addressed in the following sections.

3. Legal Issues of Cloud Computing

It takes only very basic examples to show the danger of over-centralization in the sphere of the Internet. In addition to the most common examples, such as Google and Facebook, there are a very large number of actors whose operations are crucial in the everyday life of many Internet users. While the level of dependency increases, the effects of not having control over the infrastructure become more apparent, although some of the implications might become very subtle. As user no longer control nor understand their infrastructure, they are increasingly controlled by those who do know how to control the infrastructure - and by those who own it.

3.1 – Centralized Control

Today, no matter how much one tries to keep it secret, there exist many mechanisms or devices that collect personal data and communicate it to third parties without the consent of the data subject [14]. Most often, however, it is actually the user who willingly communicates information to a variety of interested parties. On the Internet, this is done on a daily basis through blogs, forums, newsgroups, mailing lists, search engines, etc.

While this is not a concern in itself, a series of problems might eventually arise if all that data were to be gathered together into one large database. If one single entity were to provide a large variety of services and the data collected through all of these services were to be processed into an integrated framework of analysis, that entity would fundamentally be able to know much more about its user-base than what has been voluntarily disclosed by each individual user.

Technically, this is already a possibility, and, as a matter of fact, this is already part of reality. Let's take a look at Google. With a mission to "organize the world's information and make it universally accessible and useful", Google offers a large variety of services, mostly for free for the end user, whose ultimate purpose is not only that of presenting information in a more organized way, but also that of gathering as much information as possible. Services such as Google Mail, Google Documents, Google Calendar, Google Maps, Google News, Google Reader, Orkut, Youtube, Picasa - and many more - are all intended to collect information about the users of that service. Even a service apparently as harmless as the Google search engine is in fact able to collect very important pieces of information. A cookie (whose expiration date is irrelevant for any practical matter) is stored into every computer so that it can be identified at every subsequent connection [15]. Although it allows Google to collect all sort of information about users, this cookie is presented as a valuable service to the users who would otherwise be unable to enjoy the benefits of personalized search results and customized advertisements.

Since most of these services are either available online or automatically synchronized whenever a user connects to the Internet, Google can keep track of every user activity performed on its system. This data can be very valuable for the purposes of mass profiling (i.e. understanding the preferences of the user-base as reflected by the behavior of each individual user) and user profiling (i.e. understanding the preferences of each individual user through the analysis of its specific interests, activities, and social surroundings) [16].

Google, however, is ultimately not interested in monitoring the activities of its users, nor in gathering information about the socio-demographics of its user-base, but rather in the maximization of profits. Profiling is necessary for Google to know what users want, so as to eventually offer them the most personalized results and the best kind of advertisements. The greater the user-base, the most accurate the profiling can be, and the higher the profits that can be extracted from a system of customized advertisement dependent upon the interests of each individual user. In this case, the fact that the end-users do not pay for the service means that they themselves are the product being sold, or rather, statistics about them are.

Various companies have built successful business models around the realization that, instead of getting money in exchange of a service, it is often more valuable to provide services for free in order attract a maximum number of users. By accepting the terms of services, users agree to share most of their data and information with Google, regardless of the privacy or the confidentiality thereof [17]. Hence, although the majority of Google's services are offered for free, users pay - willingly or not - with their own data, which is only later turned into profit by Google AdSense or other forms of advertisement.

In this context, the scope of the Cloud is extremely important. By offering such a wide variety of services, Google is able to obtain different pieces of information which pertain to different fields of endeavor. When users search for something on the web, Google can learn about their interests; when users read their emails on Gmail, Google can learn more about their personal or professional life; when users check out a location on Google Maps, Google can learn where each user has been or wants to go. The greater the scope of the Cloud, the greater is the amount of data that can be gathered together and the more valuable is the information that can be obtained with the processing and correlation of such data [18].

