Forgetting personal data and revoking consent under the GDPR: Challenges and Proposed Solutions

Eugenia Politou, Efthimios Alepis and Constantinos Patsakis

Department of Informatics, University of Piraeus, Piraeus, Greece
Email: {epolitou,talepis,kpatsak}@unipi.gr

Upon the GDPR's application on 25 May 2018 across the European Union, new legal requirements for the protection of personal data will be enforced for data controllers operating within the EU territory. While the principles encompassed by the GDPR were mostly welcomed, two of them; namely the right to withdraw consent and the right to be forgotten, caused prolonged controversy among privacy scholars, human rights advocates and business world due to their pivotal impact on the way personal data would be handled under the new legal provisions and the drastic consequences of enforcing these new requirements in the era of big data and internet of things. In this work, we firstly review all controversies around the new stringent definitions of consent revocation and the right to be forgotten in reference to their implementation impact on privacy and personal data protection, and secondly, we evaluate existing methods, architectures and state-of-the-art technologies in terms of fulfilling the technical practicalities for the implementation and effective integration of the new requirements into current computing infrastructures. The latter allow us to argue that such enforcement is indeed feasible provided that implementation guidelines and low-level business specifications are put in place in a clear and cross-platform manner in order to cater for all possible exceptions and complexities.

Keywords: GDPR, Privacy, The Right to be Forgotten, Data Protection

Received 23 June 2017; revised 03 November 2017

1. INTRODUCTION

On 27 April 2016, after four years of drafting, lobbying and negotiations among the EU Member States and many affected organizations [1][2][3], the EU General Data Protection Regulation (GDPR) has been agreed and finalized, whereas on 4 May 2016 its final text published in the Official Journal of the European Union [1]. Following a two year implementation period, the GDPR will be applied across the European Union from 25 May 2018.

The GDPR's introduction aimed at replacing the Data Protection Directive 95/46/EC (DPD) [2] introduced in 1995 and, being a directive, left some room for interpretation during its transposition into individual national laws. In addition, the rapid change in data landscape caused by the explosion of ubiquitous and mobile computing and the big data era, had led to the necessity for another update to the regulatory environment within the EU. Yet, the radical changes brought in by the GDPR are impacting severely businesses operating within and outside the EU territory. Most importantly, as a regulation and not a directive, it will immediately become an enforceable law in all Member States and hence, it will contribute to the harmonization of current data protection laws across the EU, enhancing at the same time both data protection rights and business opportunities in the digital single market.

The regulation accomplishes its objectives, on the one hand by strengthening the well established data protection principles already specified in DPD, like consent and purpose limitation, and on the other, by encompassing new principles such as the right to be forgotten, the right to data portability, the obligation for data protection impact assessments, and privacy by design, among others. Since its first draft in 2012,

[1] http://www.eugdpr.org/gdpr-timeline.html

[2] https://edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en

[3] http://www.jimmcguigan.co.uk/EJCNews_July2013DATA\%20protection_bothstories.pdf

much debate has been taken place among scholars and law experts about the fundamental changes that introduces. Two particular GDPR principles, however, rocked the boat of legal, academic and business world: the reintroduced concept of consent along with its revocation as well as the newly introduced right to be forgotten (RtbF). They both caused prolonged controversy due to their pivotal impact on the way personal data would be handled under the new legal provisions and the drastic consequences of enforcing these new requirements in the era of big data and the Internet of Things (IoT). In this respect, the purpose of this work is twofold: firstly to review all controversies around the new stringent definitions of consent revocation and the RtbF in reference to their implementation impact on privacy and personal data protection, and secondly, to evaluate existing methods, architectures and state-of-the-art technologies in terms of fulfilling the technical practicalities for the implementation and effective integration of the new requirements into current computing infrastructures.

The rest of this work is structured as follows. In Section 2 the notions of privacy and data protection are discussed. Then, in Section 3, after a short presentation of the GDPR, the changes imposed by the new definitions of consent (3.1) and the RtbF (3.2) are analysed in terms of their theoretical and practical approaches in the academic literature (3.1.1-3.1.2 and 3.2.1-3.2.3 respectively), while already proposed technical solutions that fit to these new privacy requirements are discussed and evaluated (sections 3.1.3 and 3.2.4 respectively). Section 4 concludes the paper by discussing the future of privacy in the GDPR era.

2. PRIVACY AND PERSONAL DATA PROTECTION

Privacy and personal data protection are two interrelated terms that are often used interchangeably but they actually constitute two discrete and different notions. The idea of privacy in Europe derives from concepts such as human dignity and the rule of law. Modern conceptions of privacy have begun to be developed following the experiences of fascism in World War II and communism in the post-war period. Within European law there is a distinction between "privacy" and "data protection" which defines these two concepts as closely related, and often overlapping each other, but not as synonymous [3]. Privacy generally refers to the protection of an individual's "personal space", while data protection refers to limitations or conditions on the processing of data relating to an identifiable individual. Nevertheless, as legal scholars note [4, 5], data protection and privacy overlap on a way whereby data protection is both broader and narrower than privacy. It is narrower because it only deals with the processing of personal data, whereas the scope of privacy is wider. It is broader, however, because it applies to the processing of personal data, even if the latter does not infringe upon privacy.

Legally speaking, privacy and data protection both represent two distinct fundamental rights under the European Law, which defines the first one as a substantive right whereas the second as a procedural [6]. The stand-alone fundamental right to data protection was declared for the first time under the Article 8 of the Charter of Fundamental Rights of the European Union, enacted by Lisbon treaty [7]. It has been pointed out that the principles underpinning the human right to data protection reflect some key values inherent in the European legal order, namely privacy, transparency, autonomy and non-discrimination [8]. Therefore, under an instrumental conception, it can be argued that the right to data protection could serve as a safeguard not only for privacy but also for all fundamental rights [4]. Additionally, the right to privacy is also a well-established right by the European Convention on Human Rights (article 8) [9].

As for the GDPR, it is expressly framed in terms of rights, with Article 1 noting that the regulation "*protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data*". Despite the fact that there is not any reference to the right to privacy throughout the GDPR text, the concept of privacy is implied in most of its recitals and articles.

In what follows, we will analyse what the right to privacy and to personal data protection entails.

2.1. Privacy

Privacy is a notion that even though it has been introduced as a right in 1890 by Warren and Brandeis [10], it was only the last three decades that it has been extensively discussed in its various forms and contexts, mainly due to the blow of computing and informational sciences. As Introna noted back in 1997 [11], privacy has emerged as a philosophical issue in the late 1960 and since then is discussed in great controversy among philosophical, legal, social and science circles. Still, no universally accepted definition of privacy exists. Privacy can be seen either as a right to be left alone [10], the "*power to selectively reveal oneself to the world*" [12], or as control over personal information or even as a freedom from judgement by others [11]. Post explains [13] that "*privacy is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether it can be usefully addressed at all*". According to Solove [14], the term "privacy" is an umbrella term, referring to a wide and disparate group of related things and cannot be understood independently from society since privacy, in its core, is a social artefact and without the context of society there would be no need for privacy.

In the same spirit but more recently, Ohm argues

[15] that the four recent trends of our high tech society, namely smartphone, cloud, social networks and big data, taken together, enable the rise of a powerful, new surveillance society which raises significant new threats to privacy, such as location tracking. Although the concept of surveillance society and its impact on privacy has appeared in many non-fiction books even before the big data and smartphone era [16, 17], there are the last technological advancements and the prevalence of pervasive and ubiquitous computing that triggered massive scepticism and worldwide dispute around the notion of privacy. A survey conducted some years ago among US adults[4] found that the majority of adults feel their privacy is being challenged and showed that people give important weight to the idea that privacy applies to personal rights and information, whereas 91% ofadults in the survey agreed that consumers have lost control over how personal information is collected and used by companies. Paradoxically enough, although in earlier times control over personal data may have been best undertaken by preventing the data from being disclosed, in an internet enabled society users are showing an increased demand for more data collection, which illustrates that they do not necessarily want more privacy as if concealment, but primarily want more control and transparency on the way their data are being used and reused [18, 19].

2.2. Personal data, big data and privacy

In general, personal data refer to the information relating to an individual. While many Data Protection Acts define personal data in more or less similar terms [2, 20, 21], the GDPR elaborates a little further on their definition (Article 4) [1]:

> "Personal data" means any information relating to an identified or identifiable natural person ("data subject"); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

This definition clarifies that personal data are any information that can be used on its own or with other information to identify, contact, or locate an individual.

2.2.1. The value of personal data

It was Clive Humby who first coined the parallelism of data to oil back in 2006 to denote that data are just like crude: *"It's valuable, but if unrefined it cannot really be used"*[5]. The World Economic Forum [22, 23] has also described personal data as a new asset class for which a complex ecosystem of entities collecting, analysing, and trading personal information has emerged. Spiekermann et al. underline in [23] that personal data are seen as a new asset due to their potential for creating added value for companies and consumers by providing services hardly imaginable without it. As Acquisti et al. describe [24], personal information has both private and commercial value and often exploiting its commercial value entails a reduction in privacy and sometimes even in social welfare overall. Stated otherwise, most privacy issues originate from the two different markets, the market for personal information and the market for privacy which are actually two sides of the same coin [24]. Solove describes in his taxonomy of privacy [14] that *all methods and practices adopted nowadays by personal data markets, such as aggregation of personal data, increased accessibility, re-identification, secondary use, exclusion, and decisional interference, constitute privacy breaches.*

2.2.2. Personal data and privacy threats

Re-identification in particular, by using and combining various available sources of information, not necessary personal ones, has been characterized as one of the major privacy threats in our modern data driven society. In [25] it is explained how any information that distinguishes one person from another, like consumption preferences or call usage patterns, can be used for re-identifying anonymous data (thus re-anonymising them). Many studies on mobile privacy have revealed privacy threats in disclosing personal information. For example in [26] it was shown that the majority of applications leaked the device ID, which can provide detailed information about the habits of a user. Plus, there is always the possibility that additional data are used to tie a device ID to a person, increasing the privacy risks. Sweeney, the pioneer of re-anonymization techniques, proved many years ago that 87% of the population in the USA had reported characteristics that likely made them unique based only on ZIP, gender and date of birth [27]. More recently, researchers uncovered the identities of sample donors using free, publicly accessible internet resources like recreational genetic genealogy databases [28]. Despite the fact that a malicious adversary can use personally identifiable information, such as a name or social security number, to link data to identity, the adversary can do the same using information that nobody would classify as personally identifiable. For example, Narayanan and Shmatikov [29] demonstrated that the re-identification of individuals is possible based on anonymous film ratings of 500 000 subscribers of Netflix. These developments led to the *failure of anonymization*, as

[4] http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/

[5] http://ana.blogs.com/maestros/2006/11/data_is_the_new.html

it has been described by Ohm [30], and along with big tech giants' data aggressive policies, eliminated long time ago any confidence that personal data are safe and protected. For example, in 2016 Google, the biggest online search engine, quietly changed its privacy policy to allow individuals browsing habits to be combined with what the company learns from the use of gmail and other tools[6], a fact that has been criticized by many Consumer Protection boards as the genesis of the "super-profiles" that enable advertisers to relate personally identifiable user information with an individual's online history[7]. This phenomenon has been further amplified recently by consolidating off-line credit card transactions as well[8].

2.2.3. Ubiquitous computing and big data

The spread of ubiquitous and pervasive computing, which continuously collects huge amount of personal data from online activities and mobile devices, intensify additionally the threat for re-identification. Ubiquitous devices such as smartphones and wearable badges, by utilising a powerful set of sensors and utilities, can monitor biometric signals or location data of their holders in order to provide healthcare interventions or customized driving directions respectively, and thus assisting users in their daily tasks. In recent years, the research area of affective computing has been also developed, a field that infers people's emotions, traits and behaviours by exploiting intelligent machine learning methods and data acquired through mobile and wearable devices [31, 32]. All these emerging areas of pervasive applications characterize the boom of big data era, where sheer volume of data are collected and processed to promote machine's intelligence by learning via example.

As fascinating as this research area may sound in terms of its technological and scientific achievements, it conceals severe implications and risks to human civil rights and specifically to human privacy and data protection rights [33]. Inevitably, privacy and ethical issues arising from the use of current pervasive applications and big data exploitation have been discussed in many scientific papers [34, 35, 31]. As O'Hara has expressively illustrated in [36] "*had the government demanded that we all carry around electronic devices that broadcast our whereabouts to a central database, that the information should be stored there indefinitely...there would have been an outcry. But in the real world most if not all of us carry such devices around voluntarily, in the shape of our mobile phones*". For instance, the vast amount of personal data that are either publicly available in social networks or can be easily extracted through the continuous personal traces left behind when one is surfing the web or using ubiquitous devices, can be collected and exploited for profiling and marketing purposes or even experimental research. Except the monetary exploitation, the processing of this sheer volume of personal data may raise concerns about the conditions under which they were collected, processed and disseminated. Apparently, privacy and big data are in many cases contradictory. Big data require massive amount of information to be collected with not a predefined and clear purpose at the time of collection. Users do not have any control on their personal information stored and analysed by the involved data controllers and the parties that participate in data dissemination may be numerous [31, 35, 37]. As Narayanan and Shmatikov state in [25], *big data guarantees the key precondition for achieving re-identification, namely the corresponding data attributes to be sufficiently numerous and fine-grained in way that no two people are similar, except with a small probability.*

2.2.4. Personal data protection

To mitigate the harms from processing personal data and the consequences in individual privacy, many workable methods and technical solutions have been introduced over the last decades. The most distinguished among them are k-anonymity [38], l-diversity [39], t-closeness [40], differential privacy [41], data aggregation [42], and data obfuscation [43]. Although each of these methods may be appropriate to per case approach, the general concepts underpinning modern privacy aware systems are based on the *"privacy by design"* principles which, despite their long history of incorporation into privacy preserved systems [44], have been formulated into concrete applicable design principles by Cavoukian in 2011 [45]. Privacy by design principles encapsulates concepts such as data minimization, purpose limitation, transparency and control, all anticipated by the data protection regulations and the GDPR subsequently. However, big data characteristics by their very nature go against these principles [35, 46]. Under data minimization and purpose limitation organizations are required to limit the collection of personal data to the minimum extent necessary to obtain their legitimate goals and to delete data that is no longer used for the purposes for which they were collected. On the contrary, big data business model encourages collection of more data for longer periods of time [47].

