Usable Assured Deletion in the Cloud

Lancaster
University

Kopo Marvin Ramokapane

B.Eng. (Hons)
M.Eng. (Hons)

Supervisors: Prof. Awais Rashid
Dr. Jose M. Such

School of Computing and Communications
Lancaster University

This dissertation is submitted for the degree of
Doctor of Philosophy

Graduate College July 2019

To the memory of Keitumetse "Kgasi" Goitsemang (my grandmother) and Atlang Brian Goitsemang (my cousin brother). Your love and support (grandmother), and enthusiasm and imaginative mind (brother) spurred me on to complete this work. I miss you.

To my Mom (Mapena Gaorere Ramokapane) and Dad (Taelo Galeitsewe Ramokapane) for staying together when I battled with cancer, for allowing me to dream, and for making me believe that my dreams were a reality. Ke a lo rata (I love you). I am proud to be your son. Thank you for being farmers.

To my sisters (Tshego Ramatshaba, Lesego D. Ramokapane, Prudence M. Ramokapane and, Gontle G. K. Ramokapane) for your love and support.

To all my nephews, you made my vacations worthwhile.

To my grandmother (Keithokile “MmaDuks” Gaumakwe), for your love and sacrifice. Ke a go rata.

To my extended family, I love you all and thank you.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. The entire research was carried out under the supervision of Awais Rashid and Jose M. Such, formerly at Security Lancaster, School of Computing and Communications at Lancaster University.

Kopo Marvin Ramokapane
July 2019

Acknowledgements

Writing this thesis has been at times quite a personal endeavour; learning a unique trade and sharing what I have learnt. To me, this is fundamentally a social process. The thesis you are holding right now is the evidence that I have gone through the process of learning, questioning and thinking about how people delete from the cloud. Consequently, I need to recognise and acknowledge the support and encouragement I have received over the past years.

PhD Advisors

First, I would like to start expressing my sincere thanks to my advisors, Awais Rashid and Jose Such. Awais and Jose have been mentors and valued colleagues. I am fortunate to have worked with both of them as my advisors. They have mentored me throughout my PhD, taught me the art of research, and made my PhD life bearable. They set me very high standards and challenged me always to outdo my best. I will always be grateful for their unconditional support during my PhD; believed in me even when I did not think I could make it. I will miss our discussions, especially the "disagreements" we had concerning my work. Lastly, I would love to say I have made peace with them for deciding to take new roles away from Lancaster while I was doing my PhD.

I would also want to thank my internal panel, Vasileios Giotsas and Charalampos Rotsos for their time and valuable feedback during my PhD. They were always prompt in giving me feedback. I want to also express my thanks to Alistar Baron for helping me deal with admin related matters after Awais and Jose left Lancaster University.

Mentors

Second, I am fortunate to have God-given parents and mentors over the years, I would like to thank the Manzanga family for adopting me since 2004. Their support and love have made me feel special. Without their prayers and support, I would have been lost through the huddles of life. I would also want to thank Dr Enoch and his family for their support and helping me accept England as my second home. Enoch has been a brother to me since I started my undergrad studies in Essex. I am indebted to Ingrid and the family for their love and support including some fantastic desserts during my visits. I am grateful for Dr Ife for helping me with the last lap of my Ph.D. He was there just at the right time to help me see this thesis through. I would also want to thank Kesed, King's and Bristol Woodies for playing a significant role in my life during my time as a student.

Friends - Lancaster

Third, profound gratitude goes to Tim and Kirsty Forber, alias 'the Forbers'. Without these two, I would not have finished my PhD or even worse been homeless. They made me feel wanted and important, most importantly they never got tired of listening to me talk for hours and hours about everything in life. The Wiercinskis, for their polish experience and 'Marcin', and the Daleys for welcoming me to Lancaster.

Friends - Botswana, Colchester, Cardiff, Leeds, Trinidad and Tobago

Fourth, I would like to thank those that have kept me sane over the past few years, including many of my friends and relatives, Ernest Sebitiela, Anysha Thomas, Katlego Ramaphane, Oteng Tabona, Antoinette Moleele, Joseph Loh Han Wei, Seleb Seepi, Nonofo Molapisi, Alec Agisanyang, Angela Jansen, Mmita Sekgobane, Doringdah Foli, Kamogelo Gobuiwang, Kgomotso Matshidiso. I am also thankful for my little brothers and sisters from Mahalapye. You have become family, you were there when I first started this journey, and have seen me through it all. I would also like to thank Kaone McAdam with her family for helping me out during my time in Lancaster.

University of Essex, Univeristy of Botswana

Fifth, I am grateful for Selwana (Ranna) Nkwe, Sumit Mittal, Onty Bagwasi, Kgotla Ditlhabi and Thapelo Remotsaletswe Malusi for setting me on the path to grad school. They challenged, helped and even taught me during my undergrad days.

Lancaster University

I would also like to thank my colleagues and collaborators, Lancaster Security group especially the former (Ben, Gaurav, Asad, Peng, Ethem, Rob) and the present residents of office B55 and B59. Their help and discussions helped me to narrow and focus my work. I also like to extend my thanks to Hamed, Samuel and Farid. I know that I have disappointed a lot of my former colleagues by finishing my PhD without drinking coffee.

Social Media Followers

I also want to thank my Facebook friends and my Twitter followers. They represent the people this thesis is about; cloud users. Interreacting with them made me think a lot about their technological needs. Some of my ideas were inspired by their posts.

Family

I want to thank my family who has sacrificed and endured my inconveniences and hardships during my prolonged absence. My sisters have had to put up with seeing me once a year and not responding to their calls and messages. Their prayers, support, believe and encouragement made it possible for me to be where I am today. To my nephews, I love you and thank you for all the time you spent with me during my vacations. I am also grateful for all my relatives for understanding my absence from all family events. Nametso Masole and Bobedi Bakae kept me going and made my vacations worthwhile.

Lastly, God.

Abstract

The prevalence of cloud and storage-as-a-service has led to users storing and sharing data through such services. However, little is understood about one key element of data management in this new landscape, i.e., data deletion and more critically assured deletion. With regards to deletion, existing research has not explored the deletion needs of users, their preferences and the challenges they face. Nor is there any understanding of the challenges faced by cloud providers should they want to offer assured deletion.

