OBLIVION ON THE WEB: AN INQUIRY OF USER NEEDS AND TECHNOLOGIES

Complete Research

Novotny, Alexander, Vienna University of Economics and Business, Austria, alexander.novotny@wu.ac.at

Spiekermann, Sarah, Vienna University of Economics and Business, Austria, sarah.spiekermann@wu.ac.at

Abstract

Unlimited retention of personal information on the web may harm individuals: employers can find youthful indiscretions on social media, and incorrectly low credit scores may haunt individuals for a lifetime. Currently, Europe revives the "right to erasure" as a first step towards a forgetting web. Early technologies implementing oblivion suffer from vulnerabilities and narrowly assume that users require information to be erased after a pre-determined time. But little is known about users' actual oblivion needs. A first study shows that users desire control over disclosed personal information to reduce pre-disclosure privacy concerns, and to delete harmful information after disclosure. In the long run, users have a need for dissociating from obsolete information that represents their past identity. A second study analyses whether oblivion-enhancing technologies (OETs) currently deployed in online services satisfy users' needs. While not all services give users assurance that disclosed information can be erased again, most provide users with some active control. But to manage the increasing volume of personal information stored, users would also require "intelligent" support with oblivion. Intelligent agents that keep track of disclosed information long-term could automatically safeguard users from information relating to a past episode in life surfacing unexpectedly.

Keywords: Information privacy, Oblivion-enhancing technologies, Right to be forgotten

1 Introduction

Ever since Viktor Mayer-Schönberger's famous book "Delete" (2009), the idea of a "right to be forgotten" has permeated the Internet: How should we deal with out-dated information on the Internet? Should we transfer the ephemerality of most of the offline world's information to the web? If yes, how would the web have to change?

Historically, most information has been of an ephemeral nature. Everyday transactions and momentous observations were not recorded, because archiving was laborious and expensive. For information relating to persons, forgetfulness was even socially institutionalized. Insolvency law, juvenile delinquency records, and credit reporting regulations include limits on the retention of data (Blanchette and Johnson, 2002). Emigration to start a new life, amnesty, and Catholic absolution are just some examples of how oblivion enables social forgiveness (Bannon, 2006; Szekely, 2012).

In contrast, the web is not built with the idea of oblivion in mind. Cheap data storage, easy information access, and powerful analytics tools favour information retention at the expense of oblivion (Mayer-Schönberger, 2009). The amount of personal information (PI) – "any information relating to an identified or identifiable natural person" (COM, 2012) – that is disclosed and recorded on the web is

growing (WEF, 2012). This puts the possibility to start a new life, and to become a new person, at stake (Kaiser, 2012). PI stored on the web transcends temporal borders (Marx, 2001); once PI is disclosed to an online service, it may be used at a later date for making unexpected inferences about the individual it describes. Employers run background checks using web search engines and social media before hiring job applicants (Antonopoulos, 2010). Data brokers link the many traces people leave online and preserve duplicates in databases for eternity (Couts, 2012). Victims of out-dated or inaccurate information pay reputation defender services to battle the inextinguishable memory of web archives, but often these attempts are of limited effectiveness (Cheng, 2012).

Against this background, thought leaders discuss whether the Internet should be able to forget. Privacy reasons, long-term data security issues, and companies' data liability (FIP, 2012) are putting increasing pressure on moving towards a "forgetting web", particularly with a view to PI. Envisaged ideas of a forgetting web do not include "scientific or technological information" (Szekely, 2012), not touching the web's function as a cultural and historic memory (Kaiser, 2012). Rather, proposals are limited to PI being forgotten. Taking up this discussion, the European Union's (EU) legislation has brought about a reinforced "right to erasure" as part of its 2012 draft of a general data protection regulation. In the regulation (Article 17), the European Parliament has proposed that individuals (users) should have by law (*i*) a right to erasure of their personal data and (*ii*) that erasure requests have to be passed on to any other parties owning a copy (EP, 2013).

At the same time, little is known about users' needs for oblivion in the online environment (Bannon, 2006; Karla, 2010). While users' perception of the *collection* of PI has been extensively researched (e.g., Krasnova et al., 2013; Rizk et al., 2010), there is a lack of research on users' needs regarding a forgetting web. Are users' requirements different before and after PI has been disclosed on online services? Which oblivion mechanisms are implemented on the web already, and do they fulfill the users' needs? And, should users or online services initiate the process of forgetting?

Ideally, online services should "intelligently" support users with the oblivion of disclosed PI. Intelligent oblivion mechanisms would discern different degrees of information importance, embed decision support for meaningful data deletion, and possibly commence forgetting automatically. They would also be capable of detecting incorrect personal records that users may want to erase actively and immediately. These and many other capabilities would be required to translate complex user expectations into what we call oblivion-enhancing technologies (OETs).

But so far, OETs are in its infancy. "Digital rubbers" promising the removal of PI from the web suffer from unresolved vulnerabilities, such as key harvesting and a single point of failure (Federrath et al., 2011). By only providing the binary states of deleted or not deleted, current OETs are not capable of capturing the complexity of forgetting (O'Hara, 2012). Generally, current research is missing an overview of the types of technologies that are both oblivion-enhancing and available for satisfying users' needs for oblivion.

In this paper, we want to fill these gaps by studying users' needs for oblivion in online services and exploring whether technologies on the web fulfil these needs. We conceptualize digital oblivion as a state in which the possibility for observers (e.g., other users of online services, service operators, or third parties) to retrieve or interpret PI outside of the temporal context of its disclosure is reduced (see Section 2). Our research strategy first attempts to elicit user needs and requirements for OETs and then compares these to the current state of deployed OETs on the web. The comparison's results allow us to derive recommendations as to how oblivion should be implemented into online services. Section 2 of this paper contrasts related work on digital oblivion with the human process of remembering and forgetting. In Section 3, study 1 develops a phase model of user needs for oblivion in online services and derives corresponding requirements for OETs. Study 2 (see Section 4) investigates whether the user needs and requirements found in study 1 are fulfilled by the OETs currently implemented in major online services, and relates OETs to the memory process.

