"My data are still mine?"

Alfredo M. Ronchi
EC MEDICI Framework
alfredo.ronchi@polimi.it

Global Forum - Eindhoven
19,20 September 2016

Introduction

November ’90 on the occasion of COMDEX Fall Bill Gates introduced the his vision “information at your fingertips”, few months later, to stress the concept, he said that the real wealth in the future will be access to information, people will no more ask “how many dollars do you own” but “how much information can you access”. In a glimpse this vision become reality and twenty-six years later “information” is still a powerful “transversal” asset: business, trade, policy, security, tourism, health, … rely on information, reliable information.

In a single generation we witnessed the evolution of information technology from mainframes exclusive patrimony of space agencies and super-calculus centres to owning in their pockets a device ten thousand times more powerful, capable of observing and recording video, audio, location, and motion. These devices can communicate with nearly any other digital device from household appliances even to cars. Collectively we have the ability to store, access, and process more data than humanity has created in its entire history. The actual “visual” trend is producing an incredible amount of photo/video documentation of our everyday life; does this mean goodbye privacy?

The so called Internet Revolution gave a boost to data creation and dissemination, MAC addresses, web logs, voluntary or unintentional applications to web sites and services, and social platforms ignited the sedimentation of personal and many times sensitive information apparently lost in the cyberspace.

Among the long list of similar examples simply refer to the one due to Herbert H. Thompson, as a professor, a software developer and an author he has spent a career in software security, on August 2008 he published on Scientific American an article entitled “How I Stole Someone's Identity” [4] providing a detailed description, in seven steps, about the way in which easily he stole the identity to another person accessing his/her bank account, social security etc. etc. This result is often achieved thanks to a combined access to different datasets, identifying a correlation between apparently anonymous unrelated data.

Information is built on top of single or aggregation of data, for quite a long time people use to think that cyberspace is a black hole without “memory” where you pour data without any side effect. So far especially young generations shared on line sensitive information in

order to access a videogame or chat with friends or more recently post images and clips about their private life.

In the “Appification” era there are almost no limits to data collection and reuse, “someone” knows exactly where you are now and where you have been, APPs may collect your medical data, fitness program, your expenses or collect and analyse your contacts, your photos or video clips. Social and communication media complete the panorama adding a “private depth” to the general fresco. In recent times crowd data collection, open data and big data more or less anonymised provided the big framework.

This is not enough, what it is not collected by APPs it will be collected in a seamless mode by IoT (Internet of Things) [2]. We live in a world in which there are already countless sensors and smart objects around us, all the time. The car we drive, the phone in our pocket, our wristwatch, the clothes we wear, are smart and connected, then the concept of “private” becomes far more ephemeral. Of course IoT will add a lot to our life but this will cost us a significant part of our privacy.

Starting from all these aspects the present document will deal with main aspects concerning ownership, moral rights, privacy, ethics, security and more.

Owning Information

Historically speaking, the idea of even owning information is relatively new. The earliest copyright laws, which granted the creator of artworks, among the other rights, exclusive rights to duplication and distribution of said work, first appeared in the early 18th century. Nevertheless it would still be hundreds of years, however, before the concept of "data" as we understand it even began to develop.

The world we contributed to create, filled up with cutting edge technologies and fully connected take us to a simple, even if uncomfortable to hear, truth: we are unable of preventing all possible data tracking. Cameras, satellites, sensors and software virtually everywhere ensure that, no matter how much technology you eschew, someone can get some data off of you. Your credit card company tracks your purchases and in one word your life style. Your phone carrier tracks your calls, social relations and geographic location. Your area's law enforcement tracks the roads and intersections you walk through or drive down every day. Local administration CCTVs or private safety cameras follow you within shops or residential buildings even inside the elevator.

Unless we decide to move to the mountains renouncing to nowadays technology, some tiny data that describes our behaviour and us will probably be tracked. No matter you may say, we have nothing to hide, but what about the use, abuse or misuse others may do?