Users' deletion needs and their preferences are diverse and vary depending on the context. However, satisfying these needs may be limited to the properties of the infrastructure - what the infrastructure permits and does not. For instance, the cloud infrastructure has various features that may pose different challenges to meeting the needs of users and providing assured deletion. These features include virtualization, multi-tenancy, high availability and On-demand elasticity.

The work presented in this thesis is the first to investigate these issues. Thus, it finds that users' motivation to delete are: privacy-, policy-, expertise- and storage-driven. They fail to delete because of the poorly designed interfaces, the way they perceive cloud deletion and lack of information about cloud deletion. Users want to have a choice in how their data is deleted, they want to be able to specify the type of deletion. Their deletion preferences are complex and may always change depending on the context of deletion, i.e., individually or socially. Regarding information about deletion, they want important information that may help them to delete or recover from failures to be easily accessible through the interface. They do not want essential information only to be restricted to privacy policies. Using these findings, this thesis provides a conceptual framework for the design of usable assured deletion in the cloud and then formulates user requirements for usable assured deletion. With regards to providers, by analysing the cloud infrastructure, this work provides a systematization of the challenges that providers face while attempting to assure deletion. It also identifies the cloud provider requirements for usable assured deletion. By considering both sets of requirements, i.e., user and provider requirements, this work provides user requirements and principles for usable assured deletion. Overall, the findings of this work formulate a solid grounding for the design and the development of cloud systems that assure deletion in a usable way. More importantly, it helps in the empowerment of users with regards to assured deletion.

Setswana

Go anama ga maranyane a dipolokelo go okile tiriso ya one. Mme le fa go ntse jalo, tiriso ya one ga e ise e tlhalogangwe thata segolobogolo go rurufatsa go sutlhwa ga dilo. Dipatlisiso tsa maranyane ga ise di itebaganye le go tlhaloganya gore batho fa ba shutlha, ba tlhoka eng, ba kopana le mathata kana dikgwetlho dife, le dikeletso tsa bone. Mo godimo ga moo, go santse go sa itsewe gore ba ba rekisang kgotsa baanamisa maranyane a, bone ba kopana le dikgwetlho dife thata thata fa ba na le keletso ya go rurufatsa go sutlhwa ga dikitso.

Batho ba na le dikeletso tse difarologanyeng fa ba sutlha, dikeletso tse di tswa fela gore motho o sutlhela eng kana o sutlha eng. Go itepatepanya kana go diragatsa dikeletso tse, go laolwa ke maranyane a a dirisiwang, ke gore, maranyane a letla eng kana ga a kgone eng. Sekai, mafaratlhatlha a maranyane a na le dintlha tsa botlhokwa tse dintsi mme go sutlha dikitso mo ditlheng tse go bake dikgwetlho tse di farologanyeng ka di tlhoka botegeniki jo bo farologanyeng. Dintlha tse di akaretsa go letlelela batho ba le bantsi go dirisa maranyane a ka nako ele ngwe, le go tlhoka go nna teng ga ditiriso tse ka nako tsotlhe le tse dingwe fela jalo.

Tiro e e mo lokwalong lo, ke yone ya ntlha mo lefatsheng go itebaganya le go tlhaloganya dilo tse ka kakaretso. Ka jalo, dipatlisiso tsa lokwalo lo di fitlhetse kana di lemogile gore batho ba sutlha ka mabaka a a latelang: go boloka diphiri, go diragatsa mokwalo wa ditumalano, go gwetlhiwa ke letlhoko la mabolokelo, fa ba bangwe bone ba sutlha ka gore ba a kgona kana ba itse gore go sutlhwa jang. Lefa go ntse jalo, ga se mongwe le mongwe yo o kgonang go sutlha fa a batla kana a na le keletso ya go dira jalo. Bangwe ba palelwe ke go sutlha ka mabaka a a lateng: go tlhoka kitso e e tsepameng, go akanya ga bone mabapi le gore maranyane a a bereka jang, le go tlhoka didirisiwa tse di motlhofo go dirisiwa fa ba sutlha. Dipatlisiso tsa tiro e di lemogile gore batho ba batla dilo tse di farogoganyeng mabapi le go sutlha, mme ka kakaretso batho ba batla go itlhopela gore dikitso tsa bone di sutlhwe jang ka nako efe. Le fa go ntse jalo, dikeletso tsa bone ga di a kwalwa mo letlapeng, ke gore, dikgona go fetoga nako le nako, segolobogolo fa ba sutlha ka bonosi kana ba dirisa maranyane a le ba bangwe, ke gore, ba le setlhopha. Fa go buiwa ka letlhoko la dikitso kana dikaedi, batho ba batla kitso ee tla ba thusang go sutlha nako le nako kana e e tla ba ntshang mo mathateng fa go sutlha go pala. Mo godimo ga moo, ba batla gore dikitso tse di nne gaufi, kana fa ba ka dibonang motlhofo teng. Ga ba batle gore ditaelo kana dikitso tse di kwalwe fela mo mokwalong wa kgotla-tsamaiso ya tlhokomelo ya maranyane.