2 Related Work

Departing from existing work that has investigated users' desires for oblivion in the narrow context of online social networks, we explore proposed oblivion mechanisms on the web. Ayalon and Toch (2013) show that users' willingness to share disclosed social media posts drops over time, particularly when important factors in the users' lives have changed. Users do not want to have social media posts deleted in bulk at a fixed date; their privacy preferences develop in a more nuanced fashion over time. While users perceive elder posts as less relevant, they also want to keep some of their old posts for reminiscence (Bauer et al., 2013). Zhao et al. (2013) found that PI on Facebook's timeline progresses through a lifecycle consisting of three phases. First, information serves the users' current self-presentation needs. Second, similar to a museum, the timeline's center exhibits the user's identity to others. Third, users archive PI for themselves; the visibility of PI is restricted to a limited audience.

Remembering and forgetting represent an inseparable dualism. Forgetting is the failure of remembering, a process which prevents the past from being successfully reconstructed (Bannon, 2006). Inspired by the human memory process, the "right to be forgotten" introduces an artificial barrier to the retrieval of PI (Barua et al., 2011; Korenhof, 2013). Even though humans and the web memorize information differently, we broadly take the analogy of the human mind to discuss current OETs on the web. The memory of information in the human mind and on the web takes place in three stages: encoding, storage, and retrieval:

Encoding is the process of storing perceptual information (e.g., visual, auditory) into a sensory cache (Atkinson and Shiffrin, 1968). Encoding resembles the disclosure and collection of PI. Much work has been devoted to technologies collecting PI, such as web tracking (Conger et al., 2013), or ubiquitous sensors (Krumm, 2011). Privacy-enhancing technologies (e.g., Hansen et al., 2004; Tene and Polenetsky, 2012) avoid the encoding of PI. Acknowledging the body of work on encoding PI, this paper takes a different focus – that of how the web may forget *after* PI has been disclosed.

Storage is the process of retaining information in the human brain over time (Atkinson and Shiffrin, 1968). Human long-term memories, though, may be lost altogether, details may disappear by abstraction, or they may change from the addition of new information (Draaisma, 2013, p. 41). Digital long-term memory is equally in danger of being forgotten because of changing file formats, degradation of storage media, and long-term management costs (Cumming and Findlay, 2010). Mayer-Schönberger (2009) proposes information ecology for putting an artificial constraint on storage duration making the storage of PI more costly than forgetting. By adding life expectancy (e.g., expiry dates) to information, forgetting could be re-introduced as the default (Conley, 2010; Karla, 2010). Digital rubbers such as "XPire!" (Backes:SRT, 2013) or "The ephemerizer" (Perlman, 2005) grant access to content only before a specified date. These centralized technologies, though, suffer from a single point of failure at the key server, and cannot prevent the harvesting of decryption keys before the expiry date (Federrath et al., 2011). Decentralized approaches such as "Vanish" are capable of irreversibly self-destructing information without requiring user action after an expiry date (Geambasu et al., 2009), or gradually eroding PI over time by making decryption an increasingly difficult task (Patsakis, 2012). Expiry dates, however, only allow binary states of stored and deleted. "Can we predict and express the complex social role of information in time stamps, or will the nuances resist encoding?" (O'Hara, 2012). Therefore, Sleeper et al. (2013) highlight the need for tools enabling the selective detection of regrettable PI stored in services such as Twitter.

Retrieval is the process of recalling memories from the long-term memory. The brain accesses information through association with other memories (Raaijmakers and Shiffrin, 1981). The past is never perfectly remembered, however. Remembering is a non-exact constructive act, a reconstruction of past events through the lens of the present, contaminated by alterations and omissions to memories after their encoding (Bartlett, 1932). Such failure to remember is unintended (Bannon, 2006). Mechanisms making the retrieval of PI more difficult include blocking access to PI, or moving PI to a

separate archive (Barua et al., 2011). On the web, hyperlinks to documents may be deleted (Korenhof, 2013, p. 9). Search engine optimization techniques confound retrieval in two ways. First, positive content may be optimized to conceal other, negative content (Beel et al., 2010). And second, negative content may be directly manipulated to be ranked down (Langville and Meyer, 2006).

The memory process sheds light on the characteristics constituting digital oblivion. First, digital oblivion has an *ex-post* nature, focusing on the storage and retrieval stages after PI has been disclosed. Second, PI must be *observable* by other entities, which give potential relevance to the PI. Information not receiving attention is potentially forgotten. Observers must not be confused with "attackers" who intentionally harm the user; typically, they have legitimate access to PI on the web. Third, users face *uncertainty*: they do not know whether, when, and by whom stored PI will be retrieved in the future. As PI may come unexpectedly into the spotlight of observers long after it is encoded, observers shall have reduced certainty of *retrieving or interpreting* PI outside of its temporal context. They shall not be able to retrieve information revealing a wrongful, biased, or unwarranted picture of a user. Digital oblivion is conceptualized as a *state in which the possibility for observers to retrieve or interpret PI outside of the temporal context of its disclosure is reduced.*

Two observations can be made about the reviewed technologies. First, proposed technologies predominantly focus on the storage phase. Solutions centre on the idea of erasing information after a period of time. Second, concepts such as digital rubbers and self-destructing information are not widely deployed in current online services. But, the web today does not perfectly retain information (Kaiser, 2012), as indicated by the high rate at which hyperlinks become invalid (Lawrence et al., 2001). We will now analyse users' needs for oblivion. To the best of our knowledge, no other work exists which compares the capabilities of OETs in today's online services to these needs.

3 Study 1: User Needs for Integrating OETs into Online Services

In this study, we investigate the needs of users for a forgetting web and derive OET requirements for online services. First we describe our method of analysis. Then, we present a model of user needs as well as discuss its implications for integrating OETs into online services.

3.1 Method

We conducted a lab study with 196 German-speaking Internet users (Spiekermann and Korunovska, 2014). The study's context was to investigate participants' willingness to disclose information online to ten open-ended questions on intimate and neutral subjects, such as their political engagement and opinion on creativity. Only German-speaking Internet users were selected and invited via email to the lab, as web and language skills may impact participants' online information-sharing behaviour and ability to express opinions. One of the open questions was on the idea of a forgetting Internet: "On the Internet many people are very creative today. Do you believe that the Internet should nonetheless be able to forget? What should be forgotten?" Of the 189 participants who answered the question, 45.0% were female, 55.0% were male, and their median age was 23. The sample represents German-speaking users who were raised in Austria (69.8%), and Germany (10.6%) (4.2% were from other countries, and 15.4% indicated no origin). Because of the long tradition of data privacy regulation in German-speaking countries (Greenleaf, 2012), participants' opinions were generally privacy-friendly.