Data and Ownership

The concept of "data" as it relates to people’s everyday life is still evolving [7]. We inherited the concept of copyright and we more recently faced the concept of privacy.

Copyright and privacy, it seems reasonable that both derive from the concept of data ownership. I take a nice picture, put a watermark on it and publish on my web page, if someone else download my picture crop the watermark and posts it on his/her website, it's a copyright infringement. Nowadays open data is one of the buzzwords most popular, if a public authority will release different sets of “open data” apparently anonymised but the

combined use of them may lead to identify your personal behaviour that's a form of privacy invasion or perhaps violation [8].

Following the same *fil-rouge* on the borderline between licit and illicit activities, simply consider an unseen observer that follows you and take notes about all the different places you visit and the time of your visits, he does nothing with this information, simply store it in his notebook, he is unseen and you will never face him and discover his activity, basically in doing so he didn't broke any law. His behaviour is unconventional but still legal. If you act in public spaces or visible by public there are no laws that state that you are the sole proprietor and owner of the information regarding your public life, the collection of this information doesn't violate any right. If we look in law the closest legal offence in such a situation is stalking even if this offence usually is directly connected with harassment but the unseen observer does not ever interfered with you so no harassment, no stalking even because the unseen observer is your smartphone and it can't be convicted of stalking you.

Cyberspace is really a Black Hole?

Some people probably consider cyber space as a kind of "outer space" no man's land not subject to human's material desires and malicious behaviours. Voluntary or involuntary personal data dissemination it is not a new phenomenon, before the Internet it was less evident and limited to some specific domains: credit card companies, travel agencies, real estate companies, car dealers, etc. etc. basically people officially owning your personal information being in a position to suggest new opportunities. Later on it was the time of "fidelity cards" and the explosion of CRM[1]. The mass diffusion of the Internet ignited the real blast of personal information collection and data harvesting. You fill up a form to install a new APP and suddenly you receive a bunch of offers and advertisement often claiming that you subscribed that service. Yes you subscribed the form to install the APP but thanks to a kind of letter chain the company in charge for collecting the forms to install the APP is the same company that manages dozens of business companies and you unintentionally subscribed the full service. Your personal information are now shared among a number of companies and you will never be sure that they will disappear from on-line data base. This last aspect, "never disappear" take us to another relevant point. Introducing the concept of data ownership we make reference to the copyright concept. If my data are mine I can even delete them isn't it?

Copyright and copyleft are two sides of the same coin, they both pertain to the intellectual property of something, but which is the most relevant... if any? Traditionally, copyright and copyleft have been regarded as absolute opposites: the former being concerned with the strict protection of authors' rights, the latter ensuring the free circulation of ideas. While copyright, which seeks to protect the rights of inventors to own and therefore benefit financially from the new ideas and products they originate, thus encouraging further product development, is associated with a vast amount of legislation globally (leading to corresponding applicative complications), few studies have been made of copyleft. Indeed, a commonly held belief about copyleft is that it begins where the boundaries of copyright end, spreading over a no man's land of more or less illegal exploitation.

[1] Customer Relationship Management

[2] DIRECTIVE (EU) 2016/680 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016, entered into force on 24

If we specifically refer to the intellectual property from the "continental" standpoint apart from the "economic" rights we find, even more relevant, some moral rights like paternity, adaptation, modification, … "withdraw". The author has the moral right to "withdraw" his work of art from private or public environment. If we keep the similarity in the field of personal data we must claim for the right to withdraw them from the "digital universe", this right is usually termed "right to obsolescence" or the "right to be forgotten". Viktor Mayer-Schönberger, the author of "*Delete: The Virtue of Forgetting in the Digital Age*" [3] traces the important role that forgetting has played throughout human history. The book examines the technology that's facilitating the end of forgetting: digitization, cheap storage and easy retrieval, global access, multiple search engines, infinite replications of information, etc. etc. If it is true that our ancestors survived to the evolution process because of their ability to transfer to future generations relevant information thanks to primitive forms of writing the dangers of everlasting digital memory, whether it's out-dated information taken out of context or compromising photos the Web won't let us forget is as well evident and already creating troubles. The supporters of a "natural" approach propose an expiration date for digital information or a progressive vanishing of data as it happens in the human world. Other experts propose to apply the moral right of the author/owner to "withdraw" his data, and here it comes the first crucial point: author, owner or subject … ? A vanishing memory offers the ability to make sound decisions unencumbered by the past, offers the possibility of second chances.