In view of the above, the concept of building trust is constantly under stake when most users today are not even aware of the data processing procedures undertaken by businesses with their personal data. As Spiekermann et al. underline in [23]: "*If they* [users]

[6] http://njtoday.net/2016/11/17/google-quietly-dropped-ban-personally-identifiable-web-tracking/

[7] http://www.csmonitor.com/Technology/2016/1220/Privacy-groups-Serial-offender-Google-deceived-consumers-with-2016-policy-change

[8] http://www.zdnet.com/article/google-well-track-your-offline-credit-card-use-to-show-that-online-ads-work/

learned about today's volume and business done with their data among third parties, they may be surprised and feel betrayed. No matter whether and to what extent first party companies have engaged in data deals themselves, they could all be hit by a backlash from users once they find out". This adverse reaction has also pointed out by Mittelstadt et al. [35] who argue that "*the tension between personal big data and privacy often triggers a "whiplash effect", by which overly restrictive measures (especially legislation and policies) are proposed in reaction to perceived harms, which overreact in order to re-establish the primacy of threatened values, such as privacy*".

For many big data enthusiasts and privacy sceptics the GDPR constitutes an emergent "backlash", an overwhelming reaction from regulators to the bursting exploitation of personal data dominating not only the way industry performs business but academia conducts research as well. Indeed, privacy concerns have split a large share of academia as privacy often contradicts modern research practices. For example, a large proportion of research community urges for loosened privacy regulation and increased trust on the research ethics arguing that the fact researchers can identify individuals and all of their actions is a necessary trade-off for high quality research [48] while others argue that the regulation does not fully grapple with the challenges posed by big data and a way forward would be the experimentation with a more flexible approach to regulation through the creative use of codes of conduct [49]. All these arguments against strict privacy regulations are based on the inevitably reality that data utility decreases when privacy increases, a fact that urges data driven business world to warn against the overly broad regulatory definitions of personal data and to highlight that regulations on data protection and privacy may preclude economic and societal benefits [50]. Stated otherwise by Ohm, *no useful database can ever be perfectly anonymous* [30]. Notwithstanding this clash, most scholars agree that there cannot exist big data without privacy since the protection of personal data is, first of all, in the interest of the big data analytics service providers who will ultimately have to cope with this challenge [51].

The GDPR, taking into account both risks and challenges that big data may bring upon citizens, introduced the new legal term of pseudonymization (Article 4(5)) in order to describe data that could be attributed to a natural person by the use of additional information, which must be kept separately and be subject to technical and organizational measures to ensure non-attribution. While the use of pseudonymization is encouraged in many occasions, pseudonymized information is still considered a form of personal data and hence, a value to protect.

Next, the basic data protection principles of the GDPR, and mainly the two newly introduced rights, will be elaborated.

3. GDPR DATA PROTECTION PRINCIPLES

According to many legal scholars, the most important contribution to EU personal data processing by the GDPR is the choice of the instrument itself, since the moderation of EU data protection through a regulation, rather than a Directive, constitutes a turning point for EU signalling a forced exit of this particular field of law from Member State level to EU level [52]. Nevertheless, disappointing many privacy advocates, the final version of the GDPR still has a large number of provisions that leave room for national interpretations and approaches depending on the culture, focus and priorities of the supervising authorities.

The main data protection principles in the GDPR are revised but are broadly similar to the principles set out in the DPD: fairness, lawfulness and transparency (Article 5(1)(a)); purpose limitation (Article 5(1)(b)); data minimization (Article 5(1)(c)); accuracy (Article 5(1)(d)); storage limitation (Article 5(1)(e)), accountability (Article 5(2)), integrity and confidentiality (Article 5(1)(f)). While the DPD constituted the international standard against which all data protection initiatives, in and out of Europe, were judged [52], the GDPR brings the novelty of explicitly imposing organizations to enshrine "*data protection by design and by default*" (Article 25) enforcing measures such as data minimization as a standard approach to data collection and use. Furthermore, the GDPR extends the provision on automated individual decision-making, to include profiling cases as a prime example of enabling individuals to control their personal data in the context of automated decision-making (Article 22) and hence acts as crucial function for mitigating the risks of big data and automated decision making for individual rights and freedoms.

Besides the above, the regulation introduces some new rights for data subjects, like the right to data portability (Article 20) which ensures interoperability of subject's data and requires data to be provided in a structured, commonly used and machine-readable format and, when required, the controller to transmit the data directly to another controller. While the portability right seems reasonable and has been welcomed by the majority of public and private organizations, there are some other rights that have raised great concerns and stipulated long debates between scholars within law, privacy and ethics disciplines. Specifically, the GDPR introduces the *right to withdraw consent* (Article 7(3)) and the *right to be forgotten (RtbF)* (Article 17), two secondary rights that derive from fundamental concepts of data protection. These two controversial rights will be extensively discussed hereafter and potential frameworks and methods will be evaluated against their feasible implementation.

3.1. Consent and revocation

Consent aims at providing legitimate grounds to data controllers for collecting, processing or even disseminating personal data for secondary use. While consenting is one among several available legal grounds to process personal data under all data protection regulations up to date, is undoubtedly the most global standard of legitimacy and most likely to engender user trust [46]. Even though consent may have various forms with similar flavours, such as informed, explicit, unambiguous or broad, each of these forms is quite diverse in nature and their use have been intensively debated for their utilization in online environments and research projects.

Informed consent can be said to have been given based upon a clear appreciation and understanding of the facts, implications, and consequences of an action. Flashing back, informed consent was a cornerstone of the Nuremberg Code ethical guidelines originated in the pre-World War II Germany and specify that informed consent is not only essential for safety, protection and respect for participants, but also for the integrity of research itself [53]. As explained by Reynolds in [54], *"to be informed, consent must be given by persons who are competent to consent, have consented voluntarily, are fully informed about the research, and have comprehended what they have been told"*. Depending on the methodology, the population, the topic under study and the level of risk, informed consent may be implied or explicit, active or passive, and written or oral. For obtaining an explicit consent participants should give consent through an explicit affirmative action, such as by answering a specific question, in written or oral form, about their willingness to participate. On the other side, broad consent involves agreeing to a broad set of potential secondary future uses under a particular governance framework and has been widely adopted as the standard practice in many genetic registries and biobanks. Broad consent is also a standard practice for most big data projects where their most innovative secondary uses can't be imagined by the time of data collection.

The academic discussions on whether the user consent in online research and marketing should be informed and explicit [55, 56, 57, 58] or broad [59, 60] is heated and the relevant literature is split, while many academics have argued in both ways [61, 62] or in favour of additional countermeasures [63, 64, 65]. Meanwhile, other conceptions of consent have been also proposed, like the collaborative consent [61, 66], the dynamic consent model [67, 65] which is actually a tool that could better facilitate the process of obtaining any form of consent, and recently, the notion of meta-consent [68].

Another rising tension for the use of consent comes from the potential benefits of big data analysis and the need for explicit or informed consent [69]. Edwards in her notable work [46] examines the issue of obtaining meaningful prior consent in the era of IoT, big data and the cloud, especially when data are collected in public, as in the context of "smart cities". As Barocas and Nissenbaum have been intelligibly expressed [69], *big data extinguishes what little hope remains for the notice and choice regime* since upfront notice is not possible in case the value of personal information is not apparent at the time of collection when consent is normally given. Let alone that new classes of goods and services usually reside in future and unanticipated uses [70, 50, 35]. This motivated many radical voices to argue against the need for consent which may jeopardize innovation and beneficial societal advances [47] and therefore, its role should be circumscribed with respect to prospective data uses and, in specific cases, consent should not be required to legitimize data use. Still, for other scholars [52, 35] consent requirements are the last defence for individuals against the loss of control on their personal information processing and thus, eliminating or reducing the need for informed consent cannot be accepted uncritically and seemingly without public debate, particularly if democratic ideals are valued.

The notion of consent revocation, or withdrawal, has also been brought into light recently, with many to argue for a right to revoke consent and for a more user friendly and personalized consent mechanism [71, 72]. Indeed, when individuals' are given the opportunity to grant consent to the use of their personal information as a primary mean for exercising their autonomy and to protect their privacy, it should be logical to exist a corresponding option to withdraw or revoke that consent, or to make subsequent changes to that consent [73, 18]. The principle of consent withdrawal within the Human Computer Interaction (HCI) context has been studied in many ethical research projects, with Benford et al. in [74] to underline that in many cases it may be difficult to fully withdraw in practice because the issue of balancing consent, withdrawal and privacy is a very demanding managed task. Whitley in [18] argues further that, since the revocation of consent can mean a variety of different things depending on the circumstances and constitutive purposes that the data are being held for, it is helpful to differentiate between revoking *the right to hold* personal data and revoking *the right to use* personal data for particular purposes. Revoking the right to hold might be implemented by marking a particular record as no longer "being live" or may require the deletion of records and, in extreme cases, it might require deleting data from backups and physically grinding the hard disks. In addition, providing auditable, privacy friendly proof of compliance when and how the revocation has been achieved is a challenge both technologically and legally [18]. For instance, the advancements towards privacy-enabled networks and infrastructures puzzles some academics [75] who afraid that the same mechanisms have been put in place to protect the

privacy of data (like de-identification) may actually make it very difficult to trace and remove individual derived data in order to allow participants to withdraw completely their consent and be forgotten. In such situations, as Kaye [75] underscores, it may be only possible to prohibit the entry of new information and samples into the system. Apart from these practical difficulties, there are also economic and public-good arguments for disallowing absolute withdrawal. For instance, in the bio-banking field complete withdrawal could lead to the wastage of resources invested in bio-repositories [75, 76] whereas the practice of archiving qualitative research data for substantive secondary analysis can be significantly challenged under the revocation mechanism for withdrawing consent [77]. Due to these immense consequences, many academics and legal experts questioning the concept of consent withdrawal.

3.1.1. Consent misuses

As Mittelstadt and Floridi argue in [35], despite the fact that informed consent has been used widely within the scientific and biomedical research domain, there exist plenty of issues raised by the collection and analysis of data from potentially "unwilling" participants in other application domains, for example data scraped from social media platforms, smartphone applications, or open web forums, which provide massive amount of personal data.

Lately, cases of bad practices regarding obtaining consent have been observed extensively. One such example is the contagion study commenced by Facebook [78] where the company manipulated users by changing their newsfeed to investigate whether the emotional state of users could be influenced by the words of other users – a form of "emotional contagion" which takes place unbeknown to the user. Therefore, the study provoked extended criticism of the Facebook research practices and resulted in its characterization as "*the best human research lab ever*"[9] [79, 80], even though the company publicly acknowledged and apologized for its fault[10]. It should be noted that in the study no user consent was ever required on the grounds that users were already given broad consent when they signed in to use the social network.

Facebook has seemingly a long history for conducting online research with its users' personal data without obtaining explicit or direct consent. Back in 2010, a company's experiment with 61 million users [81] resulted in changing the real-world voting behaviour of millions of people for the US midterm elections. As Gleibs emphasizes in [53], even though the statistical effect of the manipulation was small, their intervention might have had the potential to change the outcome of the Congressional elections in 2010, and although the study influenced behaviour with good intentions, the techniques employed could be used to influence political protest or anti-democratic behaviour in countries with little democratic traditions [53]. In the study, no consent was obtained from the participants on the grounds that the experiment was not intrusive to people's lives, it bore minimal risk to the participants, it didn't affect their rights, and the research couldn't had been possible otherwise [53].

In another research study [82], a group of Danish researchers publicly released a dataset of nearly 70 000 users of the online dating site OkCupid, including usernames, age, gender, location, relationship (or sex) preference, personality traits, and answers to thousands of profiling questions used by the site[11]. Researchers excused themselves for not obtaining users consent by stating that the data were already public. Nevertheless, as Michael Zimmer explains in [83]: "*just because personal information is made available in some fashion on a social network, does not mean it is fair game to capture and release to all*".

Although the number of research projects not obtaining consent for exploiting personal data is quite large, the failures of the past (e.g. Harvard's discontinued sociology research project using Facebook sensitive data without consent and compromising participants' privacy [84]) seem to have alarmed the Institutional Review Boards for approving and engaging in the complexities of the research on social networks [83]. Aside research however, and for offering suicidal users a second chance, Samaritans, a leading suicide prevention charity, launched few years ago the Radar App[12], an app designed to tell Twitter users which of the people they follow might be feeling low by using an algorithm to identify key words and phrases in their tweets that indicated distress or a mentally vulnerable state and notifying their followers accordingly. The app provoked mass media uproar and wide criticism due to the data protection and privacy issues raised as it was allowing the sharing of personal information with other untrusted people without the subject's knowledge or consent[13]. Finally, few weeks after its launch the app was permanently suspended[14].

In the medical domain, and in the NHS UK in particular [85], the use of patient data without offering clear, specific, free and informed consent, not even unambiguous and effective opt-outs, while misleading about the level of anonymization of their data and the likelihood of re-identification with the argument that

[9] https://www.forbes.com/sites/kashmirhill/2014/06/28/facebook-manipulated-689003-users-emotions-for-science/#197c5800197c

[10] http://www.cbsnews.com/news/researcher-apologizes-for-facebook-study-in-emotional-manipulation/

[11] https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science/

[12] https://www.samaritans.org/how-we-can-help-you/supporting-someone-online/samaritans-radar

[13] https://www.theguardian.com/society/2014/nov/07/samaritans-radar-app-suicide-watch-privacy-twitter-users

[14] https://www.samaritans.org/how-we-can-help-you/supporting-someone-online/samaritans-radar#10mar

research is part of their "care", has been considered as a stretching of the law. Ultimately, these practices led to the closure of the UK national program Care.data which aimed at the integration of all patient data in a single platform[15].

The above examples are just a small fragment illustrating the chaos and uncertainty that dominate industry, academic and public institutions in obtaining and revoking consent for personal data use in ubiquitous computing systems and big data analytics. Although this gap is usually expected to be addressed by ethical guidelines and policies, it is the terrestrial legislation the one that will enforce common handling against various interpretations of national policies. Therefore, the long awaited GDPR regulation had raised great expectations in dealing with such sensitive issues.

3.1.2. Consent under the GDPR

Since the early years of the GDPR introduction, the consent obligations for research, imposed by the regulation, have been extensively criticized by the academic and medical community [86, 87, 88, 89]. Upon its final publication in 2016, extensive analysis has been conducted by law scholars, not only in Europe but worldwide, and most of the hesitations about compromising research were dropped given the "research exemption" anticipated by the regulation (Recital 33). Moreover, the GDPR specifies derogations for research without consent in cases of medical research conducted "in the public interest" or for compliance with legal obligations (Recital 51). Nevertheless, when consent is to be used, consent presumed by failure to opt-out or by change pre-ticked boxes will no longer be permitted because consent should need to be provided by a *"clear, affirmative action"* (Article 4.11) [90].