For the purpose of this paper, we interpreted the answers to the question in the course of a qualitative analysis. Interpretive methods allow for an understanding of phenomena through the assigned meaning of users and are commonly applied in IS research (e.g., Hughes and Jones, 2003; Kanungo and Jain, 2009). Participants expressed needs for oblivion in 169 cases and described their motivations for a forgetting Internet. 16 observations of user needs have been excluded because users desired deletion of information other than PI, which has no impact on a person's privacy or identity. For instance, one participant named "costs of data storage" as a reason to delete arbitrary data from online services.

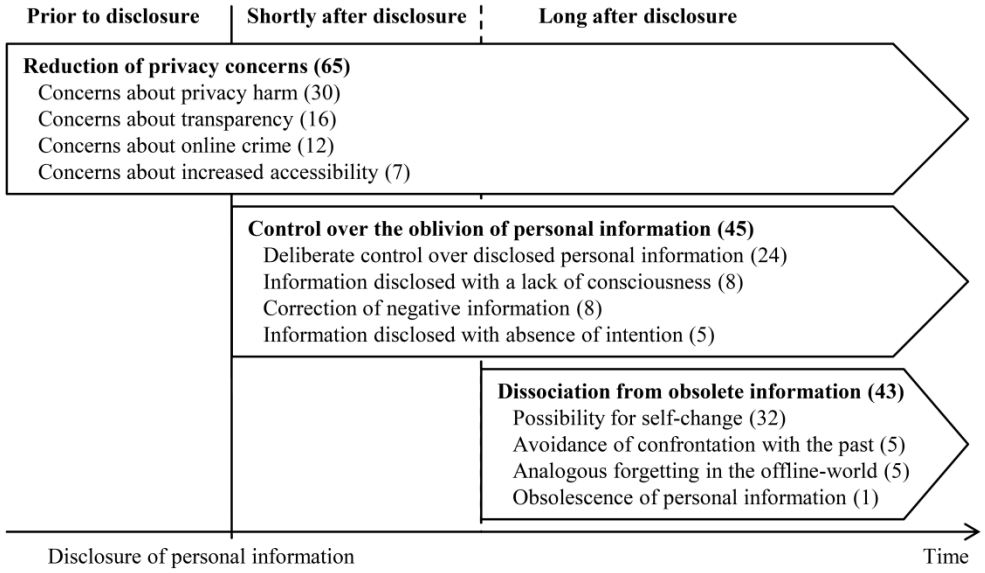

Figure 1. Needs of users for oblivion in online services (number of user need observations).

Within the remaining 153 cases, participants referred to similar needs with different labels. These were grouped based on their meaning. For example, "deletion of wrong allegations" and "possibility to correct errors" were grouped as part of a factor called "correction of negative information". Taken together, twelve distinct factors for oblivion were derived that were then grouped into three dimensions of users' oblivion needs: privacy concerns, control over disclosed PI, and dissociation from obsolete information (Figure 1 contains in brackets the number of need observations assigned to the factors). The needs expressed by participants are uniformly distributed across these three dimensions (χ^2=5.804, p=0.055), although slightly more users desired the reduction of privacy concerns. We made the additional observation that participants associated their needs with three distinct temporal phases: prior to disclosure, shortly after disclosure, and long after disclosure. All identified need dimensions could be assigned to these temporal phases.

To ensure the reliability of our interpretive procedure, we had a second analyst replicate all the steps of interpretation and grouping. While interpreting the 153 observations of user needs, disagreement emerged on 19 occasions. These were reconsidered in an audiotaped consolidation session until consensus was reached. For instance, disagreement arose about whether a participant stating that "[people] document their last alcohol consumption in pictures and employers are searching before an interview exactly for such stuff" represents a privacy concern about increased accessibility or about privacy harm. Because specific negative consequences are looming, the analysts agreed that it was a concern about privacy harm. The reasoning for the factors, the dimensions, and the resulting phase view on user needs for oblivion in online services are presented in Section 3.2.

3.2 A phase view on user needs for integrating OETs into online services

Three user need dimensions call for integrating OETs into online services: privacy concerns, control over disclosed PI, and dissociation from obsolete information (see Figure 1). These dimensions can be arranged on a temporal continuum consisting of three phases: prior to the disclosure of PI, shortly after disclosure, and long after disclosure. Before and after the disclosure of PI, online services not implementing OETs may cause privacy concerns of being exposed, experiencing privacy harm, or becoming a potential victim of PI misuse (Solove, 2006). Shortly after PI has been disclosed, users need control over the oblivion of PI, particularly when PI has been disclosed with a lack of consciousness or must be corrected. In the long-term after PI disclosure, users often do not want to be associated anymore with their information. The boundary between the phases of shortly after and long

after disclosure is fluent (dashed line in Figure 1). For instance, PI disclosed longer ago will have a different meaning to different users in different service contexts: Some will find information obsolete after a week, while others will not want to be confronted with years-old information from their past.

3.2.1 A need for oblivion due to privacy concerns

Users mentioned *privacy concerns* of disclosing PI in a web that is not capable of forgetting. The relative dominance of observations of privacy concerns (65) can be attributed to online users becoming increasingly aware of privacy issues; privacy concerns have been found to deter users from disclosing PI in online environments (Bélanger and Crossler, 2011; Krasnova et al., 2013). Despite their concerns, users still disclose PI on the web (Dinev and Hart, 2006). Their concerns extend to the phases after PI has been disclosed.

Users are concerned in all three temporal phases – about PI that will be disclosed, has been disclosed recently, and was disclosed a long time ago. "Not everything should be stored forever, as the fear of posting something on the Internet would be higher; one should have the possibility to decide on one's own whether something is stored or deleted after some period of time," a participant summarized. Concerns about *privacy harm* refer to real-world consequences that impact the user's life. Participants are concerned about a negative reputation and occupational problems. One stated that "the consequences for affected people are undeniable; some even commit suicide." *Transparency* refers to the uncomfortable feeling that others know private details of one's life (Solove, 2006). Some perceive a sense of online services storing PI of too many aspects of their life; one even expressed feeling like a "transparent individual". Others are concerned about PI being misused for *online crime*, such as in e-banking scams. *Increased accessibility* is the unexpected expansion of the audience who has access to one's PI. For instance, participants indicated that they are concerned that "future employers will comb through social networks" to gain access to PI not initially intended for them.