Laws and Regulations

As it appears from the previous paragraphs ownership of data is not yet a well-defined legal concept. We all agree about privacy and intellectual property infringement but personal data even if clearly belonging to the same "galaxy" are not properly identified and protected.

If this represents the state of the art in general it might not always be the case. Individual nations and international organizations are attempting to establish rules governing who can collect what data and what they're allowed to do with it. Germany, in fact, has a legal concept known as "informationelle Selbstbestimmung" or informational self-determination. What does informational self-determination mean? An individual has the right to decide for him or herself what information can be used by whom and for what.

UNESCO Information for All Programme (IFAP) [16] invested some resources to better focus ethical aspects with regard to the information society; the outcome of such studies is the definition of Infoethics. Quoting UNESCO IFAP: "The international debate on information ethics (infoethics) addresses the ethical, legal and societal aspects of the applications of information and communication technologies (ICT). Ethical principles for knowledge societies derive from the Universal Declaration of Human Rights and include the right to freedom of expression, universal access to information, particularly that which is in the public domain, the right to education, the right to privacy and the right to participate in cultural life. One of the most challenging ethical issues is the inequity of access to ICT between countries, and between urban and rural communities within countries.

Along with the benefits of a digitally connected world come the threats of misuse and abuse. Already countries are building mechanisms to protect their people against these risks, for example to ensure the safety of children on the Internet, but clearly a lot more

needs to be done to address the ethical implications of the information society. In collaboration with its partner institutions, IFAP seeks to do so."

The threats of misuse and abuse are again one of the major concerns. More recently personal information ownership and ethical aspects connected to open data represented one of the key subject on the occasion of the UNESCO IFAP [17] International Conference Media and Information Literacy for Building Culture of Open Government [6], held in Khanty-Mansiysk, Russian Federation, on 7-10 June 2016.

Some of the most relevant legal implications explored on the occasion of the Khanti Mansiysk event were interaction among stakeholders requires related competencies such as reliable information access and retrieval; information assessment and utilization; information and knowledge creation and preservation; and information sharing and exchange using various channels, formats and platforms. To be effective and fruitful, such interaction should be based on trustworthiness of governmental information; mutual respect and compliance with standards of ethics; and privacy and security. Though these essential competences are brought together by the concept of media and information literacy, no agenda has hitherto spotlighted the duty of using available R&D achievements to make open government more effective.

EU Data Protection Directive and personal data re-use

In recent times (April 2016[2]) the European Commission has issued a data protection Directive [10,11]. One of the improvements is the geographic coverage of the Directive. The new regulation will apply if the data controller or processor (organization) or the data subject (person) is based in the EU. Furthermore (and unlike the current Directive) the Regulation will also apply to organizations based outside the European Union if they process personal data of EU residents.

An additional interesting aspect is represented by the definition of "personal data". According to the European Commission "personal data" is any information relating to an individual, whether it relates to his or her private, professional or public life. It can be anything from a name, a photo, an email address, bank details, "posts" on social networking websites, medical information, or a computer's IP address. This is a relevant step forward in privacy issues. As clearly stated in the title of the Directive a specific focus concerns data re-use. Nowadays on line applications, APPs and open data represent the typical environment for data re-use.

What laws and legal implications may occur to "entities" re-using open data? This question pertains the problem we can summarise as "Transparency & Openness v/s Privacy, Security & Ownership". If we take into account a governmental organisation we can refer to ethics and Integrity within the organization. Usually speaking about governmental bodies we assign them high ethical standards, respect to dignity and organizational integrity.