Besides these derogations, overall the GDPR creates additional hurdles for consent over what was required by the directive[16]. Particularly, the conditions for obtaining consent under the GDPR have become stricter since consent has to be, not only informed and specific, but unambiguous as well. Whereas the earliest drafts proposed by the European Commission specifically had introduced the requirement of "explicit" consent for processing all kinds of personal data, the final document clarifies that explicit consent is required only for processing sensitive personal data (Article 9(2)). For non-sensitive data, however, a freely given, specific, informed and unambiguous consent will do and this allows the possibility of implied consent if an individual's actions are sufficiently indicative of their agreement to processing [91]. Additionally, consent has to be easily withdrawn and not to be assumed from inaction. Inevitably, these requirements translate to the amendment of many current data protection notices. However, it has been underscored [52] that the GDPR document evidently constitutes the next-best option in order to warrant a significant level of protection as it appears relevant with the contemporary processing needs. On the other hand, since data that do not pertain to natural persons are beyond the scope of the GDPR, it is argued that it fails to protect individuals in the case of automated algorithmic decisions that do not target individuals but affect their lives [33]. This is the case of "tyranny of the minority", a term introduced by Barocas and Nissenbaum [69] that describes the choice forced upon the majority of the population by a consenting minority who will to disclose information about themselves but this information may implicate others who happen to share the more easily observable traits that correlate with the traits disclosed. As they exmplain, *"the value of a particular individual's withheld consent diminishes the more effectively one can draw inferences from the set of people that do consent, when this set approaches a representative sample. Once a dataset reaches this threshold, analysts can rely on readily observable data to draw probabilistic inferences about an individual, rather than seeking consent to obtain these details"*.

Notwithstanding this deficiency, the GDPR anticipates for a right to withdraw (revoke) consent, a fact that has been warmly applauded since this explicit reference to the right to withdraw consent was missing from the DPD [73, 92]. Under this right, the data subject has the right to withdraw consent at any time, but the revocation is foreseen only for future processing of personal data and therefore the data controller should not use his data for future assessments and processing, i.e. the revocation is not retroactive, meaning that it does not apply for processing that had taken place before withdrawal: "*The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal*" (Article 7(3)). Hence, this notion of non-retroactive revocation is not affected by the progress of informational privacy infrastructures and neither devaluates already conducted research.

This also complies with the Opinion on the definition of consent [93] published in 2011 by the Article 29 Working Party[17], which specifies that withdrawal is exercised for the future, not for the data processing that took place in the past, in the period during which the data was collected legitimately. Decisions or processes previously taken on the basis of this information can therefore not be simply annulled. However, *if there is no other legal basis justifying the further storage of*

[15] http://www.nationalhealthexecutive.com/Health-Care-News/nhs-england-to-close-caredata-programme-following-caldicott-review

[16] https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-3-consent/

[17] The Article 29 Working Party, set up under Article 29 of Directive 95/46/EC (DPD), is an independent European advisory body on data protection and privacy bringing together the European Union's national data protection authorities. As from 2018, under the newly adopted GDPR, the Article 29 Working Party (WP29) will be transitioned into a new legal framework, the European Data Protection Board (EDPB).

the data, they should be deleted by the data controller. Article 7, however, does leave open for interpretation whether this provision about consent affects, apart the processing – which does not–, the storage of the data themselves on which the withdrawal applies, and therefore it does not clarify if it requires the erasure of the data upon their revocation of consent under which they were first collected. Nevertheless, following the introduction of the RtbF in section 3.2.3 this issue unravels.

Supplementary to the right to revoke consent, two more powerful rights have been foreseen under the GDPR, the right to object (Article 21) and the right to restriction of processing (Article 18). Although the right to object is specified also in DPD where compelling legitimate grounds must me demonstrated by the data subject in order to object to the processing of personal data, under the DPGR the definition of the right to object is significantly expanded since the burden is put on the data controller to demonstrate compelling legitimate grounds when a data subject is objecting to processing based on public interest (Article 6(1)(e)) or the legitimate interests of the controller (Article 6(1)(f)). By exercising the right to restriction of processing data subjects have the right to restrict the processing of personal data when the conditions specified in Article 18(1) apply, and consequently the data may only be stored by the controller but they cannot been further processed.

3.1.3. Current efforts for revoking consent

For the functional implementation of feasible consent mechanisms, many frameworks, both legal and technical, have been proposed over the past few years. An indicative portion of them are presented here.

Within the medical field, an option, from legal perspective, for implementing informed consent efficiently is not to implement any constraints at consent at all! As oxymoron as it sounds, this concept is adopted by the Portable Legal Consent (PLC), a US legal framework for consent in research developed by the Consent to Research project[18]. The project is aimed at developing a process through which individuals can make an informed choice about participating in research through the clear communication of risks, benefits, and consequences [50]. It allows participants who are willing to relinquish control of their personal information to attach a one-time research consent to their health and genetic data, which they upload themselves onto the web site [94]. Participants may withdraw their data from the database at any time, but they are clearly advised that once data are uploaded it may not be possible to remove it from all sources (for example, from researchers who have already downloaded, shared, or used the data). This Portable Legal Consent requires participants to go through rigorous consent processes and demand honesty and trust from both researchers as well as participants [53].

Almost a decade ago, researchers, in an attempt to provide a technical solution for granting and revoking consent under the DPD requirements, proposed an approach that provides for a verifiable and revocable expression of consent and allows services to gain a proof of consent even for aggregated personal data [95]. The solution builds on a digitally signed hash tree and reuses PKI mechanisms, especially certificates and certificate revocation, in order to cater for changes in the expression of consent and to allow the vanish of a once established consent, all accomplished without the need of a direct relationship or the iterative involvement of the data subject. However, as explained in [95], the solution does not avoid the non-consented processing of data.

Giving and revoking consent effectively has been the scope of many research projects, like EnCoRe (Ensuring Consent and Revocation), a large, cross–disciplinary project in UK. The project investigated how to improve the rigour with which individuals can grant and, more importantly, revoke their consent to the use, storage and sharing of their personal data by others [96]. One of the main research goals was to ensure revocation compliance throughout the supply chain, i.e. if an individual revokes consent with the original service provider this needs to pass through the supply chain to all other service providers who are handling the data, whereas the original service provider (and individual) need assurance that the revocation has taken place through the chain. It was within the EnCoRe project that the notion of dynamic consent was first coined by Professor Kaye and her team [97] as a way to provide dynamic and granular options for revocation in system design [96]. As authors explain, in dynamic consent the reliable storage and enforcement of consent choices is achieved by cryptographically "wrapping" the individual's dynamic consent preferences with samples/information provided. This is possible because of machine-readable disclosure policies or "sticky policies" that are attached to data [98, 99]. Sticky policies are attached metadata that define conditions and constraints describing how the data should be treated and they are strictly associated to users' data driving access control decisions and privacy enforcement [100]. This package of "wrapped information", which contains specific consent provisions, travels with the participant's data as these are shared or accessed for different purposes [67]. Under EnCoRe pilot this "wrapped information" embraced new homomorphic encryption techniques [101], which allowed information to be processed in its encrypted state while permitting the results of the processing to remain encrypted [102, 103].

Urquhart et al. in [104], acknowledging the unequivocal place of consent in the IoT era of embedded physical devices, proposed a route forward for changing

[18] http://weconsent.us

how consent is obtained by using the concept of "trajectories", already used within the HCI studies for understanding and designing complex user experiences. In their work, they are taking different elements of the trajectories framework; time, actors, space, interface, and map them onto designing consent processes that enable mechanisms for informing, obtaining and withdrawing consent.

The use of privacy agents, a dedicated software which would act as a "surrogate" of the subject and automatically manage on his behalf his personal data, has also been proposed for dealing with the management of data subject's explicit consent [105]. The recommended architecture for these "privacy agents" is based on formal (mathematical) semantics, a fact that enables the definition of the expected behaviour of privacy agents without any ambiguity and thus can make privacy rights protection more effective. The extensive use of state-of-the-art privacy agents, which enable people to configure their privacy preferences and exchange these preferences with data controllers through personal information policy exchange protocols, is analysed in [106].

As a proof of concept, recently researchers from the UK designed and developed an Apple mobile health app [107] for demonstrating the requirement of supporting informed consent and withdrawal in research projects. They implemented a custom-built module for consent, similar to the ResearchKit provided by Apple, whose functionality supports gaining informed consent, displaying template forms upon first launch, and allowing the collection of digital signatures. Although the researchers provided the choice for a complete data withdrawal from the study, they designed this functionality as a multi-step process in order to avoid situation where users withdraw data by mistake.

The approach of OPERANDO[19], an EU funded project aiming at implementing and validating an innovative privacy enforcement framework, is to create a vault where users store their sensitive data and selectively share them with Online Service Providers (OSP). To this end, users access a dashboard which allows them to manage which OSP accesses what data and when, and easily revoke or grant access to the data. To facilitate OSPs OPERANDO allows them to query the stored data with the use of the OData standard[20] and enforce the user's privacy policies in the results of each query.

In the same context, various web standards have emerged for the specification and implementation of consent procedures in online environments. One such standard is the User-Managed Access (UMA) [108] which has been approved by the Kantara Initiative[21]. UMA proposes an OAuth-based architecture that enables conforming applications to offer stronger consent management capabilities and an asynchronous, centralized protocol for consent. While UMA has been under development for several years, its specifications have now been stabilized and support multiple implementations and a widening variety of use cases. The authorization policies that anticipated to be used in conjunction with the UMA is the eXtensible Access Control Markup Language (XACML)[22], which defines a declarative fine-grained, attribute-based access control policy language, an architecture and a processing model describing how to evaluate access requests according to the rules defined in policies.

Kantara Initiative also supports the standardization effort of Consent Receipt[23], a form of signed receipts in a JSON Web Token Format which can be used to improve existing consent mechanisms against the requirements specified by regulations, and in particular the GDPR. Consent Receipts can facilitate people's use of consent when communicating with the data controllers as well as when withdrawing consent. In conjunction with UMA mechanisms, consent receipt can be used as a tool for demonstrating effective personal control over data. According to its specifications [109], in some respects a consent receipt could be described as a reverse cookie, in that, both the individual and the organization have a record of the consent, and the individual can use the receipt to track and profile the organization and/or service along with consent and information sharing preferences. Thus, people can track sharing with third parties, like third parties can track people. Inspired by the Kantara's standard, consent receipt prototypes were also deployed under the Personal Data Receipt project of Digital Catapult Centre in the UK[24] [110].

Beyond the above, other national wide efforts across EU countries have also been initiated aiming at the development of nationally and internationally interoperable models for personal data management. Examples are the MyData Initiative in Finland[25], which specifies reference architecture in order to provide a rigid framework for consent and data authorization management via a standard and interoperable mechanism[26], and the Consent Group within the Personal Data and Trust Network in the UK[27], which aims at using next generation standards in consent and facilitating the use of consent-based trust in digital framework use cases.

Evidently, since the publication of the final text of the GDPR and the new requirements brought upon the

[19] https://www.operando.eu/

[20] http://www.odata.org/

[21] https://kantarainitiative.org/confluence/display/uma/Home

[22] https://www.oasis-open.org/committees/xacml/

[23] https://kantarainitiative.org/confluence/display/LC/2017/01/08/Consent+Receipt+Specification+v1.0+public+comment+and+IPR+review+period?src=contextnav

[24] https://www.digitalcatapultcentre.org.uk/project/pd-receipt/

[25] https://mydatafi.wordpress.com/

[26] http://hiit.github.io/mydata-stack/ or https://hiit.github.io/mydata-stack/stack.html\#intro

[27] https://pdtn.org/pdtn-group/consent/

Privacy Shield agreement[28], the technical discussions on the feasibility of granting and revoking a simple, informed, unambiguous and declarative consent, along with its receipt as a proof of discourse, have been intensified. The Real Consent Workshops, which is the combined effort from big standard initiatives like Kantara and Digital Catapult, intent to delve into the gap between the type of consent people find meaningful and what we have online today[29]. Real Consent efforts are focused on both technology and policy, are comprised of open standards, best practices, and most importantly, aim at developing a collection of assets to address the current challenges around consent. In parallel, the Open Consent Framework is an approach to operationalize standard notices with a trust framework, i.e., a notary function where trusted third-party organizations can register/generate their notices and in which additional layers of technology can be added to provide more advanced and more trusted user functionality[30].

3.2. The right to be forgotten

Although the non-retroactive definition of consent revocation does not allow the forgetting of past processes and inferences carried out based on personal data once they were collected, the GDPR introduces the concept of the Right to be Forgotten (RtbF) or the "Right to Oblivion" for allowing the retroactive erasure of the actual personal data themselves. Forgetting previously collected personal data, obtained either because the user has once submitted them or because an online service has sneakily scrapped them, has been for long a disputable and controversial matter the European Commission attempted to untangle with legislation. Given the notably infeasibility for users to maintain control of their data, their diffusion and their subsequent uses once they were collected, the right aims at counterbalancing this luck of transparency on personal data processing.

3.2.1. Forgetting and the need to be forgotten

The right evolves from the need for forgetting which, according to Bannon [111], is a central feature of our lives, yet it is a topic that has relatively little serious investigation in the human and social sciences. He outlines that judicious forgetting is of fundamental value both for individuals and societies, a necessary human activity and not simply a bug in the design of the human. Although we most often live under the assumption that remembering and commemorating is usually a virtue and forgetting is necessarily a failing, many scholars argue otherwise [112, 113]. Memory processes have always contained both the practices of forgetting and of remembering since our memory is a combination of what we remember about our past, what we may have forgotten about it, and what we wish to forget. Even within the justice system, we see the development of practices that require certain kinds of deliberate forgetting after a period, in order to allow people to have a new start in life and not be haunted by an indiscretion many years earlier [111].