3.2.2 A desire for control over the oblivion of personal information

People's desire for control is deeply rooted in psychology (Langer, 1983). Our results reflect this desire: Users want *control over the oblivion of personal information* once it has been disclosed in online services. They want to actively decide what PI should be forgotten, both shortly after it has been disclosed and in the long term. "One should be able to remove things disclosed by oneself, by the time one notices it was a mistake." Participants wanted protection of their personal rights, as by removing denigratory content from online services: "I think of videos others upload to YouTube to expose people. Such videos should be deleted and made inaccessible." The need for deleting such content persists over time. It may emerge immediately after disclosure or several years later.

Participants stated that deleting is a form of *deliberate control* – an act of volition over PI. "Through the process of deleting, the will [of users] is communicated," participants highlighted. One participant said that being in control of the destruction of self-created information would make him feel good. PI is disclosed with *absence of intention* if the true will of the user on whether to disclose is not respected in the actual act of disclosing the information. Participants refer to unwanted information disclosure, to actions they committed in their childhood, and to PI that is already believed to be deleted. Often, unwillingly disclosed information has its roots in a *lack of consciousness* of its consequences: Some users "trifle with their private data" and "disclose without a thought in mind." As a result, participants ask for the *correction of negative information*. They desire mechanisms for the removal of false allegations and incorrect and immoral information.

3.2.3 A need for dissociation from obsolete information

The third dimension of oblivion needs is related to people's desire for identity construction (Eickelpasch and Rademacher, 2004). People want to update their personal identity from time to time and present themselves to others in different roles (Kaiser, 2012). This need is reflected by our results:

Users mention a need for *dissociation from obsolete information* that was disclosed a long time ago. The desire for online services to enable *self-change* emerges by the time disclosed PI does not mesh with a renewed self-concept: "Man changes constantly, his opinions change, his environment changes. Things which are of interest to a 17-year-old today may be mortifying in ten years […] Thus, there shall be a possibility to withdraw published items." Additionally, PI disclosed a long time ago could *confront users with undesired past* memories that they have already put out of their minds: "A non-forgetting Internet is definitely dangerous, as one may be confronted with some things years later, which no longer apply". Participants transferred the ephemeral nature of information from the *offline* world to the online world and argued that negative information should *analogously* be forgotten by the web. The concern is not that data is old as such, but rather that the PI is *obsolete*.

3.3 Requirements for oblivion-enhancing technologies

The users' needs for oblivion found in this study point to three requirements for oblivion-enhancing technologies in online services (see Table 1). First, online services should be capable of erasing all stored PI, free of residues. Erasing serves the user needs of 'reduction of privacy concerns' and 'control over the oblivion of PI'. It avoids ex-ante discouragement of PI disclosure and continuously assures users that disclosed information can be prevented from turning into a source for harm in the storage stage. Erasing could protect users from the privacy harm of wrong, negative, and unwillingly disclosed information. Erasure functionalities of OETs should therefore embrace all PI disclosed to the service, not only subsets of it. Information should be erased residue-free and must not remain in other storage locations, as this would give users a false sense of security.

Second, online services should provide comprehensive long-term control over storage and access to PI. Users have a need to *actively* control which PI is stored and who is able to retrieve it. Comprehensive control options over PI could reduce user concerns about transparency and about increased accessibility to PI by unintended audiences. Users' need for deliberate forgetting of PI suggests that control should be comprehensive and fine-grained. It should not only allow users to keep or erase PI, but to determine the degree of accessibility on a detailed level. In addition, control options need to be granted long-term to users, in the phases shortly and long after disclosure. To provide users with deliberate control, OETs need to be easily accessible through a simple user interface.

Third, online services should have intelligent, automated capabilities to forget outdated PI. Long after disclosure, OETs that address the need of dissociation from obsolete information are required. Users wish to analogously transfer the creeping ephemerality of information from the offline to the online world. These results suggest that users would like to have online services in which PI is *passively* forgotten after some time. Thus, online services should provide automated support to users in the process of forgetting. Indeed, previous work has predominantly taken a passive approach; proposed mechanisms deny access to PI after a fixed period of time (e.g., Geambasu et al., 2009; Karla, 2010; Patsakis, 2012). But users require more "intelligent" support for dissociation from obsolete information. Choosing the right expiry date is difficult and intelligent OETs may have to choose the date autonomously. They also should remind users of stored PI that likely does not fit their current personal identity. OETs need to automatically detect obsolete information that is no longer relevant and suggest adequate actions to users. It needs to be studied whether existing types of OETs in online services conform to these requirements derived from the user needs found in this study.

No.	Requirement to the online service
R1	Online service is capable of erasing all stored PI free of residues
R2	Online service provides comprehensive long-term control over the storage and access to PI
R3	Online service has intelligent, automated capability to forget outdated PI

Table 1. Requirements for oblivion-enhancing technologies in online services.

4 Study 2: Oblivion-enhancing Technologies in Online Services

This study explores the types of OETs available to users in online services and whether they are fulfilling the requirements found in study 1. We first describe the method of analysing the available OETs and then present those in relation to the stages of the memory process (see Section 2).

4.1 Method

To investigate available OETs, we analysed popular online services. Our selection was guided by the storage and retrieval stages of the memory process. Although online services both store and retrieve, a service's primary purpose is typically one of these. While some services focus on storing or archiving information (e.g., web archives, web logs), others are predominantly used to retrieve information from or share information with others (e.g., search engines, social networks). Initially, a set of six popular, current online services was selected based on heterogeneity of type (e.g., search engines, e-commerce sites) from within the top 30 of the Alexa Top 500 ranking (Alexa, 2013): Google Search (1), Facebook (2), Wikipedia (6), Twitter (10), Automattic WordPress (16), and Pinterest (26). We continued to include services of further types (e.g., reputation cleaning, location-based social networks) from the Alexa ranking not among the top 30, but containing additional OETs. Finally, a set of 13 services was chosen for analysis, 6 focusing on the storage and 7 on the retrieval stage (see Table 2). Investigating additional services beyond the 13 selected would be unlikely to reveal further insights, as the same types of OETs occurred repeatedly within our sample.