Tiro e, e dirisa ditshwetso tsa patlisiso go dira lenaneo kgotsa mokgwa o o tlhomameng wa tsamaiso ya go dira maranyane a a tla thusang batho go sutlha motlhofo mo mafaratlhatlha a mabolokelo, le go papamatsa kana go kgaogana dintlha tse di tshwanelwang ke go selwa morago fa go diriwa maranyane a. Mabapi le dikompone tse di rekisang kana tse anamisang maranyane a, dipatlisiso tsa tiro e di tlhalosa ka botlalo dikgwetlho tsa go sutlha, ke gore mathata a dikopone tse di kopanang nao fa ba batla go sutlha kana go rurufatsa gore ba sutlhile dikitso ka karetso. Mo godimo ga moo, e supa dintlha tse dikopone tse di tshwanentseng go di sela morago go tlhomamisa gore fa ba sutlhile sentle go seka ga nna le ditlamorago tse di maswe. Go leba dintlha tse tsotlhe, tiro e supa *dintlha tsa motia tse ditshwanetseng go elwa tlhoko gore batho ba kgone go sutlha. Go garela, tiro e ke motheo kana tshimolodiso ya go dira maranyane a a tsenyang go sutlha mo teng, kana go kaa botlhokwa jwa go dira gore go sutlhwa ga dikitso go diriwe motlhofo. Se se botlhokwa thata ka tiro e, ke gore e fa badirisi ba maranyane a dithata tsa go nnetefatsa gore dilo tsa bone di sutlhwa sentle ka boammaruri jaaka go dumalanwe.*

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Overview

Cloud computing has become the backbone of hosting and delivering online services over the Internet. As a result, computing resources have become cheaper, more powerful and ubiquitously available to users. Since cloud computing enables a whole collection of computing resources such as applications, storage space and processing power at low cost, the amount of data stored and processed in the cloud has significantly increased. It is expected to quadruple by 2020, attracting 2.3 billion users in the process (Cisco 2018 (accessed August 15, 2018)).

One fundamental aspect of cloud computing is that data is outsourced to third-party owned computing resources – relieving users of the burden of hardware and software maintenance costs. Users do not own the technology or the services; instead, they rent time, processing power or storage space from the cloud provider. The emergence of cloud computing reflects a paradigm shift from personal computing. One of the defining characteristics of the personal computing model is that users maintain physical control over their files and data. They are also responsible for maintaining (e.g., upgrading their hardware and software) their computers. However, with cloud computing, users have no physical access to their machines and data. Instead, they use Application Programming Interfaces (APIs) for access. The cloud enables users to access their data from anywhere in the world, and they can enjoy built-in version control. Since most processing is done remotely, users can also access their data through less powerful devices like smartphones. Furthermore, most cloud services and consumer-oriented services are generally "free", in the sense that users are not required to pay in order to have access to the services.

While many users choose specific cloud services with the mindset that they understand the benefits and the risks they are taking (e.g., someone using cloud services to back up their valuable photos, but knowing that the provider may get hacked), in many situations users do not know how the provider will handle their data (Marinescu 2017). Providing such assurances is still a concern for the cloud. There are two significant challenges, (1) assuring deletion and (2) doing it in a usable way. Assured deletion aims to provide users with guarantees over deletion data upon request. Current cloud systems are limited and offer no assurances or control over deletion of users' data. Nonetheless, incomplete deletion can lead to unintentional disclosures (Rahumed et al. 2011). The costs of data disclosures are high and may include financial losses and loss of reputation for both cloud providers and consumers. Moreover, not being able to delete has also been associated with negative psychological well-being (Sweeten et al. 2018).

Table 1.1 Summary: Assured cloud deletion properties.

Property	**Description**
Inaccessibility	Not allow deleted data to be accessible after deletion.
Timeliness	Allow data to be deleted from the system immediately or within a reasonable time after the owner or user requests it to be deleted.
Adequate feedback	Provide adequate and relevant feedback to the users during and after data deletion.
Completeness	Completely delete all data stored in the system including all copies and its metadata when the user requests for deletion.
Fine-grained deletion	Allow users to specify which data should be deleted, and only that data should be deleted without affecting other data stored in the system.
Choice	Give the owner of the data the option to specify how their data should be deleted.
Proof of deletion	Be able to provide proof that deleted data is no longer stored in the system and it is completely deleted.

Assured deletion has always been associated with the inaccessibility of data (Diesburg and Wang 2010, Reardon et al. 2013). However, as this thesis will later show in Chapter 7.1, there are other properties of assured deletion. These include inaccessibility, timeliness, adequate feedback, completeness, fine-grained deletion, choice of deletion, proof of deletion. This work assumes that a system that offers assured deletion should meet some of these properties. Table 1.1 summarises these properties.

Satisfying these assumptions is a tall order. In reality, cloud providers use contractual agreements and privacy policies to fulfil these requirements. Contracts and SLAs leave users with no technical evidence but to rely on trust – the provider will delete data as agreed (Campbell et al. 2012). Moreover, since users do not have direct access to the infrastructure, they cannot technically verify deletion for themselves. It is also not clear whether current cloud deletion mechanisms meet the deletion needs of users. Thus, the work presented in this thesis aims to investigate the following:

- What are the assured deletion challenges for cloud providers,

- What do users struggle with,
- What do users want, that is, deletion preference and how to be informed about deletion,
- How do user and cloud provider challenges relate, what are the resulting requirements, and what principles should be followed for usable assured deletion in the cloud.

1.2 Motivation

Data disposal is an essential aspect of data management and protecting sensitive data. Without deletion, data may still be vulnerable to abuse either through theft or litigation. It is, therefore, not only crucial to data owners but to service providers as well.

For cloud service providers, such guarantees are needed to comply with the data regulations of various countries and regions. For instance, article 17 of the European General Data Protection Regulations (GDPR)[1] obligates data controllers "service providers" to obscure or delete users' personal data upon request by the user (Rosen 2011, GDPR 2018 (accessed October 15, 2018)). Moreover, assured deletion can help providers meet their users' requirements and expectations. For example, satisfying the needs (i.e., completely deleting data) of an enterprise that relies on deletion for business. Furthermore, in some cases, deletion assurances can also become a differentiator in the market for a cloud provider. For instance, a cloud service provider can offer assured deletion as a service to its customers. Again, not being able to offer assured deletion may lead to undesired consequences for the provider (e.g., loss of reputation due to disclosures) after a data breach.