While this is likely to help Google increase its profit, the collection and processing of user data into a common integrated framework can also benefit the users when it comes to increasing the quality of the service. Many users are therefore not merely agreeing, but even eager to share their personal data and information with Google in order to obtain a more customized and integrated service. Google Calendar is more valuable because it can be integrated with Gmail for e-mail reminders and notifications and with Orkut for discovering new events and remembering the birthdays of some friends. As the value of a service increases not only with the number of users connected to that service but also with its degree of integration with other services, the wider is the portfolio of services offered by Google, the most users will be attracted to these services.

The problem arises when the information given to separate (and apparently independent) services is actually aggregated together by one single entity (either because it is the common provider of said services, or because it has acquired the data from third parties). Even though the information had been voluntarily provided by users, aggregated data might provide further information about users, which they did not necessarily want to disclose.

3.2 – Privacy & Confidentiality

There is an inherent security risk in the use of the Internet to transfer sensible information and personal data. As a general rule, information wants to be shared, as most of the value that can be extracted from it emerges from the usage and communication thereof. However, whenever it is published on the Internet, the privacy and confidentiality of information is necessarily put at risk [19].

These risks have been considerably increased with the deployment of Cloud Computing, which requires more careful attention to be given to its actual or potential consequences on the privacy of personal information and confidentiality of business or governmental information. What kind of information can be shared into the Cloud? Can anything be kept private in the Cloud?

Individuals are generally free to share information in the Cloud, even though they are often not fully aware of the terms of services set out by the Cloud providers and of the consequences of storing information in the Cloud [20].

In the case of an institution, privacy laws may sometimes prohibit or limit the disclosure of personal information to third parties. The possibility for a business or corporation to share information in the Cloud is subject to a series of standards established by different bodies of law, whereas government agencies are restricted by internal rules and public regulations on data protection. For instance, in the USA, the Health Insurance Portability and Accountability Act (HIPAA) establishes a series of rules regulating the use and disclosure of identifiable health information, which can only be transferred to a service provider that promises to comply with the same set of standards (often incompatible with the terms of services established by a cloud provider). Similarly, the Violence Against Women Act precludes domestic violence service providers from disclosing information without the consent of the data subject, unless compelled by statute or a court (Public Law 109-162 as amended by Public Law 109-271); tax preparation laws provide statutory and regulatory protection that limits the disclosure of tax return information without the taxpayer's consent (Internal Revenue Service rules - 26 U.S.C. § 6713 and § 7216; 26 C.F.R. §301.7216); whereas the disclosure of personal information concerning the financial situation of a consumers by a financial institution is precluded under the Gramm-Leach-

Bliley Act (15 U.S.C. § 6802); and the disclosure of video rental and cable television subscribed records is protected under the Video Privacy Protection Act (18 U.S.C. § 2710) and the Cable Communications Policy Act (47 U.S.C. § 551).

Despite the fact that the law may restrict the ability of these institutions to rely upon the services of a Cloud provider by introducing a series of procedural and/or substantive barriers, many corporate and governmental institutions (and in particular municipal governments) do store their data remotely on databases and file systems operated by a third party by contract. The problem is that information stored in the infrastructure of a third party may have weaker protection than information that remains in possession of users.

The chances for inadvertent exposure increase substantially with every new intermediary and with every new layer of abstraction. While securing the infrastructure is obviously very important, it is not sufficient if the interface or application running on that infrastructure has not been properly secured as well. Although users need a way to log into the system in order to transfer data from or into the Cloud, this could constitute a significant security risk unless proper access control and secure transfer protocols have been adopted. Likewise, even though users are made to access the services by password, unless there is filesystem level encryption of the data with a key held only by the user - which is impractical in most cases - the operator of the service or anybody else who gains physical access to the servers can peer into the stored data. In more extreme cases, attacks on the hardware can be used to extract information that is resident in runtime memory [21].

In most cases, security issues are due to lack of or poor application of cryptography and a general lack of tradition for security. Various campaigns have tried to remedy this, such as the Tactical Technology Collective's ONO Robot campaign, Survival in the Digital Age [22], and the Electronic Frontier Foundation's HTTPS-everywhere campaign [23].