Online service	**Service type**	**Service's focal stage in the memory process**	
Amazon	E-commerce	Storage	Storing purchasing behaviour
Automattic WordPress	Web log		Storing author opinions and views
Internet Archive Waybackm.	Web archive		Archiving website snapshots
Miles and More	Customer loyalty		Storing transactions across many industries
Pinterest	Personal pinboard		Collecting and organizing personal interests
Wikipedia	Web encyclopaedia		Storing descriptions and claims about persons
123People	People search	Retrieval	Access multiplier to PI available on the web
Facebook	Social network		Sharing and retrieving personal content
Foursquare	Location-based network		Retrieving the position history of friends
Google Search	Web search		Making PI accessible to a broad audience
Reputation.com	Reputation cleaning		Preventing the retrieval of PI
TripAdvisor	Customer reviews		Sharing and retrieving customer opinions
Twitter	Microblogging		Sharing and retrieving personal opinions real-time

Table 2. Online services selected for analysis.

Multiple data sources have been used to elicit OETs. Terms of services, privacy policies, articles of publicly available knowledge bases provided by the services, published patents, and court decisions were analysed. Documents describing OETs were retrieved directly from the services' websites and by searching in an online patent database (Google Patents) for each service's name. In the documents, any mechanism that reduces the possibility of retrieving or interpreting PI outside of its original temporal context was considered to be an OET. For eight services, a test account was registered to validate the described OETs. The OETs have been analysed across online services; technological mechanisms that have identical effects have been captured within the same type of OET. For example, mechanisms to erase a blog post in one service and to permanently delete the account at another service have been captured as "erasing" technologies.

In study 1 we found that users' oblivion needs are driven by the desires to control their privacy and dissociate from obsolete information. These need dimensions suggest that users will both appreciate deliberate forgetting mechanisms – the possibility to *actively* control the availability of their

information – and embrace non-deliberate forgetting that occurs *passively* within online services. In our analysis, we therefore distinguished whether OETs can be triggered actively by users or are passively initiated by the online service.

4.2 Oblivion-enhancing technologies

In the analysed online services, eight types of OETs have been found (see Table 3). These are either triggered actively and deliberately by users, or passively by an automatic mechanism intrinsic to the online service. Two types of OETs, erasing and anonymising PI, address forgetting in the storage stage of the memory process. One type of OET (modifying) operates in both the storage and retrieval stages by hindering access to PI through the deletion of stored information. The remaining five types of OETs increase the costs of retrieving or interpreting PI out of its original temporal context.

Focal stage in the memory process		Storage						Retrieval						
Stage of OET in memory process	**Oblivion-enhancing technologies**	Amazon	Automattic WordPress	Internet Archive Waybackmachine	Miles and More	Pinterest	Wikipedia	123People	Facebook	Foursquare	Google Search	Reputation.com	TripAdvisor	Twitter
Storage	Erasing		A		A/P	A			A	A				A
	Anonymising								P	P				
Storage / Retrieval	Modifying		A	P		A	A				P		A	A
Retrieval	Removing	A	A	A/P		A	A	A	A	A	A/P	A	A	A/P
	Hiding	A	A			A			A	A				A
	De-referencing		A			A			A	A			A	A
	Displacing	P				P				P	P	A		P
	Time-stamping	P	P	P	P		P		P	P	P		P	P

A ... OET actively triggered by user, P ... OET passively triggered by service

Table 3. Oblivion-enhancing technologies available in the analysed online services.

4.2.1 Available OETs for users

We present the OETs in order of their capability to reduce the certainty of observers being able to successfully retrieve earlier disclosed PI: *Erasing* is destroying information in a way that it cannot be recovered. Data is permanently deleted from storage; with erasure, all copies of data vanish forever. Erasing mechanisms offered by services require active user action. WordPress, Pinterest, and Foursquare users can permanently erase single blog posts, pins, and check-ins, respectively. Actively erasing content piece by piece may be burdensome for users who have accumulated a lot of items. Foursquare and Twitter accounts can be erased permanently. The erasure of a Facebook account takes up to 90 days and not all user information is stored within the account. Of the 13 services, passive erasing is only done by Miles and More after legally prescribed deletion periods.

Anonymising is eliminating identifiable information; identifiable information allows an individual to be singled out from a group of individuals. In contrast to erasing, what is deleted in anonymising is not the information as such, but any stored cues that help identify a single natural person. To ensure sufficient anonymity, the group of indistinguishable individuals has to be large enough (Sweeney, 2002). Two services anonymise information passively. Facebook anonymises the data collected by social plug-ins on third-party websites (liked and shared content) by removing the user's name and aggregating it with other users' data. Foursquare anonymises and aggregates the location data of its

users to share them with third parties such as restaurant managers. Online services make rare use of anonymisation and could make anonymising OETs accessible directly to users.

Modifying is changing the content of PI in a way that it gets more difficult or impossible to access the original information content. Stored PI is changed to obstruct retrieval. Wikipedia, WordPress, and Pinterest enable users to actively edit content. TripAdvisor and Twitter allow changing the display name of user profiles; parts of the stored PI are changed to make retrieval of the whole profile more difficult. Modifying also occurs passively when services distort information. The archiving of web pages in the Internet Archive's Waybackmachine or Google's cache is necessarily incomplete. Dynamic web pages and non-archived media lead to a modified snapshot of the content, not including all originally available information or functionality. While distortion is an unintentional effect in most services, online services could also integrate OETs that intentionally modify information.

Removing is excluding information from an accessible set while the information possibly endures at other locations. Retrieval from one storage location is prevented, but information remains stored at others. In many services, removing is actively accessible via the user interface. Facebook, Twitter, WordPress, Foursquare, and TripAdvisor enable users to flag self-created posts or comments as "deleted", but these are not erased from databases. Reputation.com sends manual removal requests to third-party data brokers and publicly accessible web person search engines, such as 123People. Active removal requests may undergo additional scrutiny. Wikipedia's institutionalized removal processes allow edited pages to be removed solely by community consensus; they may be restored anytime. De-indexed content of WordPress, Facebook, and Pinterest remains accessible on the web, but it can no longer be retrieved via search engines. Web pages can be directly removed from Google's web cache or the Internet Archive's Waybackmachine by blocking web crawlers in the web server's robots.txt. Removal is also passively inherent to some services. The update frequency of archived information in the Internet Archive's Waybackmachine and in Google's cache depends on how often web pages are automatically accessed. Reputation.com's "privacy services" monitor a person's online reputation; once PI reappears in third-party data brokers' databases or web person search engines, new removal requests are automatically generated. Similarly intelligent OETs that monitor a person's PI could be directly integrated into online services, also.