For cloud tenants (i.e., users), it is essential to receive assurances that data will be destroyed or deleted as agreed when a deletion request is made. Users delete to get rid of their old files and preserve their privacy (Khan et al. 2018). Also, being able to delete data from the cloud may give cloud customers a sense of control over what they store in the cloud. After the 2018 Cambridge Analytica and Facebook[2] scandal (Valdez 2018 (accessed October 22, 2018)), an online campaign was launched on social media (i.e., *#DeleteFacebook*) to encourage users to delete their Facebook accounts after data misuse reports. However, it is difficult for users to verify that their data was deleted. This scandal also highlights the need to empower users with regards to data retention and, critically, data deletion (Reilly 2017 (accessed October 22, 2018, Venkatadri et al. 2018)). Users need to know how long their data is kept by the provider and that it will be deleted when they desire so. Besides control,

[1] http://www.privacy-regulation.eu/en/article-17-right-to-erasure-%27right-to-be-forgotten%27-GDPR.htm

[2] http://www.facebook.com

users delete to deal with regrets or forget painful experiences (Wang et al. 2011b, Johnson et al. 2012, Almuhimedi et al. 2013).

In the past, there have been incidents that emphasise the need for assured deletion in the cloud. These incidents are as follows.

The need to understand challenges faced by cloud service providers.

Digital Ocean reported that a bug in their application programming interface (API) allowed new users to read data belonging to former users from their virtual servers (Ocean 2014 (accessed August 14, 2018)). This failure resulted in users' data being exposed to other users unknowingly. Digital Ocean claimed that this only affected 3% of their virtual servers but online sources suggest this could be more. In 2016, two years after the bug was reported to have been fixed, several users suggested that the problem still existed. In response, a Digital Ocean support team member explained that the fix was still being propagated throughout all regions and all servers. Therefore, it is imperative to understand assured deletion from the provider's side — investigate how providers can assure deletion or the challenges they face in an attempt to offer such assurances in the cloud.

The need to understand the deletion needs and challenges of users.

On the 29th April 2014, a user named Jan Čurn lost 8343 photos and videos backed up on Dropbox (Curn 2014 (accessed August 15, 2018)). Jan accidentally discovered this disaster two months later when he was looking for an old presentation that he wanted to use for defending his PhD thesis. He could not find his presentation, and the folder where it was saved was empty. Upon realising this, Jan contacted the Dropbox support services. He was told that he deleted his files some months before and since their retention time had already elapsed (60 days) it was difficult to recover them. Upon further investigation, Jan learnt that he had mistakenly deleted his files when his Dropbox client crashed while he was trying to desync some folders to free some space in his local machine. When the client restarted, it deemed desynced folders deleted and removed them from the server. Although he never recovered everything, Dropbox managed to recover 1463 of his files. This example shows that the current cloud deletion mechanisms leave users unable to detect when the cloud system has failed or encountered a problem while deleting. Their everyday misinformed deletion decisions leave them with unknown consequences. However, research has not explored this field, user deletion practices, preferences, and needs have not yet been considered.

These high-profile incidents highlight the need and the importance of providing assured deletion in the cloud. Digital Ocean's experience highlights the challenges that exist for providers willing to offer assured deletion, or comply with regulation (e.g., 2014 EU ruling over the right to be forgotten). Dropbox's deleted files incident suggests the need to prove and communicate deletion, and clear policies on data retention. Dr Jan Čurn's example shows the challenges users face and the disconnect that exists between data deletion that is offered and users' expectations.

1.3 Limitations of the State-of-the-art

The previous section expresses the importance of assured deletion in the cloud. Unfortunately, this area of the cloud has not been given paramount attention. Prior research efforts have been made in various areas to provide cloud users with assurances such as proof of data availability (Juels and Kaliski Jr 2007, Ateniese et al. 2007, Benson et al. 2011), data integrity (Bowers et al. 2009a, Singh et al. 2012), data location (Benson et al. 2011, Watson et al. 2012), and ownership (Halevi et al. 2011). Another primary concern for data in the cloud has been preventing an attacker from accessing sensitive information (Wang et al. 2011a, Priebe et al. 2014).

With regards to assured deletion, existing efforts (Tang et al. 2010, Rahumed et al. 2011, Mo et al. 2014b, Xue et al. 2018, Dong et al. 2018) have been restricted to enterprises and proposing different solutions that leverage the use of encryption schemes to assure deletion. Before outsourcing data to the cloud, it is encrypted and the encryption keys remain with the data owner. At the end of data use, the owner deletes the data together with its encrypting keys. Assured deletion is then artfully achieved when deleted data is no longer accessible, and the encryption keys have been securely destroyed. These solutions do not lack shortcomings:

- they assume a dishonest cloud service provider, a provider that has no interest in offering assured deletion. They have focused on assuring deletion when the provider is malicious or has no intentions of providing assured deletion. Prior work has failed to investigate assured deletion from an honest provider's side – a provider with interest in providing assured deletion. As a result, there is no evidence on how such providers can assure deletion or the challenges they face in assuring deletion. Moreover, the existing literature lacks requirements that these providers should satisfy in order to assure deletion.

- they focus on the enterprises that seek assured deletion but have neglected a user, the deletion needs and challenges of cloud users. These studies do not examine users' current deletion practices – what their needs are, the challenges they face and their deletion requirements. Current literature lacks user requirements for assured deletion despite these being useful for fulfilling the deletion needs for users.

- they focus on providing one assured deletion property, inaccessibility of data. They neglect other properties of assured deletion which may be critical for both users and service providers. In fact, proposed solutions satisfy this property by using encryption despite evidence of usability (Kamara and Lauter 2010, Curtmola et al. 2011, Erway et al. 2015) and performance issues (Cao et al. 2014). Moreover, existing research does not offer requirements for assured deletion for a cloud provider.

Currently, cloud service providers use Service Level Agreements (SLAs) to assure deletion, but they are not well suited for addressing this challenge. SLAs do not provide any technical proof – users rely on trust that the provider will honour the agreement (Campbell et al. 2012). Contracts and SLAs play a vital role in how the provider will offer cloud services. The service provider details what services they will offer and how they will offer such services including the penalty if the contract is not respected. However, SLAs on their own are limited; as previously mentioned they are highly dependent on trust (Campbell et al. 2012, Huang and Nicol 2013, Campbell et al. 2018). Moreover, SLAs cannot give any (quantifiable) assurance. There is also an issue of usability (Burkon 2013). Nonetheless, in other areas of the cloud, significant efforts have been made to provide technical assurances. For instance, to verify that the cloud provider has not violated the service agreement on data location, Albeshri et al. (2012) and Watson et al. (2012) proposed different protocols to provide a geographic assurance for the data owners – that their data remains in the same physical location specified in the SLA. When data is moved or copied to a different location, the system will alert the data owner. However, with regards to deletion, existing literature lacks the following:

- standard deletion verification techniques. While these techniques are necessary to eliminate the need to rely only on trust (i.e., that the provider will delete data as requested), prior research has not yet provided such verification or the requirements that need to be fulfilled to offer such assurances. Moreover, prior work lacks evidence on the challenges that providers may face providing technical proves. Assured deletion requirements are crucial for a more viable solution that can meet users and providers deletion needs.