Data re-users' main concern is rights and dignity of others. Majority of open data re-users are NGOs who often declare missions that are directly linked to rights of certain social

[2] DIRECTIVE (EU) 2016/680 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016, entered into force on 24 May 2016.

groups. Having responsible data policies send a clear signal to all stakeholders that organization does in fact care about its affected groups, especially those vulnerable.

More in general taking into account both governmental bodies and data re-users an additional aspect concerns reputation in front of donors, partners, and customers. Institutions and organisations having data re-use policies in place does send a clear signal to donors, partners, customers and other stakeholders that the organization treats its activities with care and high ethical standards.

Internet "prosumers" initiative: My data belongs to me

Concerns about data ownership and potential re-use do not only pertain international institutions or governments, it is an issues coming even from the grassroots. In 2014 the World Summit Award (WSA) [15], an organisation closely linked with WSIS grouping hundreds of "digital authors" coming from more than 170 countries around the world, launched "My data belongs to me" an initiative through its global multi-stakeholder network to push forward personal data ownership and big data issues at UN discussions. On the occasion of open discussions, such as the one held on he occasion of WSIS Forum 2014 in Geneva, the WSA invited participants to share views on issues with the current system of data use, the need for permission-based access, and steps for further action. This initiative underlines the consciousness about the ownership of personal information too many times shared among social platforms and business services.

Responsibilities in data re-use

Waiting for a sounding definition of data ownership it is worth to consider the responsibilities in data re-use. Re-using data organisations have the duty to ensure people's rights to: consent, privacy, security and ownership during the processes of: collection, analysis, storage, presentation and re-use.

Consent is a relevant "keyword", it means to explicitly provide the consent to use and manage private information provided in order to access a specific service. The request for "consent" must incorporate a clear and complete description of the use and aim of such data collection. Such a request may incorporate the description of future re-use of such dataset. If the potential use and re-use of data is articulated in different aims and steps the consent must be requested in the so called "granular" way that means that the platform will request a sequence of different consent that should be provided or not care of the citizen, in the field of APPs this is usually known as Warsaw Declaration on "appification of society" [5] (September, 2013).

How is usually ensured the right to consent? One of the typical approaches is "informed consent"; this is the mechanism through which people agree to provide information for research or data collection projects. The informed consent policy it is very well known in the medical sector, you read and sign the informed consent form before a surgical operation or a specific therapy but even more frequently when you apply to download eHealth APPs that will collect some physical parameters to perform their duties.

Informed consent find is basis on three components:

1. Disclosure of research objectives and any risks or negative consequences of participating capacity of individuals to understand the implications of participating voluntariness of their participation;

2. Informed consent includes plain language, easy-to-understand explanations of the types of data to be collected;

3. The purposes of collecting data, the intended and potential unintended uses of that data, who has access to and control over the data, risks of data leakage to third parties, and any benefits to participation in data collection.

Once data are collected and utilised for the specific proposes stated by the request for consent it might happen that the same data will be useful for different purposes how can we manage? Even if people use to think that once available data is re-usable without limitations, re-use of data collected for a different scope basically requires a new request for consent specifying the new purposes.

This is a real problem that affects major part of open data collected by public bodies and not only them. Imagine extending that same principle of specific consent to anything that anyone is able to “capture” regarding your life. Suddenly, you'd have to sign a legal release every time you swipe your credit card, take a taxi or walk through a store equipped with security cameras.

The question of who owns your data is not an easy one to solve. It becomes particularly problematic because you potentially create “public” data (whether or not it gets recorded) every time you leave your house entering “public” space. The number of steps you take, whether you look ahead or at the ground, what types of clothes you wear, and any number of decisions you make in view of other people are all potential data, this happens when airports security activate passenger’s shadowing or free Wi-Fi connections asking for your identity, e.g. typing your mobile phone number to gain access to the Internet, track your position.