Yet, regardless of the degree of protection promised by the cloud provider, the security and confidentiality of information is ultimately determined by the weakest link in the chain. Insofar as data is transferred through several intermediaries, only one of them needs to be violated for any malicious user to obtain the relevant information.

In addition, the laws of certain countries oblige cloud providers to communicate to the authorities any information that constitutes evidence of criminal activities. This means that government agencies can, under certain circumstances, require the disclosure of personal or confidential information. This information can be more easily obtained from a third party than from the original owner, who, in the absence of proper notice, does not have the opportunity to object. For instance, in the USA, although the Electronic Communications Privacy Act (ECPA) provides a series of protections against the access by governmental agencies to personal information held by third parties (18 U.S.C. § 2510-2522 and § 2701-2712), these protections have been subsequently weakened by the USA PATRIOT Act, which entitles the FBI to compel, following a court order, the disclosure by Cloud providers of any record stored on their servers (50 U.S.C. § 1862).

Finally, the international character of the Cloud adds an additional layer of complexity. Information stored in the Cloud can be subject to a variety of different laws according to the location where it is being stored or transmitted. A Cloud provider might avail itself of the services of other Cloud providers located in different jurisdictions, or, if the Cloud provider has data centers in multiple countries, it can transfer data between centers depending on economic factors such as the price of electricity. This means that a file being served from Luxembourg one moment could be served from the Philippines the next. The difficulty for the user to know with certainty which law applies to the information stored into the Cloud raises a number of data protection questions.

According to the European Union's Data Protection Directive, national data protection laws apply to all information located in the territory of a Member State, regardless of its origin or destination [24]. However, it is often difficult to determine in advance and with certainty the actual location of information stored in the Cloud. In particular, a crucial problem that emerges from the international character of the Cloud is the question of forum-shopping. Unless it has been contractually precluded to do so, a Cloud provider can theoretically move information from one jurisdiction to another in order to benefit from the most favorable laws. This can be used as a means for any service provider that does not want to respect domestic regulations on data protection. In order to overcome this problem, the European Union introduced the rule that data cannot be transferred to countries outside the EU which do not provide an “adequate level of protection” [25]. This is likely to reduce the possibilities for Cloud providers to outsource their services in the EU, because, even if data is merely being processed in a Member State, it might be difficult to export it after it has entered the EU.

As Cloud Computing is being adopted by an increasingly larger number of businesses and individuals, the underlying technology and infrastructure is continuously evolving, but the law does not seem to follow the pace. Given that commercial Cloud providers are more interested in making profits than in protecting the interests of their user-base, users should be wary of their privacy online and of the confidentiality of their data stored in the Cloud. Given the degree of legal uncertainty that is emerging in the sky, there is a real need for the law to be reformed in order to better accommodate current and future users concerns in terms of data security and privacy.

4. Conclusion

There are many consequences to the deployment of cloud computing: some intended, others unintentional; some good, and others bad. Many are already noticeable and measurable, while others can only be foreseen by analyzing the trends that have been set.

There is a trend fueled by the shift of control from end-users towards increasingly centralized services providers. As many such services, and in particular social networks, carry heavy privacy and confidentiality burdens, the threats to the privacy of end-users increases. Service Level Agreements and privacy policies are useless in the face of events which are irrevocable, such as the exposure of private data. Users of Sony's PlayStation network know all too well that this danger is not a hypothetical [26]. Smaller networks catering to more local communities distribute the risk and limit the scope of potential damage, but at a steep utility tradeoff.

The advantages offered by cloud computing are clear: infrastructure providers can benefit from strong economies of scale, whereas Internet service providers can benefit from enhanced flexibility and scalability of costs. From the perspective of end-users, the main advantages are the possibility to access data from anywhere and at any time - regardless of the device they are connected from - and the ability of avail themselves of the computing power and storage capacity of the cloud. Further, it allows for outsourcing the obligation of maintaining complicated infrastructure and having to maintain up-to-date technical knowledge, while externalizing the cost of purchasing and running the infrastructure.