- the ability to give users confidence when using cloud services. This part of the literature has not focused on other factors that need to be considered with technical verification to satisfy users' deletion needs. Moreover, there is no research evidence on what users' need regarding such technical verification, for instance, how users wish to receive such verifications.

In addition to SLAs, privacy policies have also been used to provide users with information about how cloud data is deleted. In general, privacy policies are expected to contain information about what data the service provider collects and how it is used, presumably providing users with sufficient knowledge to help them make informed decisions on whether they should disclose their information or stop using the service (Ion et al. 2011). Concerning deletion, they are expected to explain how data is deleted, the service provider's deletion practices including retention policy and recovery terms. Nonetheless, research (McDonald and Cranor 2008, Kelley et al. 2009) has criticised the majority of these policies – citing that they are long, unreadable and contain irrelevant information hence failing to inform users and leaving them helpless. Others have found that most of these policies are usually disconnected from privacy controls and do not match service provider's data handling operations (Anthonysamy et al. 2012; 2013). Regarding deletion, prior research fails to provide the following:

- outline what information concerning deletion should be made available through privacy policies. Existing research has not investigated privacy policies to see what information about deletion is missing or should be included to help users make better decisions with regards to cloud deletion.
- what information is essential to users with regards to deletion or where such information should be provided. Prior work does not contain any leads about where or when such information should be presented to users. Currently, there is no evidence on which method of communication works better for providing users with information about deletion or user requirements for such information.
- evidence on whether current information about deletion meets users' needs. Studies on whether current information about deletion meets the deletion needs of users are lacking.

1.4 Research Questions and Approach

The overall aim of this work is to investigate how assured deletion in the cloud can be achieved in a usable way. To achieve this aim, this thesis asks the following research questions fundamental to usability:

- **RQ1:** What technical aspects of the cloud infrastructure impact assured deletion, and what technical challenges do cloud service providers face?
- **RQ2:** What are users' general cloud deletion practices – what motivates users to delete, what factors underpin their failure to delete and their coping strategies?
- **RQ3:** What do users want regarding deletion – their deletion preferences and how they want to be informed about deletion?
- **RQ4:** How can cloud deletion mechanisms be designed and developed for users, what requirements need to be satisfied, and what principles should be followed when satisfying these requirements. Also, what concepts are important and how do they relate to one another.

To address these questions, this thesis adopts a user-centred design approach. A user-centred design approach is focused on designing computer technologies so that they are easy and pleasant to use (Dix et al. 2003).

RQ1: In order to answer RQ1, it is essential to understand the cloud infrastructure and how data is deleted from the cloud. Consequently, two scenarios (honest and dishonest cloud provider) are used to understand cloud deletion from the provider's side. Using these scenarios, assured deletion requirements are formulated. Due to the lack of literature on cloud infrastructures, an OpenStack platform (i.e., open source cloud platform) is deployed and examined to understand how data is deleted and the challenges of deleting from the infrastructure. The findings of this study are presented in Chapter 3 of this thesis.

RQ2: The second research question aimed at understanding users' deletion practices: what motivates them to delete? Can they always delete, if not, what are the factors underpinning users' failure to delete? What coping strategies do users deploy to work around their challenges? To answer these questions, 26 semi-structured interviews were conducted, and grounded theory analysis (Charmaz 2014, Stol et al. 2016) was used for data analysis. The interviews were conducted in person face-to-face (except one – skype interview) at Lancaster University. The method and the findings of the study are discussed in Chapter 4 of this thesis.

RQ3: The objective of this research question was to understand what type of data users delete from the cloud and how they wish to have these data deleted. Also, to understand what information about deletion is important to users, how and when such information should be made available to users. Using participatory research approach (Bergold and Thomas 2012), three activities were developed, and 20 active cloud users were invited to take part. To analyse this data, thematic analysis approach (Braun and Clarke 2006) was used. This method and the findings of this study are presented in Chapter 5 of this thesis.

RQ4: The last research question focused on bringing the results of RQ1, RQ2 and RQ3 together through a conceptual framework. This thesis adopts a theoretical framework analysis approach (Jabareen 2009) to identify and show how the major concepts relate to each other. The resulting framework is presented in Chapter 6.

Method Limitations

The interview method (used in Chapter 4) is costly and time-consuming. Interviews, transcribing and analysing data often take time and are dependent on the researcher's abilities and skills. This may negatively influence the results (Neuendorf 2016). Participatory action research (used in Chapter 5) may suffer from individual domination effect, that is, one or more participants may dominate or lead the discussions within the group overruling others. Interviews and participatory methods are dependent on reported behaviour, which may be different from the actual behaviour. These limitations are explained in detail in each of the chapters where the methods are used.

While the methods mentioned above delivered on the aim of this thesis, other methods could have been used to help provide more comprehensive insights to the results. These methods include laboratory studies (Jasanoff et al. 2001), ethnographic (Reeves et al. 2008) and logging studies (Newton 2000). This work could have included laboratory and logging studies where participants could be asked to complete a task (e.g., delete an item from a folder), and the actual behaviour could be observed. This could have revealed the differences between participants' reported behaviour and their actual behaviour. Ethnographic studies such as observing how participants' deletion preferences change over time would have enriched the deletion preference findings by revealing the actual behaviour over time. This could have also revealed the actual group behaviour within shared folders. These methods could have also been useful with regards to data triangulation in verifying the findings.