The importance of forgetting–whether by individuals, groups, organizations or even nations–has been observed and commented by scientists, historians, politicians, philosophers, writers and poets through the ages. French distinguished philosopher Ricoeur in *"Memory, history, forgetting"* [114] discusses the necessity of forgetting as a condition for the possibility of remembering and affirms that the "power to forget" is necessary to all actions, describing it as the very power that allows the one possessing memory and history to "*heal wounds, to replace what has been lost, to recreate broken forms out of itself alone*". In *"The End of Memory"*, Volf [115] clarifies that the injunction to remember carries within itself an allowance for forgetting: "*Remember, yes; but for how long?*" he questions and extols "how to remember rightly" so that memory might be able to rest. German philosopher Friedrich Nietzsche in his essay *"On the use and abuse of history for life"* [116] demonstrates how it is generally completely impossible to live without forgetting.

As Mayer-Schönberger describes in widely cited work [117], forgetting performs an important function in human decision-making. It allows us to generalize and abstract from individual experiences. It enables us to accept that humans, like all life, change over time. In the medical domain, recent research on hyperthymesia condition, an "unusual autobiographical memory" [118] disrupts the lives those living with it as they describe their never stopping memories to be "exhausting" and a "burden". The condition has been the inspiration for many fiction and movie artifacts[31]. Among these the most notable is Borges' infamous fiction book *"Funes the Memorious"* [119] that describes a man who is suffering by a similar condition and he is being haunted by his inability to forget anything, and as a result his life has been a misery.

Although in the paper-and-ink world, as explained in [120], the sheer cumbersomeness of archiving and later finding information often implied and promoted a form of institutional forgetfulness–a situation with parallels to human memory, in the digital world our digital records constitute an array of potential memories, the very existence of which may compromise our ability to forget, or move on [121, 122, 123]. Whereas in the past forgetting was the default, due to the cost and rigour embroiled in remembering, digital age changed this assumption and caused the balance of remembering and

[28] www.privacyshield.gov

[29] http://www.real-consent.org/, https://kantarainitiative.org/real-consent-workshops-the-consent-tech-bubble-grows/

[30] https://pdtn.org/workshop-highlights-creating-real-consent/

[31] https://en.wikipedia.org/wiki/Hyperthymesia

forgetting to be inverted and thus forgetting to be the exception [117]. As de Andrade argues in [6], *"the past is no longer the past, but an everlasting present"*, while Burkell [123] underlines that our ability to construct and maintain our own identities is threatened by digital systems that "remember" everything about us: thus, there is value in, and a need for, forgetting and being forgotten. Therefore, in the context of informational systems, we should view forgetting as a feature and hence, to try to use technology to augment forgetting in human–computer interaction in order to "teach" computers to forget [111].

Yet, computer scientists have not given a lot of thought on the phenomenon of forgetting as the capacity of modern computers to store everything and never forget has been considered always as a want-to-have feature. However, according to Blanchette and Johnson [120], who were among the first computer scientists to have envisioned the need for forgetting within the information systems, privacy policies must address not only collection and access to transactional information, but also its timely disposal as part of a broader and comprehensive policy approach [120]. In this regard, Dodge and Kitchin [124] had been arguing, almost a decade ago, that rather than seeing forgetting as a weakness or fallibility, it must be seen as an emancipatory process that will free pervasive computing from burdensome and pernicious disciplinary effects. Currently, digital memories, comprising of the vast amount of data currently collected as we go about our everyday lives, make possible a comprehensive reconstruction of our words and deeds and, even if they are long past, they strongly suggest we are moving into a panoptic society as they create a temporal version of Bentham's panopticon [125], constraining our willingness to say what we mean [117, 120, 126]. Search engines, most notably Google Search, stand at the heart of this panoptic architecture of the Internet [126] as web enables the retention of large quantities of personal micro-information over time, which can provide for an extremely detailed reflection of our past. Rosen remarks [127] "*the fact that the Internet never seems to forget is threatening, at an almost existential level, our ability to control our identities; to preserve the option of reinventing ourselves and starting anew; to overcome our checkered pasts... The Internet is shackling us to everything that we have ever said, or that anyone has said about us, making the possibility of digital self-reinvention seem like an ideal from a distant era*". Within this context, Solove in his book [128] explains how the free flow of information on the Internet can make us less free: "*Information that was once scattered, forgettable, and localized is becoming permanent and searchable. Ironically, the free flow of information threatens to undermine our freedom in the future. These transformations pose threats to people's control over their reputations and their ability to be who they want to be. The more freedom people have to spread information online, the more likely that people's private secrets will be revealed in ways that can hinder their opportunities in the future*". Indeed, examples of the devastating consequences of digital forgetfulness are spread across the literature. Indicative cases are the teacher who lost her job over a photo of her holding a glass of wine posted on Facebook[32], and more recently, the case of Harvard University who withdrew acceptance of 10 freshmen over to their offensive postings in a group Facebook chat[33].

Undoubtedly, with the rising of WEB 3.0 era [129] the explosion of data on the web has emerged as a new problem space. Semantic web technologies integrated into, or powering, large-scale web applications and Linked Data best practices for publishing and connecting structured data on the web [130] contribute to the boosting of personalization and contextualization of information. The dominance of intelligent search services and the efficient inferences produced by Artificial Intelligence algorithms pave the way for the endless information dissemination and the vitiation of forgetfulness. Consequently, the need for forgetting in digital age has begun to occupy more and more computer scientists who are beginning to realize that forgetting is an essential part of HCI systems. For example, in recent years attempts have been made for modelling forgetting in robotic devices in order to support a more realistic and natural digital illusion of life experience [131]. Additionally, the value of intentional forgetting at situations in which people may be highly motivated to forget has been studied so as to provide implications for designing complex practices associated with problematic disposal of digital possessions [132].

3.2.2. About the CJEU decision

Amidst the social and philosophical discussions on the criticality of forgetfulness in digital era, in 2014 the Court of Justice of the European Union (CJEU) tried to tackle the need for forgetting through the infamous Google Spain decision which forced Google to take down harmful personal information from its search results [133, 92]. Although the final settlement ordered Google to remove the relevant link at first only from its corresponding Spanish domain and later from all European Google sites, later, the French privacy authority CNIL requested irrelevant and outdated contents to be removed from all non-European sites as well and therefore fined the company for non compliance. This provoked an intense debate between CNIL and Google which appealed to this request and the final decision is being longingly awaited[34].

[32] http://www.cbsnews.com/news/did-the-internet-kill-privacy/

[33] https://www.washingtonpost.com/news/morning-mix/wp/2017/06/05/harvard-withdraws-10-acceptances-for-offensive-memes-in-private-chat

[34] https://www.bloomberg.com/news/articles/2017-01-25/google-argues-right-is-wrong-in-clash-with-french-

Yet, contrary to general impression, it has been pointed out by many scholars [112, 133, 92] and the Commission itself that the Google Spain judgement does not create a RtbF, as the CJEU could not enforce a right that does not exist in the current legislation, but simply applies the RtbF which was already present (although not explicitly mentioned) in the existing legal framework, extending the lawfully published information right and the right to object. Still, the decision planted the seeds to affirm something that goes in the direction of the RtbF [92].

Although plaintiff's original intention was to remove the disputed information from the online archive where it was originally posted, CJEU ruling aimed at the technological intermediary and not the original publisher of information and thus, the information was legally retained in the online archive whereas the links to the information removed from Google Search. This was considered by some, such as Gorzeman and Korenhof [126], as an elegant solution: history is still retained and accessible but the access is less easy. Forgetting invoked by the CJEU's decision may in time challenge historians with the retrieval of information in order to get an accurate view of past societies, but this difficulty is not automatically an impossibility, something that would be the case if the information were thoroughly deleted on the storage level or not encoded at all [126]. The court clarified that Google has to carefully balance the request for removing search results and of all the rights involved, including the public's right to have access to information. This would limit the application to cases only where the information to be deleted is both damning and irrelevant. On top, as Mayer-Schönberger emphasizes in [134], search engines don't have to redesign themselves to comply as Google is already handling millions of deletion requests for copyright violations every month. Indeed, the CJEU decision has had such an impact that, since it was handed, more than 734 484 requests filed for being de-listed from Google search[35] and 43,1% of them were satisfied.

In relation to the alleged censorship imposed by the CJEU decision, O'Hara's comments in [135] that the CJEU decision, although makes life complicated for the big corps, isn't targeted at particular types of information or data subjects, and therefore cannot be considered censorship. Baum [136] also explains that, "*the court's ruling simply takes us back to the time when, if you wanted to find out something about someone, you had to dig for it; you had to know where to look for it. Censorship would be if the offending records themselves were expunged, and that is not what the court ruled*". Indeed, as the disputed information was not to be deleted from the web, and hence censored, the decision was ultimately not about the fundamental balance between privacy rights and expression rights when dealing with personal information over the web [112]. As the impact analysis of this decision is still ongoing and its distinction with the RtbF under the GDPR is rather vague, many scholars argue that the relationship between the regulation's RtbF and the CJEU's reasoning will clearly require careful elaboration hereafter [52]. Nevertheless, as Mayer-Schönberger highlights [134], the CJEU decision has not definitely solved the challenge of comprehensive digital remembering.

privacy-czar

[35] https://www.google.com/transparencyreport/removals/europeprivacy/

3.2.3. The right to be forgotten under the GDPR

Back in 2012, EU, in an attempt to respond to the challenges posed by digital remembering and having as ultimate goal to give control of personal data back to individuals, proposed the RtbF in its recently adopted regulation. The right evolves from the national law in many European countries like France where the Right to Oblivion is anticipated. According to some legal experts, the RtbF enshrined in the GDPR has more a symbolic importance than a substantive effect as it does not actually represent a revolutionary change to the existing data protection regime but its roots lie within the DPD and in particular within the right to erasure and the right to object, although the GDPR is more analytical in defining the right and the conditions under which it shall be invoked [6, 137, 138]. For instance, the condition of withdrawing consent in order the RtbF to be triggered has not been encompassed in any national or European data protection law so far [137].

Admittedly, this right as introduced in the Article 17 of the GDPR is a breakthrough on the EU legislation domain because does not only encompasses the right to erase (or "to forget") but it also embraces the right "to be forgotten". While the first specifies the need for a controller to delete data, the latter implies the need for data to be deleted *from all possible sources* in which they reside. According to extended legal analysis [52, 92], the right is a novelty and has a broader scope than any of the existing rights whereas its unique feature, which makes it different from the rights granted by the existing legislation, is its retro-activity. Article 17(1) provides several situations where a person has the right to ask personal data to be erased by the data controller. Of particular interest is sub-paragraph b, which allows the person to withdraw his or her consent. In other words, based on the GDPR, withdrawal of a previously given consent is sufficient to have personal data erased by the controller. Under the regulation, an individual can request erasure of his personal data *from every data controller* who is processing the data and not only from the one who processed the data in the first place (Article 17(2)). The fact that the consent was provided only to the original controller does not appear to be relevant since the obligation for erasure arises when the person withdraws consent, without any

specification on the controller who received it.

From the above, it is evident that the enforcement of this right would pose major technical issues due to the practicalities involved in knowing all the controllers who are processing the personal data in question. Even in the case where controllers do have knowledge of the third parties processing some data that they collected, it places upon them the additional obligation to inform those third parties about the erasure request, given that Article 17(2) states that "*... the controller shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data*". Hence, controllers are required to implement technical solutions to allow the tracking of personal information and to prove its efficient removal in the case of request for erasure under the RtbF. And although the first may not be considered a difficult task, since many controllers keep links of their copied information, the burden to prove that the erasure has been implemented successfully from all available sources is still technologically questionable. The fact that the regulation does not provide a clear and unambiguous definition of the RtbF regarding its non-trivial practicalities of enforcing such a deletion when secondary uses apply, i.e. personal data have been disseminated to third parties or they have been anonymized or pseudoanonymized, led many to argue that its future enforcement is reasonably doubted [139, 140]. Nevertheless, the GDPR provides a convenient exemption from the obligation to inform all recipients of any rectification or erasure when this *"proves impossible or involves a disproportionate effort"* (Article 19). Yet, this exemption has also raised some concerns regarding the effectiveness of the RtbF as its scope of applicability is not always obvious [141].

Inevitably, the right provoked plenty heated debates and fierce discussions within law, philosophy, social, humanitarian and computing disciplines and has been lengthy explored in surveys, proposals and academic writings. Xanthoulis in [139] asserted that the RtbF should be conceptualised as a human right and more specifically as an expression of the broader right to privacy, whereas de Andrade in [6] presents the RtbF as a branch of the right to identity, which is the right to be different, not from others but from oneself, i.e. from the one(s) we were before. Therefore, the RtbF – as part of the right to personal identity – is intimately connected to the ability to reinvent oneself, to have a second chance, to start over and present a renewed identity to the world. Following this line, Burkell [123] explores the consequences of the digital record of our lives for identity and argues that the RtbF may be, above all else, a psychological necessity that is core to identity – and therefore a value that we must ensure is protected. Yet, the RtbF has been met with intense resistance from both businesses and free speech advocates due to its collision with other rights and protected interests[36][6, 52, 19, 47]. They questioned the regulation's incentives and emphasized the difficulty on achieving a delicate balance between the involved rights, namely the right to privacy and the right to freedom of expression which, along with the right to privacy, is also contained in the European Convention on Human Rights (Article 10) [9]. Google's chief privacy counsellor remonstrated that the RtbF represents the biggest threat to the free speech and expression on the Internet[37][142] because it is not limited just to personal data that people provided themselves through an unambiguous consent agreement, but instead, it applies to all possible cases of personal data may be found online[38] [19]. Further, the RtbF has also been labelled by some as censorship and disastrous for the freedom of expression[39] whereas some argued that "*a right to be forgotten is about extreme withdrawal, and in its worse guise can be an antisocial, nihilist act*" and "*is a figment of our imaginations as it neglects the role society plays in individual's life*"[40]. On an informational level, scientists have pointed out that enforcing the RtbF would lead to preventive actions like anonymization of databases per default, something that would cause an unacceptably high amount of information loss [143].