This does not, however, come without costs. Exporting data to the cloud means that users can no longer exercise any kind of control over the use and the exploitation of data. Data stored in various data centers can be processed without the knowledge of users, to be further redistributed to third parties without their consent. If everything has been stored in the cloud, the cloud provider can ultimately determine everything that users can or cannot do. As most Internet users are no longer in charge of their own data and are no longer capable of managing their own infrastructures of production, storage, and distribution, the control is in the hand of few corporate entrepreneurs.

Just as, after the industrial revolution, governments have been urged to exercise their authority for the creation of labour and consumer protection laws, today, during the digital revolution, governmental intervention has become necessary in order to promote civil liberties and to protect fundamental rights on the Internet, at least with regard to those risks which cannot be properly addressed through the adoption of clearer policies by cloud providers and better practices by users.

References

[1] Cloud Computing can be implemented at various levels of abstractions and deployed either internally of externally. In the common sense of the term, Cloud Computing refers to the concept of a "public Cloud" as a service offered by a third-party that dynamically provides a series of resources accessible on-demand through the Internet, often via web applications. This can be contrasted to the concept of a "private Cloud" as a service for private networks allowing a company to host applications or virtual machines on its own premises.

[2] During its early phases, the Internet was often regarded by many pioneers and visionaries as a potential implementation of a pure market economy characterized by free exchange of information, low transaction costs and very few barriers to entry. See, e.g. Eric Schlachter (1994), Cyberspace, the Free Market and the Free Marketplace of Ideas, in Hastings Communications and Entertainment Law Journal (Comm/Ent) [16 Hastings Comm/Ent L.J. 87]; Yannis Bakos (1998), The emerging role of electronic marketplaces on the Internet, in Communications of the ACM, Volume 41 Issue 8; James C. Bennet (2001), The End of Capitalism and the Triumph of the Market Economy, in Network Commonwealth: The Future of Nations in the Internet Era.

[3] While they all provide users with a way to communicate with each other, different platforms provide different means of communication. Some allow threaded messaging while others only allow linear messaging. Some restrict the number of characters allowed in messages, for example 140 on Twitter, 450 in Facebook public status updates and 10000 in OkCupid private messages, while others do not impose any such practical restrictions.

[4] Hugi was originally operated by Síminn, the former state telecoms company which was privatized in 2005 with the sale of 98.8% of its shares to Skipti. It is now operated by Skjá miðlar ehf. For more information, see www.hugi.is

[5] Developed in the PHP programming language with MySQL as a database, and not providing much in the way of Web 2.0-style services beyond the level of user interaction presumed in such a setting; there is no post-loading processing which accesses server data, as through AJAX or other asynchronous HTTP requests.

[6] It can be expected that if Hugi had not been "neglected" similar updates would have followed there, although perhaps not with as great rapidity. In conversation with Hugis webmaster, in May 2011, it was said that, although Hugi had seen better times, a large cause of its decline was the neglect of the site's original owner.

[7] As of 2011, Iceland ranks first in terms Facebook penetration, with over 65.76% of the population on Facebook or 203 140 in total. For more updated statistics, see http://www.socialbakers.com/facebook-statistics/iceland

[8] In a small and informal questionnaire (n=30) amongst former users of Hugi, when asked whether, all other things being equal, they would prefer a service such as Facebook, but with their personal data hosted within Iceland, exactly half said they would; when asked if they would prefer a service where their data was hosted on their own private computer, 64% said they would. Younger people, in particular, seem less concerned with sovereignty over their own data, while older users appear more concerned about the locality of their data. Yet, all of those questioned said that the size and international aspect of Facebook mattered either much or very much.

[9] In the same questionnaire amongst former users of Hugi who also use Facebook, 82.15% claimed that Facebook and Hugi serve different roles, with the rest claiming that they only partially serve the same role.

[10] As for 2011, Facebook is valued at roughly 80 billion dollars (according to a recent private-market transaction on SharePost, an online marketplace for private investments) and has over 500 million users; meaning that each user's contribution, if we ignore the network effect, is about $160. Of course, given the nature of network effects, the most recent user added is always the most valuable. With 7% of humanity registered on the world's largest social network, the only way for Facebook to increase shareholder value is to aggressively reach out to an ever-growing group of users, while minimizing the risk that current users leave.