1.5 Contributions

The research reported in this thesis draws on results and methods from the fields of computer science, human-computer interaction (HCI), and human decision making. Through this interdisciplinary approach, this work attempts to inform platform design and data deletion policies around cloud computing by proposing requirements and design principles for assured deletion. To close the gap in the literature (see Section 1.3), the work presented in this thesis contributes to the understanding of cloud deletion in the following ways:

- **A set of requirements for assured deletion in the cloud.** These requirements enable cloud service providers to offer assured deletion to customers while not violating their SLAs. Fulfilling these requirements is helpful in two ways: (1) it gives cloud consumers the control regarding how they want their data to be deleted, and the confidence that their data privacy is protected through deletion; (2) it helps the service provider to comply with regulations and offer control to users; and (3) it helps to meet the deletion needs of users. Through a systematic review of the literature and in-depth analysis of the OpenStack platform, this work presents assured deletion challenges faced by cloud providers while attempting to offer assured deletion. It identifies critical cloud features which pose challenges to assured deletion. Altogether it offers a systematisation of requirements and challenges of assured deletion in the cloud, and a reliable reference point for future research in developing new solutions to assured deletion.

- **An empirical understanding of users' deletion practices, perceptions and coping strategies.** Through a qualitative approach (i.e., 27 semi-structured interviews) this work explores the everyday deletion decisions of cloud users across different devices and platforms. It establishes an understanding of what motivates cloud users to delete, their mental models of cloud deletion, the challenges they face, their coping strategies and what they want concerning deletion. In doing this, it discusses the implications of these dynamics for the design of cloud deletion mechanisms.

- **An empirical understanding of users' deletion preferences.** Using participatory research method (20 participants - four groups each containing five members), this work examines user cloud deletion preferences and the information that may support their deletion needs in the cloud. It specifically identifies the following: (1) how cloud users classify cloud data, individually and in group settings (2) how they want different data to be deleted, and what factors affect those choices, and (3) the feasibility to design deletion mechanisms that satisfy the deletion needs and desires of users. To that end, it also examines users' requirements regarding information about deletion. It considers: (1) what deletion information is important to users, (2) when do users want this information to be presented to them, and lastly (3) where do users prefer to find such information. This is the first study to consider cloud users' deletion preferences and needs.

- **Conceptual framework, requirements and design principles for usable cloud deletion.** Provides a conceptual framework for the design of usable assured deletion. It also identifies six (6) user requirements for usable assured deletion. Based on these requirements, this work also proposes seven (7) basic principles for usable

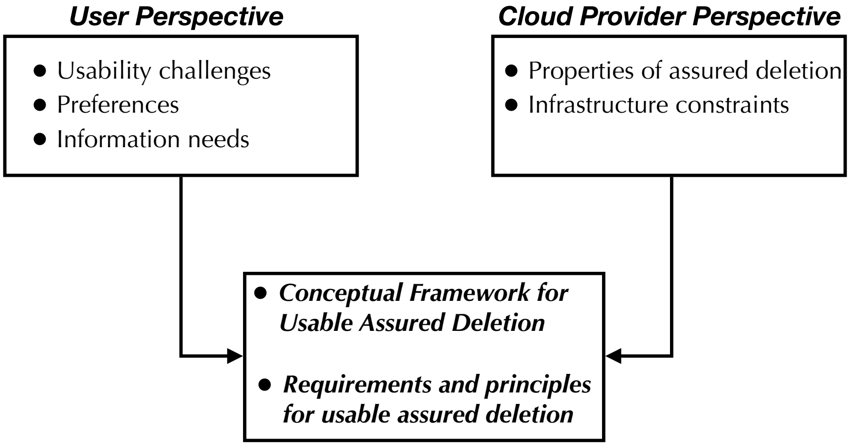

User Perspective
• Usability challenges
• Preferences
• Information needs
Cloud Provider Perspective
• Properties of assured deletion
• Infrastructure constraints
• Conceptual Framework for Usable Assured Deletion
• Requirements and principles for usable assured deletion

Fig. 1.1 Summary of the contributions of this thesis.

assured deletion that may be considered when developing a system that satisfies such requirements. These user requirements and principles are grounded in the results of the two user studies conducted as part of this work. The principles aim to inform what and how cloud deletion mechanisms and interfaces should be developed and deployed to help users delete from the cloud. The framework shows what concepts are important and should be considered during development of usable assured deletion mechanisms.

Figure 1.1 summarises the contributions of this thesis.

1.6 Novel Aspects Of This Work

The novel aspects of the work presented in this thesis are as follows:

- Unlike prior work, this work is the first to explore assured deletion from an honest cloud provider perspective, that is, a provider that desires to offer assured deletion. It highlights what makes it difficult for providers to assure deletion in the cloud. This work also considers other assured deletion properties which have not been considered in existing literature. It is the first to formulate and propose assured deletion requirements for cloud service providers that should be satisfied to assure deletion. Overall, it provides a comprehensive analysis of how salient cloud features present challenges to assured deletion in the cloud.
- Existing literature has not investigated the deletion needs of cloud users. Prior work has only focused on enterprises that desire assured deletion, neglecting users. This work provides empirical evidence on what users need concerning cloud deletion. It

provides insights on what motivates users to delete from the cloud, what challenges exist, and what mitigation strategies users develop when faced with such challenges. This research is also the first to study cloud deletion mental models and how they impact users' deletion practices. By doing this, it connects the gap between what providers offer and what users expect when deleting from the cloud.

- This research studies users' cloud deletion preferences. Prior studies have not considered the deletion preferences of users, that is, what type of data users delete and how they want such data to be deleted. This work demonstrates that users desire different types of deletion under different sharing conditions and deletion motivations. In doing so, it demonstrates the need for new cloud deletion mechanisms that will allow users to choose how they want their data to be deleted.
- Prior to this work, no study has explored information about cloud deletion, that is, what information about deletion is essential to users and where such information should be made available to users. This work investigates what, where, and when such information should be presented to users. It provides requirements for such information which is useful for informing users' decisions and helping them to recover from failures while deleting from the cloud.
- Prior studies have not considered user requirements nor principles for usable assured deletion. The work presented in this thesis formulate and proposes six (6) user requirements and seven (7) principles for usable assured deletion. These requirements should be considered in the development of cloud systems that meet the deletion needs of users, and the principles are to guide how such requirements should be satisfied.