Hiding is reducing the accessibility of information to a smaller audience. It is similar to removing in that information is rendered irretrievable for some, but remains accessible for others. It is different from removing in that the audience to which information is restricted can be selected. Facebook's audience selector gives users active, fine-grained control of the visibility of content to other users. Twitter, Pinterest, and WordPress users also can confine the visibility of content to a pre-defined audience. Foursquare offers the option to hide check-ins from venue managers but keep them accessible to friends. Amazon's wish lists and shared recent purchases can be hidden from the public profile. Currently, the services do not provide passive hiding mechanisms, for instance by automatically suggesting to users which audience would be appropriate for which type of content.

De-referencing is disconnecting the link to or between information, or changing the address of information so that retrieval gets more difficult or impossible. Rather than deleting information content, only references to this content are cut off. All analysed de-referencing mechanisms require active user action. WordPress allows bloggers to change a blog's address by moving it to another host or transferring it to another user account, thus invalidating existing hyperlinks to the blog. On Facebook, liked information objects and tagged persons on pictures can be "unliked" and "untagged", respectively, dissociating them from an individual. Foursquare, Pinterest, TripAdvisor, and Twitter accounts can be disconnected from other social media accounts to prevent further correlation of two user profiles to each other. Users' need for dissociation is well-supported by de-referencing OETs. For instance, a person's private and business information can be disentangled.

Displacing is the out-ranking of information by other information in terms of visibility. Information is added to swamp PI with further, more desirable information. As an active process, professional writers

at Reputation.com create positive content about an individual to displace negative search results from high ranks. As a passive process, many services displace information organically. Google Search perceives poorly linked, older material as less relevant; older check-ins and pins on Foursquare and Pinterest are pushed down in feeds. Recommendations on Amazon change based on time, preferences, and other customers' behaviour until an item disappears completely. Influencing the relevance rank of positive or negative information has been successfully applied in marketing (Langville and Meyer, 2006). Search engine optimization is a laborious and costly process, though. Because of PI's privacy impact, services could artificially accelerate the displacing process for PI.

Time-stamping is the marking of information with its disclosure date, which allows the information to be arranged in its original temporal context when it is retrieved. Based on the information's temporal context, observers may discount the information or focus their attention on it (Mayer-Schönberger, 2009). Amazon, Wordpress, Wikipedia, Facebook, Foursquare, Google Search, Tripadvisor, and Twitter automatically tag content with their publication date and display it next to the content. Facebook's timeline arranges profile information along a chronicle. Using Internet Archive Waybackmachine's calendar view, users can browse web page snapshots taken at different times.

4.2.2 Do available OETs satisfy user needs?

Based on the results of study 2, we can analyse to what extent current OETs satisfy the requirements derived in study 1 (see Table 4). Addressing the requirement that all stored PI can be erased free of residues (R1), we find that only 6 out of the 13 analysed services offer active erasing mechanisms. Where active erasing OETs are provided, not all PI stored by the service can be erased. For instance, Facebook allows users to erase all PI stored in an account, but does not erase PI which is stored outside of the account, such as private messages. In contrast, 12 out of the 13 services offer an OET to remove PI, but the web may continue storing that PI at other storage locations.

Requirements		**Fulfilment of requirements by OETs deployed in online services**
R1	Capability to erase all stored PI free of residues	6 out of 13 services have erasing OETs implemented Not all stored PI is covered by erasing capability of services
R2	Comprehensive long-term control over storage and access to PI	All services provide some active control over PI to users OETs are difficult to access, reducing actual user control
R3	Intelligent, automated capability to forget outdated PI	2 out of 13 services automatically anonymise PI after a time 1 out of 13 services enables automated long-term PI management

Table 4. Fulfilment of the requirements found in study 1 by the analysed online services.

With regard to the requirement of comprehensive long-term control over the storage of and access to PI (R2), we find that all analysed online services provide at least one active long-term control option. However, some OETs are difficult to access, reducing the actual control users have. Frequently, the option to de-reference a user account from social media is hidden in the account's menu structures. Removing web pages from Google Search, 123People, or the Internet Archive's Waybackmachine requires easier user interface access, as currently web crawlers need to be blocked in the web server's robots.txt or manual removal requests need to be filed. Hiding mechanisms, in contrast, can be configured over the user interface. They give access to one audience group by simultaneously blocking other observers. Retrieval-stage OETs such as hiding are enhancing user control by allowing fine-grained control of forgetting, although only 6 out of the 13 services provide such OETs.

Addressing users' need for dissociation from obsolete information, services should provide intelligent, automated capabilities to forget outdated PI (R3). Automatically anonymising users' PI after a time is a promising mechanism for satisfying this need. However, only 2 of the 13 analysed services are deleting PI related to single users while retaining aggregated information. Users also lack intelligent support for dissociation from obsolete information. Only one service (Reputation.com's monitoring mechanism) is autonomously managing users' PI over time. The service notifies users when undesired

PI appears on the web, and reacts automatically. In most online services, though, passive forgetting happens by coincidence rather than by intentional support.

5 Discussion and Conclusions

Our two studies contribute by providing a model of users' needs for oblivion in online services and analysing whether different types of OETs deployed in today's online services satisfy those needs. The results highlight that deployed OETs do not provide intelligent capabilities to forget outdated PI, and do not give users sufficient long-term control over the storage of their PI. Online services need to implement OETs that operate in both the storage and retrieval stages of the memory process. But, closing the gap between users' needs and available OETs in the online environment will be challenging due to three current developments on the web.

First, the increasing amount of PI users disclose on the web (WEF, 2012) conflicts with users' desire for control. Users are already reaching the limits of manual control over the PI they have disclosed over time. Hardly any online services provide users with intelligent support for managing PI in the long-term, or enable organic forgetting of PI. Intelligent OETs should provide three capabilities: tracing disclosed PI for the user, filtering critical information, and suggesting appropriate counteraction. As a prerequisite to control, OETs need to keep track of all PI users have disclosed to different online services. Then, OETs need to evaluate which PI could harm the users' privacy or no longer fit the users' self-identity. For such information, OETs need to take appropriate counteraction at the online services that store the PI. Based on data mining and machine learning, future work should develop such autonomous, intelligent user agents (e.g., Le Métayer and Monteleone, 2009).