Along the same lines, Rosen [142] condemned any efforts at regulating the Internet, and search engines in particular, as he asserted that any kind of regulation of the Internet violates the inherent code of its freedom and thus, the RtbF will bring chilling effects on the Internet era which *will not be as free and open upon the application of the right*. On the other side, scholars and privacy experts argue that free speech is already being selected and restricted by search engines themselves [144, 145]. Google's global privacy counsel argues "*history should be remembered, not forgotten, even if it's painful. Culture is memory*"[41] whereas other eminent theorists state otherwise [139, 114, 115], i.e., that forgetting is a necessity for the evolvement of history remembering considering that cultures seem to have been built over the course of time through a process of selective remembering and forgetting, not through total remembering. In a more compromised approach, Mitrou and Karyda point out [146] that, while the RtbF cannot be synonymous with a right of a total erasure of history, the interests of social and historical inquiry does not legitimize keeping every

[36] https://www.theguardian.com/commentisfree/libertycentral/2011/mar/18/forgotten-online-european-union-law-internet

[37] http://peterfleischer.blogspot.gr/2011/03/foggy-thinking-about-right-to-oblivion.html

[38] http://peterfleischer.blogspot.gr/2012/01/right-to-be-forgotten-or-how-to-edit.html

[39] http://www.wired.co.uk/article/right-to-be-forgotten-blog

[40] https://www.theguardian.com/commentisfree/libertycentral/2011/mar/18/forgotten-online-european-union-law-internet

[41] See footnote 35

piece of personal information regardless the rights and interests of the persons affected.

With regard to the conflict between privacy and freedom of speech, Solove in his notable work [128] argues that "*we must protect privacy to ensure that the freedom of the Internet doesn't make us less free... we must balance the protection of privacy against freedom of speech*" and "*Both are essential to our freedom. Freedom of speech is a fundamental value, and protecting it is of paramount importance. Yet, privacy often furthers the same ends as free speech. If privacy is sacrificed at the altar of free speech, then some of the very goals justifying free speech might be undermined*" [128]. And Lindsay explains in [147], "*privacy is not necessarily the opposite of freedom of expression – if people feel assured they have some control over their information, they are more likely to share it. On the other hand, if people know that what they say and do online will be accessible to all, and for all time, they may be more likely to self-censor*".

Other academics [126, 148] proposed a more conciliatory position, arguing that since the RtbF draws more heavily on the mechanisms of human forgetting which provides for a big greyscale (in contrast to erasure which is black-and-white), an individual can have a need for different grades of forgetting rather than plain erasure, and therefore the RtbF, instead of the plain erasure of information, could rely on the level of encoding or retrieval of the information. Although this position it may considered by some as censorship at the level of information retrieval [126], and hence the least heavy yet most effective means to get the minimum amount of censorship overall, it still resembles the decision concluded by the CJEU to remove the links and not the information itself, and therefore it cannot be considered as an actual forgetting.

In a large part of academia, the RtbF was considered as a highly qualified right as it attempts to restore some balance in favour of individuals by providing tools for controlling their personal data, while at the same time certain conditions are foreseen to be satisfied in order the right to be applicable, such as when data are no longer needed, or where data are collected or processed with a person's consent and that consent is later withdrawn. The RtbF is also subject to important exemptions and safeguards, such as the cases where it may conflict with the freedom of expression and information, or for journalistic purposes and the purposes of academic, artistic or literary expression (Article 17.3 and 85). Claims that the RtbF, as introduced in the GDPR, will stifle the press are therefore untrue, since there is an expressed exemption for journalists, as well as an exemption for individuals engaged in purely personal or household activities [147]. In this respect, Mantelero in [137] highlights that the oppositions the RtbF received concerning the suppression of freedom of speech represent a sort of paradox. On the one hand, big IT companies are trying to promote the idea that sharing information is a social norm and that privacy and forgetting are outdated concepts, but on the other hand, the same companies are progressively collecting an enormous amount of data in order to profile individuals and, above all, to extract predictive information with high economic, social, political and strategic value.

Another broad area of criticism against the establishment of the RtbF comes from the fact that it may impose considerable obstacles in data transfer between EU and third countries. As a first step for resolving the impending implementation and interoperability issues resulting from the enforcement of the RtbF, Ambrose in [149] analysed the options non-EU countries and data controllers (like the USA) have to react to the establishment of such a right, while Bennett in [150] discussed how a reconciliation between the USA and EU on the RtbF can be achieved. On this matter, Voss and Castets-Renard proposed a coherent worldwide taxonomy of the RtbF [138] in order to identify its various forms within different countries and to measure the extent to which there is a convergence of legal rules internationally. Ultimately, the Privacy Shield Framework[42] between the EU and the USA will have to deal with this issue drastically.

3.2.4. Implementation challenges and proposed solutions

The implementation of the RtbF in the digital environment is not a straightforward task and can't be achieved without affecting the value of already collected data stores. Technically speaking, the effective implementation of Article 17(2), which require controllers to take *"reasonable steps, including technical measures"* to inform third parties when a data subject has requested the erasure of previously published personal data relating to them, may be proved burdensome or even impossible in any number of scenarios [49]. Legislators deliberately avoided the idea of recommending specific technical frameworks or privacy preserved methods for implementing the legal requirements introduced by the GDPR. Instead, they followed a technology-agnostic approach by specifying the functional requirements in a highly abstracted level, as far as their underlying implementation is concerned, and as such they didn't bind the provisions of the law with current trends and state-of-the-art technologies in computer science. The ultimate purpose of this approach was to allow the GDPR's adjustment to future technical innovations. Yet, upon the GDPR's application across the European Union on 25 May 2018, businesses and organizations should have operational-ready implementations of the requirements in a transparent and efficient manner. To this end, we discuss below some technical methods and frameworks, existing either in business or academic environments,

[42] https://www.privacyshield.gov/

and we highlight their weaknesses and strengths in terms of implementing user's full control over their personal data, and particularly their effective erasure from third party controllers where the data have been disseminated. As the GDPR dictates, the triggering events for the erasure to take place would be the invoking of the RtbF under any of the conditions (a)-(f) described in the article 17.1, which also includes the case where the data subject exercises his right to withdraw a previously given consent.

While, beyond any doubt, academia and industry are currently working vigorously towards the design of technical solutions and the conformance of current infrastructures to the new requirements for forgetting within digital environments, existing frameworks need to be evaluated for compliance with the RtbF requirement and, if needed, to be amended accordingly. Unfortunately though, recent exercises have demonstrated that state-of-the-art technologies used in large cloud mainframes face technical constraints which may affect the lawful implementation of the RtbF. For example, it has been underlined in an exercise regarding the compliance of Data Lake Enterprise Architecture Model with the GDPR [151] that the immutability of Hadoop is a phenomenon which does not allow files to be physically updated or deleted. Instead, a new instance of the file is created and automatically becomes an active one whereas previous instances of files are not deleted, only flagged as not active. This property of HDFS files to remain always undeleted, prevents Data Lake architecture from achieving compliance with the RtbF which requires assured deletion of personal data.

While there are some recent studies proposing technical solutions for a feasible implementation of the RtbF, most of them concentrate to the problem of implementing a RtbF compatible with the CJEU decision, that is a right to delist personal information from search engines, and not the one enforced by the GDPR to entirely remove it. For example, in a recent undertaken analysis [152] of the RtbF in the context of the CJEU decision and its impact on search engines, it was suggested an implementation approach based on Personal Data Management Architectures (PDMA). However, the approach does not deal with the situation of permanently deleting information from all of data controllers holding this information. On the contrary, hereafter we focus on discussing indicative methods and frameworks that can been employed for implementing the permanent and parallel deletion of personal data upon request.

In this context, and in spite of earlier studies on establishing theoretical foundations for the design of mechanisms for forgetting of personal information [153], there exist numerous arguments against the feasibility of deleting information on the Internet, based mainly on the easiness of copying information, and hence, on the difficulty or impossibility to ensure that information can be ever completely erased [147]. Indeed, while industry has heretofore developed tools for facilitating users in their personal data administration, Novotny and Spiekermann in [154], after studying 13 available online services, concluded that, while half of them provide for some erasing mechanisms, none of them provide intelligent capabilities to forget outdated personal information. For example, Google, has long ago introduced Google Dashboard in the spirit of providing users with the capabilities of viewing, managing and deleting their online personal information like web searches, shared docs etc. Undoubtedly, the sheer amount of information the users view when they browse their dashboards usually brings chilling effect about the extent of the collected information by Google who knows more about their Internet activity than they do[43]. Nevertheless, although the deleting option of this information is a kind of relief, there is not any evidence of permanently erasure of these data from company's servers. On top of that, one can never know if data were replicated to other sites or services. Other industry efforts enabling the full control of personal data include personal data storages like TeamData[44], which enables users to securely and collaboratively manage data in their workplace by providing individuals with an online "data vault" where their work information is being stored or shared, while at the same time privacy by design principles are followed to assert privacy. Still, the platform handles work-related data and not strictly personal.

While the Internet is spread with tech counselling articles and services on how one could delete all online personal information from various web services, there is not an automatic way to ensure the erasure of outdated or erroneous personal data from all of the services they may have been disseminated once they were uploaded. Even the "Web 2.0 Suicide Machine"[45] initiative launched in 2010, a tool that allows users to "suicide" their electronic selves in the social networks by automatically removing user's private content and friend relationships from these sites, operates with a handful of sites and does not guarantee full online removal. Given the futility of deleting online personal data, professional reputation managers, implementing strategies that rely on techniques of burying offending information rather than removing it, have emerged recently. Nonetheless, these services address only the tip of the information iceberg [123]. In the light of this, and as more and more reputation queries are being processed by a handful of de facto reputation brokers, scholars have proposed a form of "reputation bankruptcy"[46], a choice which will allow individuals to

[43] https://fossbytes.com/google-tracking-dashboard-myactivity/ and http://www.seochat.com/c/a/google-optimization-help/google-dashboard-an-overview/

[44] https://teamdata.com/

[45] http://suicidemachine.org/

[46] http://blogs.harvard.edu/futureoftheinternet/2010/09/07/reputation-bankruptcy/

wipe their online reputation slates clean and start over after a predefined number of years.

The reputation bankruptcy idea evolves from the theoretical work for introducing forgetting in informational systems, suggested by Bannon in the early 21st century [111]. Bannon envisioned that private messages might be marked so that it is not possible to forward them without author permission or that all social messages to be designed to fade away over time. He imagined various kinds of electronic tagging systems for messages that could time-stamp data and may contain something like a "sell-by" date in order to explore augmentation means for all human activities, both remembering and forgetting [111]. Similarly, Solove [128] imagined a world in which digital-storage devices could be programmed to delete photos or blog posts or other data that have reached their expiration dates, and he suggested that users could be prompted to select an expiration date before saving any data.

In this respect, Mayer-Schönberger [117] elaborated in the concept of forgetting through expiration dates for information. He described the various structural, legal, and technical components of expiration dates and how they would work together, while he offered a spectrum of possible implementations based on how thoroughly policy-makers and the public desire to revive forgetting [117]. Meanwhile, computer scientists have initiated research on privacy preserving ubiquitous computing frameworks and policies that enable enforcing limited retention periods for personal data storage [155, 156, 157] and for assuring complete deletion of data and files [158, 159]. Following these lines, in [160] researchers demonstrated a relational database wherein once a tuple has been "expired" – any and all its side-effects are removed, thereby eliminating all its traces, rendering it unrecoverable, and also guaranteeing that the deletion itself is undetectable. Nevertheless, as it has been mentioned in many exploratory essays [49, 19, 147], the practicability of this theoretical principle is far from evident. Critics of the data expiration idea argue that even if auto-expire tools existed, they would do nothing to prevent the usual privacy problems when someone copies content from one site and moves it to another not supporting the auto-expire function[47]. This is the reason, as Mantelero notices in [137], why the idea of fixing a general time limit for mandatory erasure has been correctly avoided in the GDPR. Time, however, has been identified as a critical factor for introducing forgetting by many scholars, like Korenhof et al. [161] who argued that we should not overlook or disregard the importance of time in weighting the opposing interests when we are shaping policy mechanisms like the RtbF. Having this in mind, Korean researchers patented and sold a technique called Digital Aging System (DAS) which attaches "aging timer" to digital personal data [162], whereas the German company Xpire[48] developed a smartphone app that enables the creation of self-destructing social posts in Facebook, Twitter, and Tumblr.

Acknowledging the importance of time and evolving on the data expiration concept, the idea of data degradation is also proposed. Privacy-aware data management by means of data degradation, whereby sensitive data becomes less sensitive over time as a result of various degradation processes [163], is based on the assumption that long lasting purposes can often be satisfied with a less accurate, and therefore less sensitive, version of the data. Data are progressively degraded such that they would still serve application purposes, while their accuracy has been decreased and thus the privacy sensitivity as well. Yet, data degradation still faces the same weaknesses as those described for expiration dates since it cannot prevent the undesirable copy of data before their initial degradation.

There are also other theoretical approaches for deleting, mainly due to the exercise of consent withdrawal, like the one presented in [76] for biobanks, that suggest the withdrawn samples and data to be parked in "limbo" or be dead-locked for a period of time and only destroyed/erased at the end of that period if the person withdrawing has not changed her mind. However, this approach cannot be legally accepted upon the GDPR's enforcement since the requirement for implementing consent withdrawal under the GDPR imposes data to be deleted "without undue delay" when an individual withdraws consent and the consent takes effect (in case neither of the exemptions described in Article 17(3) apply).

Pursuing the feasibility of forgetting in digital systems after a period of time, researchers, less than a decade ago, introduced Vanish [164], a very prominent technology for enforcing forgetting that causes sensitive information, such as emails, files, or text messages, to irreversibly self-destruct, hence "vanish", automatically after they are no longer useful, and all that without employing any centralized or trusted system. Vanish ensured that all copies of certain data become unreadable after a user-specified time even if an attacker obtains both a cached copy of the data and user's cryptographic keys and passwords. Instead of relying on data controllers to delete the data stored "in the cloud", Vanish encrypted the data and then "shattered" the encryption key. To read the data, the computer had to put the pieces of the key back together, but these "eroded" or "rusted" as time elapsed, and after a certain point the document couldn't longer be read. Vanish leveraged the services provided by decentralized, global-scale P2P infrastructures and in particular Distributed Hash Tables (DHTs), for encrypting user data locally with a random encryption

[47] See footnote 35

[48] http://getxpire.com/xpireApp

key not known to the user. Then it destroyed the local copy of the key and sprinkled bits of the key across random nodes in the DHT. Moreover, it did so without any explicit action by the users or any party storing or archiving the data, in such a way that all copies of the data vanished simultaneously from all storage sites, online or offline. Unfortunately, while Vanish seemed a very promising solution, scientists managed to break it not long after its initial publication [165], and although the research team tried to tackle the problems identified in a following work [166], they never seemed to have succeeded their original goals. Yet, in the following years an increasing amount of encouraging follow up research works has been carried out for the quest of new improved versions of the Vanish prototype without facing its vulnerabilities [167, 168, 169, 170].