This looking from the perspective of privacy but at the same time public institutions must respect the values of transparency and openness. The contraposition of such duties, transparency & openness versus privacy, security & ownership, finds its solution in the ethical and responsible re-use approach. This contraposition of duties may be schematized in a very effective way considering the right to privacy patrimony of those without “power”, while the need for transparency and openness is for those who have “power”.

So in extreme synthesis we have some principles: transparency & openness together with do no harm! The main concepts to be considered are: the right to consent and the respect of privacy, security & ownership. The concepts of privacy, security, commercial or state secrecy can be secured following the “do not harm” principle. Data re-users must do all within their powers to not cause any harm to any of the stakeholders that can rise as a direct or indirect result of open data re-use. To conclude if we consider the process from the data stages point we find: collection and storage, analysis and presentation.

The role of Privacy and risk related to breaches

Responsible and ethical data re-use is around the concept of privacy, legal requirements, risks and mitigations associated. Privacy is concerned with control over information, who can access it, and how it is used.

Privacy has many dimensions, from concerns about intrusive information collection, through to risks of exposure, increased insecurity or interference in their decisions that individuals or communities are subjected to when their 'private' information is widely known. Privacy is generally linked to individuals, families or community groups, and is a concept that is often used to demarcate a line between a 'private' and 'public' sphere.

Article 12 of the Universal Declaration on Human Rights [9] states "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation"

Let us take into account more closely privacy risks and their mitigation, key risks related to privacy are: disrespect to privacy can cause humiliation, embarrassment or anxiety for the individual, for example from a release of health data, it might be concluded that an individual accessed treatment for a sensitive sexual health condition; can have an impact on the employment or relationships of individuals; can affect decisions made about an individual or their ability to access services. This specific point might lead for instance to: their inability to obtain insurance; result in financial loss or detriment; can pose a risk to safety, such as identifying a victim of violence or a witness to a crime.

As usual when we have to deal with risks we analyse them in order to find mitigation actions. Let us start taking into account a basic privacy risk assessment determining any specific unique identifying variables, such as name, cross-tabulation of other variables to determine unique combinations that may enable a person to be identified, such as a combination of age, income, and postcode. In addition acquiring knowledge of other publicly available datasets and information that could be used for list matching. Of course this procedure will not ensure 100% privacy because new data sources might be open to public access completing the puzzle. As an example think about the typical concerns related to some on line personal feedback or better on line vote, how to ensure single vote from right holder citizen and at the same time disjoin his/her identity from the expressed vote.

Risk assessment: mapping

We all know that security and privacy are subject to risk as already stated thus it is important to identify and mitigate risks associated with privacy and security concerns. In order to reach this goal, as a first approach, we can perform the following steps: identify the persons at risk in the event of personal information exposure (not restricted to the data owner or collector), identify knowledge assets that can be extracted from the data collected (discrete data points, meta analysis of data points, mash up of the collected data and external data sources); evaluate the importance of each knowledge asset to the potential goals/harms (little or no relevance, significant relevance, crucial). This approach, many times, will lead us to identify the crucial nodes that, if adequately protected, will ensure no harm. The level of privacy risk will be dependent on the likelihood that identification could occur from the release of the data and the consequences of such a release. Anyway mitigation is many times linked to de-identification.

Aspects connected to Security

In the previous paragraph we mentioned non-only privacy but even security. Security is somewhat linked to privacy, adapt security protocols and tactics to encompass:

1) Digital information security;

2) Physical and operational security;

3) Psychosocial well being required for good security implementation.

Nowadays the key concept is “holistic security”, a “global” approach to security integrating all the different aspects and problems. A specific interest is devoted to digital security.

Digital security is more than focus on software or tools; integrating emotional well-being, personal and organizational security. Good implementation of digital security tools and tactics requires attending to the practitioners’ psychosocial capacities to recognize and respond dynamically to different threats to themselves and to participants related to project data collection and communications (intimidation, social engineering.)