[11] Natural monopolies are justified by large economies of scale: a producer's cost curves decrease when the scale of production increases. Network effects describe the increase in value of a good or service derived from the standardization of that good or service. While natural monopolies often comes together with network effects, like in the case of the telephone network, network effects do not necessarily lead to natural monopolies, like in the case of the Internet network.

[12] Decentralized protocols are ubiquitous on the Internet. Giving an exhaustive list would be unpractical, but common examples include the Domain Naming System protocol (DNS), the SMTP protocol for e-mails, Bittorrent and Gnutella for file-sharing, Skype (which uses centralized coordination servers but attempts to make calls directly between peers), and, finally, FreeNET, i2p and TOR for anonymous navigation.

[13] Diaspora is currently using a client-server model, although it aims to eventually have federation support, which is a form of devolved P2P (currently used by Jabber/XMPP and IRC, amongst others). Federation means that, although the distinction between clients and servers remains, there are multiple servers that act as peers amongst themselves. The objective of Diaspora is to allow every user to decide whether to participate to the federation as a server, or whether to act as a mere client. For more details, see https://joindiaspora.com/

[14] Spyware programs (which are a form of malware) are malicious software that collects personal data about users without their consent. As users perform tasks such as browsing the Internet, spyware programs collect information about users and their behavior. Although commonly acknowledged in the digital world, similar devices are commonly deployed in the physical world, in the form of eavesdropping, interception of written communications, video surveillance through CCTV, and, most recently, identification via biometric data and geo-localization by means of GPS tracking and networking technologies. For a more detailed overview of the mechanisms and the consequences of pervasive surveillance in modern societies, see, e.g. David Murakami Wood (2008), Towards Spatial Protocol: The Topologies of the Persavise Surveillance Society, in Alessandro AUrigi and Fiorella De Cindio (Eds), Augmented Urban Spaces: Articulating the Physical and Electronic City; Ashgate Publishing.

[15] Every time a user connects to Google's search engine, a cookie is stored on the user's device, with an expiration date of two years. The expiration date is pushed ahead of two years whenever that cookie is accessed by any of Google's sites and it is detected that the cookie is about to expire. By virtue of this cookie, Google is able to store an almost permanent and unique ID on every user's device, as Google will either keep the same unique ID in the cookie, or at least be able to associate the old ID with any new ID that is issued. Although Google claims that the purpose of the cookie is to remember user preferences, the cookie is also be used for the purposes of profiling. See http://www.google.com/privacy/privacy-policy.html - "When you visit Google, we send one or more cookies to your computer or other device. We use cookies to improve the quality of our service, including for storing user preferences, improving search results and ad selection, and tracking user trends, such as how people search. Google also uses cookies in its advertising services to help advertisers and publishers serve and manage ads across the web and on Google services."

[16] Mass profiling is more concerned with the general trends and navigation patterns of the user-base than with the actual preferences and activities of each individual user. User profiling focuses instead on the personal and distinctive characteristics of users and is therefore more likely to infringe upon their right to privacy. For an overview of the various techniques used for the profiling of users in a Cloud environment, see, e.g. Olfa Nasraoui and Carlos Rojas (2003), From Static to Dynamic Web Usage Mining: Towards Scalable Profiling and Personalization with Evolutionary Computation, in Workshop on Information Technology, Rabat, Marocco, and, in particular, Gang Ren; Tune, E.; Moseley, T.; Yixin Shi; Rus, S.; Hundt, R. (2010), Google-Wide Profiling: A Continuous Profiling Infrastructure for Data Centers, in Micro, IEEE, volume 30, issue 4.