Enforcing fine-grained data management obligations and improving the accountability of responsible parties as specified by policies and regulations is the focus of another research approach named Information Flow Control (IFC) [171, 172]. IFC is a data flow control model that enforces policy against every flow in the system. To achieve IFC, tags are linked with data and entities in order to represent various properties and policies concerning the data flow. The tags are collected into two labels: a) a secrecy label representing the data's privacy/confidentiality/sensitivity; and b) an integrity label representing the data quality/provenance/authority. IFC can assist with the erasure concerns coming under the GDPR's RtbF requirement as data flows are audited, and thus it is possible to determine where data has gone and to ensure that the deletion requests are directed to all relevant entities.

Extending the concept of IFC for managing personal data in the mobile environment, Enck et al. [173] proposed TaintDroid, an efficient, system-wide dynamic taint tracking and analysis system capable of simultaneously tracking multiple sources of sensitive data within the Android environment. TaintDroid provides real-time analysis by leveraging Android's virtualized execution environment to monitor the behaviour of third-party Android applications and to automatically label (taint) data from privacy-sensitive sources while transitively applying labels as sensitive data propagates through program variables, files, and interprocess messages. When tainted data leave the system, TaintDroid logs the data's labels, the application responsible for transmitting the data, and the data's destination. Although authors' primary goal was to detect when sensitive data leave the Android system, we firmly believe that TaintDroid, along with other equivalent IFC models for cloud environments [172], could be used to provide visibility on how applications treat private data, and simultaneously, to satisfy forgetting requirements under the RtbF.

More recently, Zyskind et al. [174] proposed the use of blockchain concept (which has already demonstrated in the financial space that trusted, auditable computing is possible) for the implementation of a platform that enables users to own and control their data without compromising security or limiting companies' and authorities' ability to provide personalized services. More specifically, they described and implemented a decentralized personal data management system and a protocol that turns a blockchain into an automated access-control manager which does not require trust in a third party. They accomplished this by combining a blockchain, re-purposed as an access-control moderator, with an offblockchain storage solution focused on privacy in which, at any given time, the user may alter the set of permissions and revoke access to previously collected data. Users are not required to trust any third-party and they are always aware of the data being collected about them and how they are used. The decentralized nature of the blockchain, combined with digitally-signed transactions, ensures that an adversary cannot pose as the user or corrupt the network, since that would imply the adversary forged a digital-signature or gained control over the majority of the network's resources. Therefore, this decentralized platform makes legal and regulatory decisions about collecting, storing and sharing sensitive data much simpler because it is possible laws and regulations to be programmed into the blockchain itself, so that they are enforced automatically. In this respect, the proposed solution is rendered as a very good candidate for implementing the RtbF requirement specified in the GDPR.

Taking into account the new data protection requirements enforced by the GDPR, Microsoft researchers [175] proposed an interoperable context-aware metadata-based architecture that allows permissions and policies to be bound to data, enabling this way any entity to handle the data in a way that is consistent with a user's wishes, including revoking a use previously granted. Within this architecture, a trusted processing container is used in order to ensure data are processed according to the policies specified by the associated metadata, even when they are temporarily separated for performance. When data leave the container, the metadata, which provide a layer of abstraction from the data, are reattached. Processing and interpretation would occur first on the metadata, whereas the data themselves can be used only when the entity has been properly validated as having the right to use the data. The architecture is flexible enough to allow for changing trust norms in order to help balancing the tension between users and business and to satisfy regulators' desire for increased transparency and greater accountability, while still enables data to flow in ways that provide value to all participants in the ecosystem. Although this metadata-based architecture is considered as a useful building block for enabling and supporting the RtbF imposed by the GDPR, metadata alone cannot guarantee that entities will abide by specified policies. Nevertheless, it can facilitate their enforcement

by making them readily accessible, and, when implemented as part of a principles-based policy framework, can enforce trustworthy data practices imposed by regulations.

4. CHALLENGES AND CONCLUSIONS

Until the GDPR's enforcement on the 25th of May 2018, many challenges have to be faced by data controllers operating within the EU to meet its legal requirements. Despite the long lasting heavy discussions, negotiations and revisions on the final GDPR text and the ample time given to organisations to apply the corresponding changes to their processes, products and services, few organisations are yet able to prove actual GDPR compliance. One of the main reasons for this is that GDPR is mostly a legal document, providing little if any technical guidance to the entities that are obliged to implement it. Although this was an intentional choice as the EU did not want to bind GDPR to explicit technologies that would favour specific platforms and solutions, this technology agnostic approach may cause unforeseen complications to organisations attempting to adapt their internal processes to the GDPR's provisions.

For instance, one of the most profound difficulties would be the conformance of existing backup procedures in order to meet the forgetting requirements. Due to the reliance on ICT, institutions are obliged to keep regular backups of their data in case of security incidents or physical disasters. A big question arising under the GDPR's forgetting legislation is how organisations should handle their backups once a user requests to remove his data. Apparently, according to the GDPR this deleting action must be performed in the backups as well, opening thus the door to potential data abuses, deliberate exploitations or even accidental mistakes. Propagating the required erasure mechanisms to backups, empower users and financial institutions to manipulate data integrity according to their needs, like hiding transactions from audit controls when deemed necessary. Depending on the organisation policies and legal framework, user data records may have to be kept in non-volatile storage. Therefore, once a user requests the deletion of his data, non-automated, and –contrary to the legal framework within the institution operates– actions have to be performed, leading to additional costs and possible legal deadlocks. Such issues may become more evident in financial institutions where records must always follow the information reliability, integrity and transparency principles.

The fact that GDPR's enforcement coincides with the integration of the Payment Services Directive 2 (PSD2)[49] perplexes further the issues for financial institutions. In this regard, Account Servicing Payment Service Providers (ASPSP) (e.g. credit institutions, banks) have to allow Third Party Payment Service Providers (TPPs) (e.g. a Payment Initiation Service Provider (PISP) or an Account Information Service Provider (AISP)) to access the payment account of a Payment Service User (PSU). On the one hand, the requirement for Strong Customer Authentication, dictated under the Article 97 of PSD2 and the Article 27(3)(a) of the Draft Regulatory Technical Standards by the European Banking Authority (EBA), arguably require ample user data and device fingerprinting methods for its implementation. On the other hand however, this intense data collection and processing may risk GDPR compliance by challenging its forgetting obligations.

Whilst, over the past few years there has been a surge in the use of biometric authentication which eliminates the need to remember passwords, this convenience is the biggest advantage and, at the same time, the Achilles' heel of these methods. While users do not need to remember anything and can use their fingerprints, iris, gait etc. to authenticate, they cannot replace these biometrics once they are lost. To this end, there is an increasing need for privacy preserving schemes that will protect users' privacy. However, biometric measurements are subject to noise in the sense that each time a measurement is made some alterations are expected to occur, making thus each measurement distinct and different from its stored template. These alterations, stemming from motion blurring, divergences in luminosity, angle or other crucial factors, render the traditional cryptographic methods for private equality testing useless. To cater for this deficiency, many protocols investigating the concept of *Privacy-Preserving Biometric Authentication* have emerged recently [176, 177, 178, 179, 180, 181]. They mainly exploit properties of partial and somewhat homomorphic encryption to hide the biometric measurements of the user which is to be authenticated and allow only matching operations against the template measurement to be performed. Nevertheless, these methods require the user authentication entity to have a stored copy of users' biometric measurements, a highly sensitive piece of information that cannot be forgotten when is to be used for authentications purposes. The use of cancelable biometrics, where a biohash and not the actual measurement is provided to the authentication entity [182, 183, 184], might be proved a reasonable solution to the problem of implementing forgetting under the GDPR obligations. Nevertheless, cancelable biometrics beat the principal purpose of using biometrics in the first place as users have to either carry an additional tag or remember a password.

Another tension arising from the alignment of current services with the GDPR, is the conflict between modern device interfaces, which are very intuitive and tailored to user needs, and the regulation's requirements for implementing a simple, specific, unambiguous and, in the case of sensitive data, informed consent. As

[49] https://ec.europa.eu/info/law/payment-services-psd-2-directive-eu-2015-2366_en

currently a vast amount of individuals' information is collected in order to personalize user experience, the GDPR, by obliging users to consent to all and every piece of their identifying information when their personal data are collected, may hinder this option. Apart from the informed case, the consent requirements in the DPD were very similar to those proposed by the GDPR. Yet, the sanctions now imposed by the regulation constitutes these requirements not only mandatory to all operating data controllers, but also remarkably costly if not implemented. Conforming however, to the GDPR's consent strict requirements may be proved not only extremely cumbersome in terms of user experience, but highly critical for the quality of the service if the relevant personal information is not provided. Notwithstanding this contradiction, recent research efforts for providing informed notices in user friendly and meaningful design choices while conforming to data protection legislations seem to be particularly promising towards overcoming this obstacle [185].

The above concerns are some indicative examples of the long list of conflicts needs to be resolved for the successful GDPR application. By all means, the issues arising from the data protection provisions introduced by the regulation, and in particular the right to withdraw consent and the RtbF, are not trivial. In fact, they lie at the heart of intensively discussed and disputed areas since the GDPR's enforcement is due to overlap with other rights and business practices.

Scholars and policy makers have extensively argued over the way that the rights to privacy, to data protection, to freedom of expression, and to be informed should be balanced in the online world under the GDPR regime and, although in principle privacy and freedom of expression have equal weight in Europe, balancing privacy-related and freedom of expression-related interests will always remain difficult [5]. To quote O'Hara and Shadbolt words [36] "*The point about privacy is that it raises hard cases; people want privacy for perfectly good reasons, and others want information for equally good reasons*".

Elaborating further, while in our digital era, data, and mainly personal data, represent not only money but also power whose exercise may affect society and individuals in an unprecedented way, this fragile and complex balance between individual rights and collective knowledge should not be exclusively entrusted to market dynamics. Instead, regulators and public authorities should always ensure a level playing field between consumers and businesses in order to guarantee the respect of the fundamental rights of individuals and the freedom of expression [137, 147]. Yet, while regulating the Internet is challenging and it is important that laws do not unduly infringe freedoms or deter innovation [147], some scholars assert that the utopian ideal of cyberspace needs to yield to human reality, and therefore the regulation of the Internet's excesses is necessary in order to gain the benefits of its substantial breakthroughs and prevent privacy harms, even though in doing so it may need to sacrifice, at least a little, important counter values, like innovation, free speech, and security [30, 186].

In this respect, we recognize that regulating the right to be forgotten and the right for withdrawing consent under the GDPR is a step towards the right direction. Nonetheless, it should not be seen as a cure-all, as Lindsay underlines in [147], or a silver bullet that will apply immediately to all domains and solve all the problems. Rather, we argue that in order the provisions to be put in place successfully and to cater for all pragmatic exceptions and complexities, while at the same time a demonstrable compliance across all involved parties to be possible, low-level implementation guidelines and business-wide requirements modelling is necessary. To this end, we firmly believe that use case specific recommendations and technology-agnostic technical standards must be provided by formal legal and technical European bodies. It is our understanding and expectation that the EDPB, in coordination with the national data protection authorities, should take appropriate measures and initiate work on that direction in the immediate future. Still, we would like to stress that any policies and roadmaps proposed and legislated for regulating and balancing the forgetting requirements specified in the GDPR should need strong support from both the business and academic community in order to be widely adopted.

Taking all the above into consideration, we proposed and discussed here some frameworks, methods and architectures that can be used in pursuing the GDPR's requirements for revoking consent and permanently deleting widely disseminated personal data. Although these methods constitute an indicative only sample of modern state-of-the-art technologies that can be adopted for enforcing the RtbF in digital ecosystems, they signify undoubtedly the potentiality and feasibility of implementing forgetting in modern IT systems. Whether these approaches will apply successfully into the digital world, it remains to be seen.

ACKNOWLEDGEMENTS

This work was supported by the European Commission under the Horizon 2020 Programme (H2020), as part of the *OPERANDO* project (Grant Agreement no. 653704).

REFERENCES

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Official Journal of the European Union, L 119 (4 May 2016), pp. 1-88.

[2] 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, Official Journal of the European Union, L 281, 23/11/1995, pp. 3150.

[3] Abramatic, J.-F., Bellamy, B., Callahan, M., Cate, F., van Eecke, P., van Eijk, N., Guild, E., de Hert, P., Hustinx, P., Kuner, C., et al. (2015) Privacy Bridges: EU and US Privacy Experts In Search of Transatlantic Privacy Solutions. Technical report. University of Amsterdam.

[4] Gellert, R. and Gutwirth, S. (2013) The legal construction of privacy and data protection. *Computer Law & Security Review*, **29**, 522–530.

[5] Kulk, S. and Zuiderveen Borgesius, F. J. (2017) Privacy, Freedom of Expression, and the Right to Be Forgotten in Europe. In Polonetsky, J., Tene, O., and Selinger, E. (eds.), *Cambridge Handbook of Consumer Privacy*. Cambridge University Press.

[6] de Andrade, N. N. G. (2014) Oblivion: the right to be different from oneself: re-proposing the right to be forgotten. *The Ethics of Memory in a Digital Age*, pp. 65–81. Springer.

[7] Charter of Fundamental Rights of the European Union, 2012/C 326/02.

[8] McDermott, Y. (2017) Conceptualising the right to data protection in an era of Big Data. *Big Data & Society*, **4**, 2053951716686994.

[9] Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR).

[10] Warren, S. D. and Brandeis, L. D. (1890) The right to privacy. *Harvard law review*, **4**, 193–220.

[11] Introna, L. D. (1997) Privacy and the computer: why we need privacy in the information society. *Metaphilosophy*, **28**, 259–275.

[12] Hughes, E. (1997) A cypherpunk's manifesto. *The electronic privacy papers*, pp. 285–287. John Wiley & Sons, Inc.

[13] Post, R. C. (2000) Three concepts of privacy. *Geo. LJ*, **89**, 2087.

[14] Solove, D. J. (2006) A taxonomy of privacy. *University of Pennsylvania law review*, **154**, 477–564.

[15] Ohm, P. (2012) The Fourth Amendment in a World Without Privacy. *Mississippi Law Journal*, **81**, 1309–1355.

[16] Brin, D. (1999) *The transparent society: Will technology force us to choose between privacy and freedom?* Basic Books.

[17] Shenk, D. (1998) *Data smog: Surviving the information glut*. Harper San Francisco.

[18] Whitley, E. A. (2009) Informational privacy, consent and the "control" of personal data. *Information security technical report*, **14**, 154–159.

[19] Ausloos, J. (2012) The "right to be forgotten"–worth remembering? *Computer Law & Security Review*, **28**, 143–152.