Second, accessible PI on the web dates back to a progressively longer history, thereby not only transcending spatial but also temporal boundaries (Marx, 2001). Temporal transcendence of PI conflicts with users' need for dissociation from obsolete PI. PI is still available long after disclosure and will be exclusively interpreted in the temporal context of its retrieval. As a result, previous work assumed that access to PI should be denied after a fixed period of time (e.g., Geambasu et al., 2009; Karla, 2010; Mayer-Schönberger, 2009). Available retrieval-stage OETs, such as time-stamping, suggest alternatives to deletion. Displaying the time context of PI could facilitate the interpretation of PI in its original temporal context. Engineers should design user interface components beyond textual time-stamps to saliently display PI's original temporal context. PI's temporal context, though, should not be disclosed in case this information could harm the user's privacy.

Third, distributed information storage and the exchange of PI between services complicate the implementation of OETs. A cooperative design is required to enforce digital oblivion across boundaries of services. Erasing backups and informing third-parties that PI has been passed on require more effort than marking information as removed in a service's own database. The proposed EU right to erasure, though, will oblige service providers to notify third parties that have a copy of PI when users erase information (Kalabis and Selzer, 2012; Xanthoulis, 2012). Thus, standardized interfaces should be developed to enable the exchange of oblivion-related communication between services.

Future research should continue to ask how to design the web for oblivion. How can the web intelligibly understand what humans want to preserve and when information should be forgotten? And to what extent should users play an active role in the process of oblivion? Digging deeper into these issues is promising. Designing the web for oblivion can assist users in preserving valuable personal memories, while not putting them in jeopardy of being frozen in the story their PI tells to others.

Acknowledgement

We would like to thank Kathryn Sterling-Kaasa for the editing of this paper.

References

Alexa. (2013). Global Top Sites. The Top 500 Sites on the Web. Available: http://www.alexa.com/topsites (Accessed Apr 18, 2013).

Antonopoulos, A.M. (2010). Privacy? Run a Background Check on Yourself. Network World, Oct 19, 2013.

Atkinson, R.C. and Shiffrin, R.M. (1968). Human Memory: a Proposed System and its Control Processes. In K.W. Spence & J.T. Spence (Eds.), The Psychology of Learning and Motivation: Advances in Research and Theory (Vol. 2, pp. 89-195). Academic Press, New York.

Ayalon, O. and Toch, E. (2013). Retrospective Privacy: Managing Longitudinal Privacy in Online Social Networks. In Proceedings of the 9th Symposium on Usable Privacy and Security, ACM, Jul 24-26, Newcastle, UK, 1-13.

Backes:SRT. (2013). X-Pire! Available: http://www.backes-srt.de/produkte/x-pire/ (Accessed Oct 10, 2013).

Bannon, L.J. (2006). Forgetting as a Feature, Not a Bug: the Duality of Memory and Implications for Ubiquitous Computing. CoDesign, 2 (1), 3-15.

Bartlett, F.C. (1932). Remembering: A Study in Experimental and Social Psychology. Cambridge University Press.

Barua, D., Kay, J., Kummerfeld, B. and Paris, C. (2011). Theoretical Foundations for User-controlled Forgetting in Scrutable Long Term User Models. In Proceedings of the 23rd Australian Computer-Human Interaction Conference (OzCHI), ACM, Nov 25-29, Canberra, Australia, 40-49.

Bauer, L., Cranor, L.F., Komanduri, S., Mazurek, M.L., Reiter, M.K., Sleeper, M. and Ur, B. (2013). The Post Anachronism: the Temporal Dimension of Facebook Privacy. In Proceedings of the 12th ACM Workshop on Privacy in the Electronic Society (WPES), ACM, Nov 4, Berlin, Germany, 1-12.

Beel, J., Gipp, B. and Wilde, E. (2010). Academic Search Engine Optimization. Journal of Scholarly Publishing, 41 (2), 176-190.

Bélanger, F. and Crossler, R.E. (2011). Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. MIS Quarterly, 35 (4), 1017-1042.

Blanchette, J.-F. and Johnson, D.G. (2002). Data Retention and the Panoptic Society: The Social Benefits of Forgetfulness. The Information Society, 18 (1), 33-45.

Cheng, J. (2012). Over 3 Years Later, 'deleted' Facebook Photos are Still Online. Wired, Feb 06.

COM. (2012). Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation). Jan 25, European Commission.

Conger, S., Pratt, J.H. and Loch, K.D. (2013). Personal Information Privacy and Emerging Technologies. Information Systems Journal, 23 (5), 401-417.

Conley, C. (2010). The Right to Delete. Papers from the AAAI Spring Symposium: Intelligent Information Privacy Management 2010, Mar 22-24, Stanford, CA, 53-58.

Couts, A. (2012). Meet the Online Snoops Selling Your Dirty Laundry and How you Can Stop Them. DigitalTrends, Mar 27.

Cumming, K. and Findlay, C. (2010). Digital Recordkeeping: are We at a Tipping Point? Records Management Journal, 20 (3), 265-278.

Dinev, T. and Hart, P. (2006). An Extended Privacy Calculus Model for E-Commerce Transactions. Information Systems Research, 17 (1), 61-80.

Draaisma, D. (2013). Das Buch des Vergessens (2 ed.). Galiani, Berlin.

Eickelpasch, R. and Rademacher, C. (2004). Identität. Transcript Verlag, Bielefeld.

EP. (2013). Draft Report on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individual with Regard to the Processing of Personal Data and on

the Free Movement of Such Data (General Data Protection Regulation). Jan 16, European Parliament.

Federrath, H., Fuchs, K.-P., Herrmann, D., Maier, D., Scheuer, F. and Wagner, K. (2011). Grenzen des 'digitalen Radiergummis'. Datenschutz und Datensicherheit - DuD, 35 (6), 403-407.

FIP. (2012). Fair Information Practice Principles. Available: http://www.ftc.gov/reports/privacy3/fairinfo.shtm (Accessed Sep 26, 2013).