1.7 The scope of this thesis

The scope of this thesis is limited to the following:

Cloud service providers are companies that offer storage services. The work presented in this thesis focuses on the technical aspects of cloud providers, that is, the technical infrastructure of a cloud service, not the business aspects.

Cloud users are individuals from any demographic with different skills and interests who use the cloud storage services offered by the cloud service provider.

Cloud storage is a scalable and dynamic storage resource that is offered by cloud providers. This is the storage that cloud users consume directly (e.g., Dropbox, iCloud, Google Drive) and excludes services that store or run on the cloud (e.g., Facebook, Instagram).

1.8 Thesis outline

The remainder of this thesis is structured as follows:

Chapter 2 presents the background and related work for this thesis, motivating and defending the focus on cloud deletion. It summarises closely related work on cloud computing, cloud deletion, user privacy and behaviour.

Chapter 3 presents identified assured deletion requirements and provides a systematic study of assured deletion challenges faced by cloud service providers while attempting to offer assured deletion. It is based on a publication that was presented at Cloud Computing Security Workshop (*CCSW 2016*) held at Austria in October 2016. It begins by introducing assured deletion and presenting the two adversarial models used to formulate provider requirements for assured deletion. It then discusses assured deletion requirements and challenges, demonstrating why it is difficult to assure deletion in the cloud, and finally presenting assured deletion conceptual architecture.

Chapter 4 is based on the publication titled "*'I feel stupid, I can't delete: A study of Users' Cloud Deletion Practices and Coping strategies*" presented at the Usenix Symposium of Usable Privacy and Security, (*SOUPS 2017*). This chapter introduces the first study on users' cloud deletion practices and perceptions. Through semi-structured interviews of 27 active cloud users and grounded theory analysis, this study attempted to understand users' deletion practices, perceptions, and coping strategies. This chapter concludes by discussing the implications of the results for design.

Chapter 5 describes a participatory action study that explored two areas: (1) factors driving cloud users' everyday deletion decisions and preferences, and (2) information about cloud deletion – what information is considered essential, where and when do users want this information made available.

Chapter 6 considers the previous three chapters (i.e., Chapter 3, 4 and 5). Based on the findings of these previous chapters, this chapter formally presents a conceptual framework, user requirements and design principles for usable assured cloud deletion. It begins by first presenting the conceptual framework and discusses how concepts are related to each other. It then presents formulated user requirements and principles for usable assured deletion. Each principle is presented in a stylised format which includes name, intent, motivation, example and the infrastructure constraints.

Chapter 7 concludes this thesis, highlighting the critical lessons learned from previous chapters and discussing open research challenges and future research questions.

1.9 Publications

Some of the material presented in this work has been published in various research venues while one paper is under review.

- **Ramokapane Kopo M.**, Rashid Awais, and Such Jose M., "Assured deletion in the cloud: requirements, challenges and future directions." *Proceedings of the 2016 ACM on Cloud Computing Security Workshop*. ACM, 2016.
- **Ramokapane Kopo M.**, Rashid Awais, and Such Jose M., "I feel stupid I can't delete...": A Study of Users' Cloud Deletion Practices and Coping Strategies. *Proceedings of the Thirteenth Symposium on Usable Privacy and Security*. 2017.

Other publications:

- **Ramokapane Kopo M.**, Mazeli Anthony, and Rashid Awais,"Skip, Skip, Skip Accept!!! A study on the Usability of Smartphone Manufacturer Provided Default Features and User Privacy." *Proceedings on Privacy Enhancing Technologies 2019.2 (2019): 209-227.*
- Abdi Noura, **Ramokapane Kopo M.**, and Such Jose M., More than Smart Speakers: Security and Privacy of Smart Home Personal Assistants, *Proceedings of the Fifteenth Symposium on Usable Privacy and Security*. 2019.
- Schien, D., Nolden, C., Bird, C., Rubia, C., Wilkinson, D., **Ramokapane, Kopo M.**, Preist, C., Chitti, P., Chitchyan, R., Exploring Implications of Capacity-Based Electricity Pricing for Peak Demand Reduction, *CEUR Workshop Proceedings*. 2019

Chapter 2

Background and Related Work

This chapter provides an introduction to cloud storage and assured deletion, then extensively discusses existing research considered by this work. The first section (Section 2.1) introduces cloud storage and its ecosystem (Section 2.1.1) – describing stakeholders and their responsibilities. Section 2.1.2 focuses on assured deletion; this section contextualises assured deletion, first considering assured deletion in an individually owned computer, and then in the cloud. It also summarises previous efforts that attempt to assure deletion in the cloud, identifying their limitations. The second part of this chapter, Section 2.2 discusses related work; it considers work on the following areas: deletion concepts (Section 2.2.1), deletion from social media (Section 2.2.1), users' preferences (Section 2.2.2), privacy policies (Section 2.2.3), users' mental models (Section 2.2.4), users' conceptualisation of shared folders (Section 2.2.5). Throughout this discussion, it is indicated how the work presented in this thesis make use of each of the discussed work. The last section (Section 2.3), provides a summary of open research questions.