[20] Data Protection Act 1998. [online] Available at: http://www.legislation.gov.uk/ukpga/1998/29/contents [Accessed 15 Jun. 2017].

[21] Data Protection Act 1988 (Data Protection Act 2003, as amended). Available at: http://www.irishstatutebook.ie/eli/1988/act/25/enacted/en/html [Accessed 15 Jun. 2017].

[22] Schwab, K., Marcus, A., Oyola, J., Hoffman, W., and Luzi, M. (2011) Personal data: The emergence of a new asset class. *An Initiative of the World Economic Forum*.

[23] Spiekermann, S., Acquisti, A., Böhme, R., and Hui, K.-L. (2015) The challenges of personal data markets and privacy. *Electronic Markets*, **25**, 161–167.

[24] Acquisti, A., Taylor, C., and Wagman, L. (2016) The economics of privacy. *Journal of Economic Literature*, **54**, 442–492.

[25] Narayanan, A. and Shmatikov, V. (2010) Myths and fallacies of personally identifiable information. *Communications of the ACM*, **53**, 24–26.

[26] Egele, M., Kruegel, C., Kirda, E., and Vigna, G. (2011) PiOS: Detecting Privacy Leaks in iOS Applications. *NDSS*, pp. 177–183.

[27] Sweeney, L. (2000) Simple demographics often identify people uniquely. *Health (San Francisco)*, **671**, 1–34.

[28] Gymrek, M., McGuire, A. L., Golan, D., Halperin, E., and Erlich, Y. (2013) Identifying personal genomes by surname inference. *Science*, **339**, 321–324.

[29] Narayanan, A. and Shmatikov, V. (2008) Robust de-anonymization of large sparse datasets. *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pp. 111–125. IEEE.

[30] Ohm, P. (2009) Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, **57**.

[31] Horvitz, E. and Mulligan, D. (2015) Data, privacy, and the greater good. *Science*, **349**, 253–255.

[32] Politou, E., Alepis, E., and Patsakis, C. (2017) A survey on mobile affective computing. *Computer Science Review*, **25**, 79–100.

[33] Oostveen, M. and Irion, K. (2016) The Golden Age of Personal Data: How to Regulate an Enabling Fundamental Right? *Personal Data in Competition, Consumer Protection and IP Law - Towards a Holistic Approach?* Springer.

[34] Bettini, C. and Riboni, D. (2015) Privacy protection in pervasive systems: State of the art and technical challenges. *Pervasive and Mobile Computing*, **17**, 159–174.

[35] Mittelstadt, B. D. and Floridi, L. (2016) The ethics of big data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics*, **22**, 303–341.

[36] O'Hara, K. and Shadbolt, N. (2014) *The spy in the coffee machine: the end of privacy as we know it*. Oneworld Publications.

[37] Yu, S. (2016) Big privacy: Challenges and opportunities of privacy study in the age of big data. *IEEE access*, **4**, 2751–2763.

[38] Sweeney, L. (2002) k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **10**, 557–570.

[39] Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007) l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, **1**, 3.

[40] Li, N., Li, T., and Venkatasubramanian, S. (2007) t-closeness: Privacy beyond k-anonymity and l-diversity. *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*, pp. 106–115. IEEE.

[41] Dwork, C. (2006) Differential privacy. *Automata, Languages and Programming. ICALP 2006. Lecture Notes in Computer Science, vol 4052.*, pp. 1–12. Springer, Berlin, Heidelberg.

[42] Li, Q., Cao, G., and La Porta, T. F. (2014) Efficient and privacy-aware data aggregation in mobile sensing. *IEEE Transactions on dependable and secure computing*, **11**, 115–129.

[43] Bakken, D. E., Rarameswaran, R., Blough, D. M., Franz, A. A., and Palmer, T. J. (2004) Data obfuscation: Anonymity and desensitization of usable data sets. *IEEE Security & Privacy*, **2**, 34–41.

[44] Langheinrich, M. (2001) Privacy by designprinciples of privacy-aware ubiquitous systems. *International conference on Ubiquitous Computing*, pp. 273–291. Springer.

[45] Cavoukian, A. (2011). Privacy by design-the 7 foundational principles (2011).

[46] Edwards, L. (2016) Privacy, security and data protection in smart cities: A critical eu law perspective. *Eur. Data Prot. L. Rev.*, **2**, 28.

[47] Tene, O. and Polonetsky, J. (2012) Big data for all: Privacy and user control in the age of analytics. *Nw. J. Tech. & Intell. Prop.*, **11**, xxvii.

[48] Daries, J. P., Reich, J., Waldo, J., Young, E. M., Whittinghill, J., Ho, A. D., Seaton, D. T., and Chuang, I. (2014) Privacy, anonymity, and big data in the social sciences. *Communications of the ACM*, **57**, 56–63.

[49] Rubinstein, I. S. (2013) Big data: The end of privacy or a new beginning? *International data privacy law*, **3**, 74–87.

[50] Hemerly, J. (2013) Public policy considerations for data-driven innovation. *Computer*, **46**, 25–31.

[51] D'Acquisto, G., Domingo-Ferrer, J., Kikiras, P., Torra, V., de Montjoye, Y.-A., and Bourka, A. (2015) Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics. *arXiv preprint arXiv:1512.06000*, **abs/1512.06000**.

[52] de Hert, P. and Papakonstantinou, V. (2016) The new General Data Protection Regulation: Still a sound system for the protection of individuals? *Computer Law & Security Review*, **32**, 179–194.

[53] Gleibs, I. H. (2014) Turning virtual public spaces into laboratories: Thoughts on conducting online field studies using social network sites. *Analyses of Social Issues and Public Policy*, **14**, 352–370.

[54] Reynolds, P. D. (1979) Ethical dilemmas and social science research. , **?**

[55] Hofmann, B. (2009) Broadening consent - and diluting ethics? *Journal of Medical Ethics*, **35**, 125–129.

[56] Ioannidis, J. P. (2013) Informed consent, big data, and the oxymoron of research that is not research. *The American Journal of Bioethics*, **13**, 40–42.

[57] Rothstein, M. A. and Shoben, A. B. (2013) An unbiased response to the open peer commentaries on "does consent bias research?". *The American Journal of Bioethics*, **13**, W1–W4.

[58] Stevenson, F., Lloyd, N., Harrington, L., and Wallace, P. (2012) Use of electronic patient records for research: views of patients and staff in general practice. *Family practice*, **30**, 227–232.

[59] Sheehan, M. (2011) Can broad consent be informed consent? *Public Health Ethics*, **4**, 226–235.

[60] Steinsbekk, K. S., Myskja, B. K., and Solberg, B. (2013) Broad consent versus dynamic consent in biobank research: Is passive participation an ethical problem? *European Journal of Human Genetics*, **21**, 897–902.

[61] Simon, C. M., L'heureux, J., Murray, J. C., Winokur, P., Weiner, G., Newbury, E., Shinkunas, L., and Zimmerman, B. (2011) Active choice but not too active: public perspectives on biobank consent models. *Genetics in Medicine*, **13**, 821–831.

[62] Katz, J. (1994) Informed consent-must it remain a fairy tale. *Journal of Contemporary Health Law and Policy*, **10**, 69–91.

[63] Brown, B., Weilenmann, A., McMillan, D., and Lampinen, A. (2016) Five provocations for ethical hci research. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 852–863. ACM.

[64] Hayden, E. C. (2012) A broken contract. *Nature*, **486**, 312–314.

[65] Mostert, M., Bredenoord, A. L., Biesaart, M. C., and van Delden, J. J. (2015) Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. *European Journal of Human Genetics*, **2**, 956–960.

[66] Bernal, P. (2010) Collaborative consent: Harnessing the strengths of the internet for consent in the online environment. *International Review of Law, Computers & Technology*, **24**, 287–297.

[67] Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., and Melham, K. (2015) Dynamic consent: a patient interface for twenty-first century research networks. *European Journal of Human Genetics*, **23**, 141–146.

[68] Ploug, T. and Holm, S. (2015) Meta consent: a flexible and autonomous way of obtaining informed consent for secondary research. *BMJ: British Medical Journal*, **350**.

[69] Barocas, S. and Nissenbaum, H. (2014) Big data's end run around procedural privacy protections. *Communications of the ACM*, **57**, 31–33.

[70] Cate, F. H. and Mayer-Shönberger, V. (2013) Notice and consent in a world of Big Data. *International Data Privacy Law*, **3**, 67–73.

[71] Luger, E. and Rodden, T. (2013) An informed view on consent for UbiComp. *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pp. 529–538. ACM.

[72] Morrison, A., McMillan, D., and Chalmers, M. (2014) Improving consent in large scale mobile hci through personalised representations of data. *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, pp. 471–480. ACM.

[73] Curren, L. and Kaye, J. (2010) Revoking consent: A 'blind spot' in data protection law? *Computer law & Security review*, **26**, 273–283.

[74] Benford, S., Greenhalgh, C., Anderson, B., Jacobs, R., Golembewski, M., Jirotka, M., Stahl, B. C., Timmermans, J., Giannachi, G., Adams, M., et al. (2015) The ethical implications of hci's turn to the cultural. *ACM Transactions on Computer-Human Interaction (TOCHI)*, **22**, 24.

[75] Kaye, J. (2012) The tension between data sharing and the protection of privacy in genomics research. *Annual review of genomics and human genetics*, **13**, 415–431.

[76] Holm, S. (2011) Withdrawing from research: a rethink in the context of research biobanks. *Health Care Analysis*, **19**, 269.

[77] Parry, O. and Mauthner, N. S. (2004) Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data. *sociology*, **38**, 139–152.

[78] Kramer, A. D., Guillory, J. E., and Hancock, J. T. (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, **111**, 8788–8790.

[79] Jouhki, J., Lauk, E., Penttinen, M., Sormanen, N., and Uskali, T. (2016) Facebook's emotional contagion experiment as a challenge to research ethics. *Media and Communication*, **4**, 75–85.

[80] Schroeder, R. (2014) Big Data and the brave new world of social media research. *Big Data & Society*, **1**, 2053951714563194.

[81] Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., and Fowler, J. H. (2012) A 61-million-person experiment in social influence and political mobilization. *Nature*, **489**, 295–298.

[82] Kirkegaard, E. O. and Bjerrekær, J. D. (2016) The OKCupid dataset: A very large public dataset of dating site users. *Open Differential Psychology*, **46**.

[83] Zimmer, M. (2010) "But the data is already public": on the ethics of research in Facebook. *Ethics and information technology*, **12**, 313–325.

[84] Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., and Christakis, N. (2008) Tastes, ties, and time: A new social network dataset using Facebook.com. *Social networks*, **30**, 330–342.

[85] Brown, I., Brown, L., and Korff, D. (2010) Using NHS patient data for research without consent. *Law, Innovation and Technology*, **2**, 219–258.

[86] Pelliccia, F. and Rosano, G. (2014) Medical research could soon be jeopardized by new european union data protection regulations. *European heart journal*, **35**, 1503–1504.

[87] Rosano, G., Pelliccia, F., Gaudio, C., and Coats, A. J. (2014) The challenge of performing effective medical research in the era of healthcare data protection. *International journal of cardiology*, **177**, 510–511.

[88] Quinn, P., Habbig, A.-K., Mantovani, E., and De Hert, P. (2013) The Data Protection and Medical Device Frameworks-Obstacles to the Deployment of mHealth across Europe? *European journal of health law*, **20**, 185–204.

[89] Ploem, M., Essink-Bot, M., and Stronks, K. (2013) Proposed EU data protection regulation is a threat to medical research. *BMJ*, **346**, f3534.

[90] Rumbold, J. M. M. and Pierscionek, B. (2017) The effect of the General Data Protection Regulation on medical research. *Journal of medical Internet research*, **19**.

[91] Lee, P. and Pickering, K. (2016) The General Data Protection Regulation: A myth-buster. *Journal of Data Protection & Privacy*, **1**, 28–32.

[92] Bartolini, C. and Siry, L. (2016) The right to be forgotten in the light of the consent of the data subject. *Computer Law & Security Review*, **32**, 218–237.

[93] Article 29 Data Protection Working Party, "Opinion 15/2011 on the definition of consent" WP 187. `http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp187_en.pdf`.

[94] Vayena, E., Mastroianni, A., and Kahn, J. (2013) Caught in the web: informed consent for online health research. *Sci Transl Med*, **5**, 173fs6.

[95] Pöhls, H. C. (2008) Verifiable and revocable expression of consent to processing of aggregated personal data. *International Conference on Information and Communications Security*, pp. 279–293. Springer.

[96] Whitley, E. A. and Kanellopoulou, N. (2012) Privacy and informed consent in online interactions: Evidence from expert focus groups. *International Conference on Information Systems (ICIS)*. Association for Information Systems.

[97] Kaye, J., Curren, L., Anderson, N., Edwards, K., Fullerton, S. M., Kanellopoulou, N., Lund, D., MacArthur, D. G., Mascalzoni, D., Shepherd, J., et al. (2012) From patients to partners: participant-centric initiatives in biomedical research. *Nature Reviews Genetics*, **13**, 371–376.

[98] Pearson, S. and Casassa-Mont, M. (2011) Sticky policies: an approach for managing privacy across multiple parties. *Computer*, **44**, 60–68.

[99] Karjoth, G., Schunter, M., and Waidner, M. (2002) Platform for enterprise privacy practices: Privacy-enabled management of customer data. *International Workshop on Privacy Enhancing Technologies*, pp. 69–84. Springer.

[100] Mont, M. C., Pearson, S., and Bramhall, P. (2003) Towards accountable management of identity and privacy: Sticky policies and enforceable tracing services. *Database and Expert Systems Applications, 2003. Proceedings. 14th International Workshop on*, pp. 377–382. IEEE.

[101] Stuntz, C. (2010) What is Homomorphic Encryption, and Why Should I Care. *Craig Stuntz Weblog, March*, **18**.

[102] Gentry, C. et al. (2009) Fully homomorphic encryption using ideal lattices. *STOC*, pp. 169–178.

[103] Micciancio, D. (2010) A first glimpse of cryptography's holy grail. *Communications of the ACM*, **53**, 96–96.

[104] Urquhart, L. and Rodden, T. (2017) New directions in information technology law: learning from human–computer interaction. *International Review of Law, Computers & Technology*, **31**, 150–169.

[105] Le Métayer, D. and Monteleone, S. (2009) Automated consent through privacy agents: Legal requirements and technical architecture. *Computer law & Security review*, **25**, 136–144.