Geambasu, R., Kohno, T., Levy, A. and Levy, H.M. (2009). Vanish: Increasing Data Privacy with Self-destructing Data. In Proceedings of the 18th USENIX Security Symposium, Aug 10–14, Montreal, Canada.

Greenleaf, G. (2012). Global Data Privacy in a Networked World. In I. Brown (Ed.), Research Handbook on Governance of the Internet (pp. 41). Edward Elgar, Cheltenham.

Hansen, M., Berlich, P., Camenisch, J., Clauß, S., Pfitzmann, A. and Waidner, M. (2004). Privacy-enhancing Identity Management. Information Security Technical Report, 9 (1), 35-44.

Hughes, J. and Jones, S. (2003). Reflections on the Use of Grounded Theory in Interpretive Information Systems Research. In Proceedings of the European Conference on Information Systems (ECIS), 19-21 June, Naples, Italy, 833-845.

Kaiser, A.-B. (2012). Rechtlich gefordertes Nichtwissen im virtuellen Raum. Der Schutz der Privatsphäre im Web 2.0. In H. Hill & U. Schliesky (Eds.), Die Vermessung des virtuellen Raums. E-Volution des Rechts- und Verwaltungssystems 3 (pp. 55-71). Nomos, Baden-Baden.

Kalabis, L. and Selzer, A. (2012). Das Recht auf Vergessenwerden nach der geplanten EU-Verordnung. Datenschutz und Datensicherheit - DuD, 36 (9), 670-675.

Kanungo, S. and Jain, V. (2009). Using Interpretive Structural Modeling to Uncover Shared Mental Models in IS Research. In Proceedings of the European Conference on Information Systems (ECIS), Jun 8-10, Verona, Italy, Paper 192.

Karla, J. (2010). Can Web 2.0 Ever Forget? Business & Information Systems Engineering, 2 (2), 105-107.

Korenhof, P. (2013). Forgetting Bits and Pieces. Presented at the 8th International IFIP Summerschool on Privacy and Identity Management, June 17-21, Berg en Dal, NL.

Krasnova, H., Eling, N., Schneider, O., Wenninger, H., Widjaja, T. and Buxmann, P. (2013). Does This App Ask For Too Much Data? The Role Of Privacy Perceptions In User Behavior Towards Facebook Applications And Permission Dialogs. In Proceedings of the European Conference on Information Systems (ECIS), June 5-8, Utrecht, NL.

Krumm, J. (2011). Ubiquitous Advertising: The Killer Application for the 21st Century. IEEE Pervasive Computing, 10 (1), 66-73.

Langer, E. (1983). The Psychology of Control. Sage, Beverly Hills, CA.

Langville, A.N. and Meyer, C.D. (2006). Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press, Princeton, NJ.

Lawrence, S., Pennock, D.M., Flake, G.W., Krovetz, R., Coetzee, F.M., Glover, E., Nielsen, F.A., Kruger, A. and Giles, C.L. (2001). Persistence of Web References in Scientific Research. Computer, 34 (2), 26-31.

Le Métayer, D. and Monteleone, S. (2009). Automated Consent through Privacy Agents: Legal Requirements and Technical Architecture. Computer Law & Security Review, 25 (2), 136-144.

Marx, G.T. (2001). Murky Conceptual Waters: The Public and the Private. Ethics and Information Technology, 3 (3), 157.

Mayer-Schönberger, V. (2009). Delete: The Virtue of Forgetting in the Digital Age. Princeton University Press, Princeton, NJ.

O'Hara, K. (2012). Can Semantic Web Technology Help Implement a Right to Be Forgotten? Computers and Law, 22 (6).

Patsakis, C. (2012). Encrypt to Forget. In Proceedings of the 7th Spanish Meeting on Cryptology and Information Security, Sep 4-7, Donostia-San Sebastián, Spain, 1-5.

Perlman, R. (2005). The Ephemerizer: Making Data Disappear. Feb 2005, Sun Microsystems.

Raaijmakers, J.G. and Shiffrin, R.M. (1981). Search of Associative Memory. Psychological Review, 88 (2), 93-134.

Rizk, R., Gürses, S. and Günther, O. (2010). SNS and Third Party Application Privacy Policies and their Construction of Privacy Concerns. In Proceedings of the European Conference on Information Systems (ECIS), Jun 6-9, Pretoria, SA, Paper 143.

Sleeper, M., Cranshaw, J., Kelley, P.G., Ur, B., Acquisti, A., Cranor, L.F. and Sadeh, N. (2013). "I read my Twitter the next morning and was astonished": a Conversational Perspective on Twitter Regrets. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Apr 27 - May 2, Paris, France, 3277-3286.

Solove, D. (2006). A Taxonomy of Privacy. University of Pennsylvania Law Review, 154 (3), 477-560.

Spiekermann, S. and Korunovska, J. (2014). The Importance of Interface Complexity and Entropy for Online Information Sharing. Behaviour & Information Technology, forthcoming, Available: http://www.tandfonline.com/doi/abs/10.1080/0144929X.2013.845910#.UyhdKk1OVaQ (Published online Jan 31, 2014).

Sweeney, L. (2002). k-Anonymity: A Model for Protecting Privacy. International Journal of Uncertainty, Fuzziness & Knowledge-Based Systems, 10 (5), 557-570.

Szekely, I. (2012). The Right to Forget, the Right to be Forgotten. In S. Gutwirth, R. Leenes, P. De Hert & Y. Poullet (Eds.), European Data Protection: In Good Health? (pp. 347-363). Springer, New York.

Tene, O. and Polenetsky, J. (2012). To Track or Do Not Track: Advancing Transparency and Individual Control in Online Behavioral Advertising. Minnesota Journal of Law, Science and Technology, 13, 281-357.

WEF. (2012). Rethinking Personal Data: Strengthening Trust. May 2012, World Economic Forum.

Xanthoulis, N. (2012). Conceptualising a Right to Oblivion in the Digital World: A Human Rights-Based Approach. Available at SSRN 2064503, May 22, 2012.

Zhao, X., Salehi, N., Naranjit, S., Alwaalan, S., Voida, S. and Cosley, D. (2013). The Many Faces of Facebook: Experiencing Social Media as Performance, Exhibition, and Personal Archive. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13), ACM, Apr 27 - May 2, Paris, France, 1-10.