Open Source Intelligence (OSINT)

Lastly let us consider a particular use of the information gathered, OSINT [18,19] is the acronym of Open Source Intelligence and refers to intelligence collected from publicly available sources. In the intelligence community, the term Open indicates overt and publicly available sources, in opposition to covert or clandestine sources and it is not related to open-source software. It is important to notice that OSINT is distinguished from research; it applies the process of intelligence to create tailored knowledge supportive of a specific decision by a specific individual or group.

OSINT includes a wide variety of information and sources:

Media: newspapers, magazines, radio, television, and computer-based information.

Web: web sites, web communities, user-generate contents, video sharing sites and blogs.

Metasource Engines: MetaCrawler, Ixquick, Dog Pile, etc.

Deep Web: no index web sites, reserved information and illegal contents.

Social Networks: Facebook, Twitter, LinkedIn, Instagram, etc.

Software OSINT: Foca, Maltego, Shodan, etc.

We already took into account Social Engineering that of course represents a relevant risk no matter how good is cyber security, the weakest link of the security chain are humans.

Conclusions

Arguably, we haven't even discovered every type of data that can be recorded. At the same time today we have only a limited idea and vision on potential risks due to “data leaks”, in some way we are still in the digital Middle Ages both for positive outreaches and drawbacks. Anyway back to “my data” until the legal infrastructure changes, though, none of that will change this one simple fact: you don't "own" data just because it's about you.

References

1. United Nations Manual on the prevention and control of computer-related crime, UN 2001
2. Chris Babel, Tackling Privacy Concerns Is Key to Expanding the Internet of Things, Wired Innovation Insights, Feb 2015
3. Viktor Mayer-Schönberger, Delete: The Virtue of Forgetting in the Digital Age, ISBN-13: 978-0691138619, Princeton University Press 2009
4. Herbert H. Thompson, "How I Stole Someone's Identity", Scientific American, August 2008
5. Warsaw Declaration, http://www.coe.int/t/dcr/summit/20050517_decl_varsovie_EN.asp ,Council of Europe Warsaw Summit May 2005
6. UK Government Service Design Manual: Open Data https://www.gov.uk/service-manual/technology/open-data.html
7. Daniel Burrus, Who Owns Your Data?, https://www.wired.com/insights/2014/02/owns-data/
8. Barb Darrow, The Question of Who Owns the Data Is About to Get a Lot Trickier, Fortune, http://fortune.com/2016/04/06/who-owns-the-data/
9. Universal Declaration of Human Rights, http://www.un.org/en/universal-declaration-human-rights/
10. Protection of personal data in EU, http://ec.europa.eu/justice/data-protection/
11. Regulation (Eu) 2016/679 of the European Parliament and of the Council of 27 April 2016, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC
12. Google - Privacy & Terms, https://www.google.com/intl/en/policies/privacy/
13. Merriam Webster: Ethic, http://www.merriam-webster.com/dictionary/ethic
14. BBC Ethics Guide, http://www.bbc.co.uk/ethics/introduction/intro_1.shtml
15. My data belongs to me, http://wsa-mobile.org/news/my-data-belongs-me-wsa-roundtable-discussion-personal-data-virtualisation-society-wsis10-review
16. UNESCO and WSIS, Ethical dimensions of the Information Society (C10), http://www.unesco.org/new/en/communication-and-information/unesco-and-wsis/implementation-and-follow-up/unesco-and-wsis-action-lines/c10-ethical-dimension-of-the-information-society/
17. Information for All Programme (IFAP), Information Ethics, http://www.unesco.org/new/en/communication-and-information/intergovernmental-programmes/information-for-all-programme-ifap/priorities/information-ethics/
18. Randolph Hock, Internet Tools and Resources for Open Source Intelligence – OSINT, http://www.onstrat.com/osint/
19. Central Intelligence Agency, INTellingence: Open Source Intelligence, https://www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/open-source-intelligence.html