[106] Spiekermann, S. and Novotny, A. (2015) A vision for global privacy bridges: technical and legal measures for international data markets. *Computer Law & Security Review*, **31**, 181–200.

[107] Rooksby, J., Asadzadeh, P., Morrison, A., McCallum, C., Gray, C., and Chalmers, M. (2016) Implementing ethics for a mobile app deployment. *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, pp. 406–415. ACM.

[108] Maler, E. (2015) Extending the Power of Consent with User-Managed Access: A Standard Architecture for Asynchronous, Centralizable, Internet-Scalable Consent. *Security and Privacy Workshops (SPW), 2015 IEEE*, pp. 175–179. IEEE.

[109] Lizar, M. and Turner, D. Consent Receipt Specification, version 1.0.0. `http://contact.kantarainitiative.org/comment/KI-CR1_0_0.pdf`.

[110] Styliari, T. C. and Nati, M. (2016) Researching the transparency of personal data sharing: designing a concert receipt. *Digital Catapult* , **?**

[111] Bannon, L. J. (2006) Forgetting as a feature, not a bug: the duality of memory and implications for ubiquitous computing. *CoDesign*, **2**, 3–15.

[112] Tirosh, N. (2015) Reconsidering the "Right to be Forgotten"–memory rights and the right to memory in the new media era. *Media, Culture & Society*, **39**.

[113] Connerton, P. (2008) Seven types of forgetting. *Memory studies*, **1**, 59–71.

[114] Ricoeur, P. (2004) *Memory, history, forgetting*. University of Chicago Press.

[115] Volf, M. (2006) *The end of memory: Remembering rightly in a violent world*. Wm. B. Eerdmans Publishing.

[116] Nietzsche, F. (1874) *On the use and abuse of history for life*.

[117] Mayer-Shönberger, V. (2011) *Delete: The virtue of forgetting in the digital age*. Princeton University Press.

[118] Parker, E. S., Cahill, L., and McGaugh, J. L. (2006) A case of unusual autobiographical remembering. *Neurocase*, **12**, 35–49.

[119] Borges, J. L. (1954) *Funes, the memorious* Avon Modern Writing No. 2. Avon Books.

[120] Blanchette, J.-F. and Johnson, D. G. (2002) Data retention and the panoptic society: The social benefits of forgetfulness. *The Information Society*, **18**, 33–45.

[121] Allen, A. L. (2008) Dredging up the past: Lifelogging, memory, and surveillance. *The University of Chicago Law Review*, **75**, 47–74.

[122] Hand, M. (2016) Persistent traces, potential memories: Smartphones and the negotiation of visual, locative, and textual data in personal life. *Convergence*, **22**, 269–286.

[123] Burkell, J. A. (2016) Remembering me: big data, individual identity, and the psychological necessity of forgetting. *Ethics and Information Technology*, **18**, 17–23.

[124] Dodge, M. and Kitchin, R. (2007) "Outlines of a world coming into existence": pervasive computing and the ethics of forgetting. *Environment and planning B: planning and design*, **34**, 431–445.

[125] Bentham, J. (1791) *Panopticon or the inspection house*. Payne, London.

[126] Gorzeman, L. and Korenhof, P. (2017) Escaping the Panopticon Over Time. *Philosophy & Technology*, **30**, 73–92.

[127] Rosen, J. The web means the end of forgetting, 2010. `http://www.nytimes.com/2010/07/25/magazine/25privacy-t2.html`.

[128] Solove, D. J. (2007) *The future of reputation: Gossip, rumor, and privacy on the Internet*. Yale University Press.

[129] Hendler, J. (2009) Web 3.0 Emerging. *Computer*, **42**.

[130] Bizer, C., Heath, T., and Berners-Lee, T. (2009) Linked data-the story so far. *Semantic services, interoperability and web applications: emerging concepts*, pp. 205–227.

[131] Gurrin, C., Lee, H., and Hayes, J. (2010) iForgot: a model of forgetting in robotic memories. *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pp. 93–94. IEEE.

[132] Sas, C. and Whittaker, S. (2013) Design for forgetting: disposing of digital possessions after a breakup. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1823–1832. ACM.

[133] Kulk, S. and Borgesius, F. Z. (2014) Google Spain v. González: Did the court forget about freedom of expression. *Eur. J. Risk Reg.*, **5**, 389.

[134] Mayer-Shönberger, V. (2014) Omission of search results is not a "right to be forgotten" or the end of google. *The Guardian*, **13**.

[135] O'Hara, K. (2015) The right to be forgotten: The good, the bad, and the ugly. *IEEE Internet Computing*, **19**, 73–79.

[136] Baum, R. M. (2014). It's Not Censorship. `http://cen.gext.acs.org/articles/92/i22/s-Censorship.html`.

[137] Mantelero, A. (2013) The EU Proposal for a General Data Protection Regulation and the roots of the "right to be forgotten". *Computer Law & Security Review*, **29**, 229–235.

[138] Voss, W. G. and Castets-Renard, C. (2016) Proposal for an International Taxonomy on the Various Forms of the "Right to Be Forgotten": A Study on the Convergence of Norms. *Colorado Technology Law Journal*, **14**, 281–344.

[139] Xanthoulis, N. (2013) The Right to Oblivion in the Information Age: A Human-Rights Based Approach. *US-China L. Rev.*, **10**, 84.

[140] Koops, B.-J. (2011) Forgetting footprints, shunning shadows: A critical analysis of the "right to be forgotten" in big data practice. *SCRIPTed*, **8**.

[141] European Data Protection Supervisor, "Opinion of the EDPS on the data protection reform package". `https://edps.europa.eu/sites/edp/files/publication/12-03-07_edps_reform_package_en.pdf`.

[142] Rosen, J. (2011) The right to be forgotten. *Stan. L. Rev. Online*, **64**, 88.

[143] Malle, B., Kieseberg, P., Weippl, E., and Holzinger, A. (2016) The right to be forgotten: towards machine learning on perturbed knowledge bases. *International Conference on Availability, Reliability, and Security*, pp. 251–266. Springer.

[144] Stuart, A. H. (2013) Google search results: buried if not forgotten. *NCJL & Tech.*, **15**, 463.

[145] Nunziato, D. C. (2005) The death of the public forum in cyberspace. *Berkeley Technology Law Journal*, **20**, 1115–1757.

[146] Mitrou, L. and Karyda, M. (2012) EU's data protection reform and the right to be forgotten: A legal response to a technological challenge? *5th International Conference of Information Law and Ethics 2012*.

[147] Lindsay, D. (2012). The "right to be forgotten" is not censorship. http://www.monash.edu/news/opinions/the-right-to-be-forgotten-is-not-censorship.

[148] Korenhof, P. (2013) Forgetting Bits and Pieces: An Exploration of the Right to be forgotten in Online Memory Process. *Tilburg Institute for Law and Technology Working Paper Series*, **4**, 6.

[149] Ambrose, M. L. (2014) Speaking of forgetting: Analysis of possible non-EU responses to the right to be forgotten and speech exception. *Telecommunications Policy*, **38**, 800–811.

[150] Bennett, S. C. (2012) The right to be forgotten: Reconciling EU and US perspectives. *Berkeley J. Int'l L.*, **30**, 161.

[151] Kadenic, V. (2015) Compliance of Data Lake Enterprise Architecture Model with the General Data Protection Regulation (GDPR). Bachelor thesis Luleå University of Technology.

[152] O'Hara, K., Shadbolt, N., and Hall, W. (2016) A pragmatic approach to the right to be forgotten.

[153] Barua, D., Kay, J., Kummerfeld, B., and Paris, C. (2011) Theoretical foundations for user-controlled forgetting in scrutable long term user models. *Proceedings of the 23rd Australian Computer-Human Interaction Conference*, pp. 40–49. ACM.

[154] Novotny, A. and Spiekermann, S. (2014) Oblivion on the web: an inquiry of user needs and technologies. *Twenty Second European Conference on Information Systems, Tel Aviv*.

[155] Hong, J. I. and Landay, J. A. (2004) An architecture for privacy-sensitive ubiquitous computing. *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pp. 177–189. ACM.

[156] Ashley, P., Hada, S., Karjoth, G., Powers, C., and Schunter, M. (2003). Enterprise privacy authorization language (epal).

[157] Langheinrich, M. (2002) A privacy awareness system for ubiquitous computing environments. *international conference on Ubiquitous Computing*, pp. 237–245. Springer.

[158] Perlman, R. (2005) File system design with assured delete. *Security in Storage Workshop, 2005. SISW'05. Third IEEE International*, pp. 6–pp. IEEE.

[159] Tang, Y., Lee, P. P., Lui, J. C., and Perlman, R. (2012) Secure overlay cloud storage with access control and assured deletion. *IEEE Transactions on dependable and secure computing*, **9**, 903–916.

[160] Bajaj, S. and Sion, R. (2013) Ficklebase: Looking into the future to erase the past. *Data Engineering (ICDE), 2013 IEEE 29th International Conference on*, pp. 86–97. IEEE.

[161] Korenhof, P., Ausloos, J., Szekely, I., Ambrose, M., Sartor, G., and Leenes, R. (2015) Timing the right to be forgotten: A study into "time" as a factor in deciding about retention or erasure of data. *Reforming European data protection law*, pp. 171–201. Springer.

[162] Lee, H. J., Yun, J. H., Yoon, H. S., and Lee, K. H. (2015) The right to be forgotten: Standard on deleting the exposed personal information on the internet. *Computer science and its applications*, pp. 883–889. Springer.

[163] Anciaux, N., Bouganim, L., Van Heerde, H., Pucheral, P., and Apers, P. M. (2008) Data degradation: Making private data less sensitive over time. *Proceedings of the 17th ACM conference on Information and knowledge management*, pp. 1401–1402. ACM.

[164] Geambasu, R., Kohno, T., Levy, A. A., and Levy, H. M. (2009) Vanish: Increasing Data Privacy with Self-Destructing Data. *USENIX Security Symposium*, pp. 299–316.

[165] Wolchok, S., Hofmann, O. S., Heninger, N., Felten, E. W., Halderman, J. A., Rossbach, C. J., Waters, B., and Witchel, E. (2010) Defeating Vanish with Low-Cost Sybil Attacks Against Large DHTs. *NDSS*.

[166] Geambasu, R., Falkner, J., Gardner, P., Kohno, T., Krishnamurthy, A., and Levy, H. M. (2009). Experiences building security applications on DHTs.

[167] Zeng, L., Shi, Z., Xu, S., and Feng, D. (2010) Safevanish: An improved data self-destruction for protecting data privacy. *Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on*, pp. 521–528. IEEE.

[168] Xiong, J., Liu, X., Yao, Z., Ma, J., Li, Q., Geng, K., and Chen, P. S. (2014) A secure data self-destructing scheme in cloud computing. *IEEE Transactions on Cloud Computing*, **2**, 448–458.

[169] Wang, G., Yue, F., and Liu, Q. (2013) A secure self-destructing scheme for electronic data. *Journal of Computer and System Sciences*, **79**, 279–290.

[170] Zeng, L., Chen, S., Wei, Q., and Feng, D. (2012) Sedas: A self-destructing data system based on active storage framework. *APMRC, 2012 Digest*, pp. 1–8. IEEE.

[171] Singh, J., Powles, J., Pasquier, T., and Bacon, J. (2015) Data flow management and compliance in cloud computing. *IEEE Cloud Computing*, **2**, 24–32.

[172] Bacon, J., Eyers, D., Pasquier, T. F.-M., Singh, J., Papagiannis, I., and Pietzuch, P. (2014) Information flow control for secure cloud computing. *IEEE Transactions on Network and Service Management*, **11**, 76–89.

[173] Enck, W., Gilbert, P., Han, S., Tendulkar, V., Chun, B.-G., Cox, L. P., Jung, J., McDaniel, P., and Sheth, A. N. (2014) TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS)*, **32**, 5.

[174] Zyskind, G., Nathan, O., et al. (2015) Decentralizing privacy: Using blockchain to protect personal data. *Security and Privacy Workshops (SPW), 2015 IEEE*, pp. 180–184. IEEE.

[175] Maguire, S., Friedberg, J., Nguyen, M.-H. C., and Haynes, P. (2015) A metadata-based architecture for user-centered data accountability. *Electronic Markets*, **25**, 155–160.

[176] Blanton, M. and Gasti, P. (2011) Secure and efficient protocols for iris and fingerprint identification. *Computer Security–ESORICS 2011*, pp. 190–209. Springer.

[177] Shahandashti, S. F., Safavi-Naini, R., and Ogunbona, P. (2012) Private fingerprint matching. *Information Security and Privacy*, pp. 426–433. Springer.

[178] Bringer, J., Chabanne, H., and Patey, A. (2013) Practical identification with encrypted biometric data using oblivious ram. *Biometrics (ICB), 2013 International Conference on*, pp. 1–8. IEEE.

[179] Bringer, J., Favre, M., Chabanne, H., and Patey, A. (2012) Faster secure computation for biometric identification using filtering. *Biometrics (ICB), 2012 5th IAPR International Conference on*, pp. 257–264. IEEE.

[180] Blundo, C., De Cristofaro, E., and Gasti, P. (2013) EsPRESSo: efficient privacy-preserving evaluation of sample set similarity. *Data Privacy Management and Autonomous Spontaneous Security*, pp. 89–103. Springer.

[181] Patsakis, C., van Rest, J., Chorań, M., and Bouroche, M. (2015) Privacy-preserving biometric authentication and matching via lattice-based encryption. *International Workshop on Data Privacy Management*, pp. 169–182. Springer.

[182] Jin, A. T. B., Ling, D. N. C., and Goh, A. (2004) Biohashing: two factor authentication featuring fingerprint data and tokenised random number. *Pattern recognition*, **37**, 2245–2255.

[183] Ratha, N., Connell, J., Bolle, R. M., and Chikkerur, S. (2006) Cancelable biometrics: A case study in fingerprints. *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, pp. 370–373. IEEE.

[184] Teoh, A. B., Kuan, Y. W., and Lee, S. (2008) Cancellable biometrics and annotations on biohash. *Pattern recognition*, **41**, 2034–2044.

[185] Schaub, F., Balebako, R., Durity, A. L., and Cranor, L. F. (2015) A design space for effective privacy notices. *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pp. 1–17. USENIX Association.

[186] Richards, N. M. and King, J. H. (2013) Three paradoxes of big data. *Stanford Law Review Online*, **66**, 41.