[17] Google privacy policy states that Google may collect all kind of personal information provided by users themselves, log in information gathered whenever users access one of the various Google's services, user communications, information gathered by cookies stored in users' devices or collected by third party applications, and location data in the case of location-enabled services such as Google Maps or Latitude. For more details on Google privacy policy, see http://www.google.com/privacy/privacy-policy.html

[18] Google's privacy policy clearly states that Google will be pooling all the information they collect from all of their services. Google reserves the right to "combine the information you submit under your account with information from other Google services or third parties in order to provide you with a better experience and to improve the quality of our services." See http://www.google.com/privacy/privacy-policy.html

[19] The advent of Internet and digital technologies introduced a series of concerns that might significantly affect users' willingness to communicate personal data and confidential information over the Internet. Given that there can be no perfectly secure mechanism to transfer information, publishing information on the web necessarily involves the risk of data loss or spill over. See e.g. Bob Blakley, Ellen McDermott, Dan Geer (2001), Information security is information risk management, in Proceedings of the 2001 workshop on New security paradigms, New York; and Eric C. Turner; Subhasish Dasgupta (2003), Privacy on the Web: an Examination of User Concerns, Technology, and Implications for Business Organizations and Individuals, in Information Systems Management, Volume 20, Issue 1.

[20] While many users do not even bother to familiarise themselves with the terms of services of the cloud computing platform they wish to use, doing so is often not an easy undertaking even for those who try to understand the consequences of entering into such agreement. Besides, it is fairly common that the provider reserves the right to vary the terms and conditions on which the service is provided without notifying the users. For more details, see Dan Svantesson, Roger Clark (2010), Privacy and consumer risks in cloud computing, in Computer Law & Security Review, 26 (4), 391-397.

[21] An interesting example is the Cold boot attack, allowing anyone with physical access to a computer to retrieve encryption keys from the operating system after restarting the machine. The attack relies on the "data remanence" of DRAM and SRAM memory in order to retrieve memory contents that remain readable for a short period after power has been removed. For more information, see J.Alex Halderman, Seth D. Schoen, Nadia Heninger, William Clarkson, William Paul, Joseph A. Calandrino, Ariel J. Feldman, Jacob Appelbaum, Edward W. Felten (2008): Lest we remember: Cold Boot Attacks on Encryption Keys, in Proceedings 2008 USENIX Security Symposium.

[22] The Tactical Technology Collective and ONO Robot produced a series of animated films to raise awareness about the digital traces users leave behind. Its main aim is to engage people in better understanding the information and communications technologies they are using, so that they can decide when and if they want to take risks. For more details, see www.onorobot.org

[23] HTTPS Everywhere is a Firefox extension produced as a collaboration between The Tor Project and the Electronic Frontier Foundation. It encrypts communications with a number of major websites using Transport Layer Security. For more details, see http://www.eff.org/https-everywhere

[24] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Article 4 (National law applicable) specifically states that each Member State shall apply the national provisions it adopts pursuant to this Directive to the processing of personal data where: (a) the processing is carried out in the context of the activities of an establishment of the controller on the territory of the Member State;; (b) the controller is not established on the Member State's territory, but in a place where its national law applies by virtue of international public law; (c) the controller is not established on Community territory and, for purposes of processing personal data makes use of equipment, automated or otherwise, situated on the territory of the said Member State, unless such equipment is used only for purposes of transit through the territory of the Community

[25] Ibid, Article 25 introduces the principles that Member States shall provide that the transfer to a third country of personal data which are undergoing processing or are intended for processing after transfer may take place only if, without prejudice to compliance with the national provisions adopted pursuant to the other provisions of this Directive, the third country in question ensures an adequate level of protection; where the adequacy of the level of protection afforded by a third country shall be assessed in the light of all the circumstances surrounding a data transfer operation or set of data transfer operations; particular consideration shall be given to the nature of the data, the purpose and duration of the proposed processing operation or operations, the country of origin and country of final destination, the rules of law, both general and sectoral, in force in the third country in question and the professional rules and security measures which are complied with in that country.

[26] In April 2011, Sony suffered a breach in the Playstation online video game network. As one of the largest Internet security break-ins, this breach led to the theft of personal data, such as names, addresses, birth dates, passwords and possibly credit card numbers belonging to 77 million user accounts. This required Sony to shut down the network, and although Sony given notice of the breach to its customers, no information has been provided as to how the data might have been compromised.