2.1 Cloud Storage

Cloud storage is a service that extends and develops from cloud computing. It offers users with scalable and dynamic storage resources on-demand with pay-as-you-go pricing model. In the last decade, cloud storage has become immeasurably popular. It is estimated that by 2020, it would be worth around USD 71.3 billion (Cisco 2018 (accessed August 15, 2018)). The concept of cloud storage started in 2006 by Amazon when they introduced Elastic Cloud Computing (EC2) and the Simple Storage Service (S3) (Marinescu 2017). In general, the idea behind the cloud was that data could be processed and stored more efficiently on more prominent and powerful systems accessible via the Internet. While this model requires users to relinquish control of their data to the cloud provider, the benefits are many and have prompted significant adoption. These include ubiquitous access to storage, low overhead file sharing, harmonisation between devices, storage size, and real-time collaboration amenities. Cloud storage services include object storage (storage that has been optimised to handle unstructured data), database storage (offering database tables as a service) and block storage (storage provided to the customer in the equivalent of raw block device). Users access these services through different interfaces (e.g., APIs) and clients. There are several types of cloud services that exist:

Private Cloud *– the infrastructure is run solely for an organisation,*

Community Cloud *– the infrastructure is run for organisations that share common cause,*

Public Cloud – the infrastructure is made available for the general public and is generally owned by an organisation that sells cloud services, and

Hybrid Cloud *– the infrastructure is a composition of two or more cloud types (e.g., private and community).*

Storage-as-a-service gives enterprises and end users the opportunity to have access to large amounts of storage at a low-cost on demand. For instance, enterprises (e.g., start-ups) can develop applications without concerns over purchasing and maintaining the storage hardware. It also eliminates the need for up-front financial commitment, businesses can only pay for what they use and can stop using the service anytime. As the creation and consumption of audio and video content have significantly increased, cloud storage gives individual users the ability to acquire enough storage resources for themselves without the need to buy expensive equipment.

Despite the benefits, this model is full of profound implications. Control over data is moved from the owner to a third-party, and data might be processed by multiple parties at multiple sites or across borders (Marinescu 2017). This can make it difficult to determine who is responsible for what action when there is a fault in the system. It also raises security and privacy issues over data stored in the cloud – testing confidence and trust. For instance, the customer has to trust that the provider will handle data in the right manner or according to the agreement. Assurances are usually established by contracts, policies, service level agreements (SLA) and terms of service. SLA is an agreement between the provider and the consumer. It specifies how the provider will offer or deliver their services to the consumer.

2.1.1 Cloud Storage Ecosystem

To understand the issue of assured deletion in the cloud, it is essential to appreciate the cloud ecosystem. Cloud storage service models usually have two major stakeholders, the service provider and the cloud consumer.

The Cloud Service Provider

The service provider offers and maintains cloud services or the cloud infrastructure. Providers make storage services available to users through APIs. They are responsible for handling consumer data and making sure that data is available when users request it. They are also responsible for protecting it from unauthorised access and abiding by regulations. There are two different types of providers that can exist to provide storage services, the provider (e.g., Amazon S3) who owns the storage infrastructure (infrastructure as a service) and the provider (e.g., Dropbox) who rents the storage infrastructure and offer storage services to consumers (provider offers interfaces to connect to the storage).

Cloud Users - Consumers

These are the users of the cloud or consumers of cloud services; they register to use the cloud services from the provider. Users can either pay for these services or use them for free. Consumers have no direct access to the storage infrastructure but can only access the infrastructure through APIs. Consumers' interactions with the storage depend on the mechanisms offered by the provider. Cloud providers can provide user interfaces (users login and use the service) or API libraries (program and integrate their systems to the cloud). Cloud consumers can be divided into two categories: the end user and enterprise. The end-user is a home computer user, they usually sign up for free cloud services, while an enterprise is a company or organisation that pays to use cloud services, they can create applications that collect and store users' data in the cloud.

End users store and access their storage accounts through mobile applications, web interfaces and synchronisation software installed in their computers. When data is uploaded, edited or deleted from the cloud through any of these devices, the operation will be synchronised throughout all the devices. Some interfaces (e.g., mobile application) offer data on-demand – data is not stored on the device but is only fetched from the cloud when needed.

Enterprises create and deploy their services on the cloud. These applications store users' data in the cloud through the use of APIs. Users of these applications have no access to the storage APIs they can only store and manage data through the applications offered. For instance, while social media network platforms are usually deployed on the cloud, users of these networks have no direct access to the storage but through the social network.

2.1.2 Assured Deletion

Deletion has long been used to manage data that has served its purpose and is no longer needed. Similar to other cloud operations, users need assurance over deletion. Without any form of assurance over deletion or a data breach that shows a failure to delete, cloud users may never know whether their data has been deleted (or it is inaccessible). This could lead to many undesired consequences. As previously mentioned, assured deletion is not only restricted to one property – inaccessibility of deleted data but other properties as well. Table 2.1 below shows different properties prior work has attempted to cover.

Table 2.1 Summary: Assured cloud deletion properties. Prior work has attempted to provide the following assured deletion properties.

Source	Inaccessibility	Completeness	Deletion granularity	Timeliness	Usability	Proof of deletion
Rahumed et al. (2011)	✓		✓	✓		
Tang et al. (2012)	✓		✓	✓		
Wang et al. (2012)	✓		✓	✓		
Habib et al. (2013)	✓		✓		✓	
Cachin et al. (2013)	✓		✓	✓		
Chaoling et al. (2014)	✓		✓	✓		
Mo et al. (2014a)	✓		✓	✓		✓
Luo et al. (2016)	✓	✓	✓			✓
Xue et al. (2017)	✓		✓	✓		✓
Lai et al. (2017)	✓		✓			
Yu et al. (2018)	✓		✓			✓

Personal Computing Model

In the personal computing model, when a file is deleted, the system would unlink access to that particular file and remove its name from its directory. However, this process does not entirely remove the file contents from the physical drive (Reardon et al. 2013). If a sensitive file is deleted in this manner, it can be recovered from the storage device and hence remains vulnerable to malicious use. This method of deletion is considered not secure (Geambasu et al. 2009). According to prior work (Rahumed et al. 2011, Reardon et al. 2012; 2013), one property of assured deletion is achieved when an adversary with access to the system is no longer able to recover deleted data from the system. Others (Geambasu et al. 2009, Zeng et al. 2010) also suggest that data is assuredly deleted when it is permanently inaccessible to anyone upon request after it has been deleted. Several deletion methods can be used to achieve assured deletion of data in non-cloud environments, these methods are different and behave differently (Diesburg and Wang 2010, Reardon et al. 2013).

Media destruction - This technique provides great confidentiality but allows no reuse of the storage media.
Overwriting the storage media - The user modifies the storage by overwriting the data blocks, which initially stored the file with new random data. Overwriting is commonly used when the data owner also owns the infrastructure. If a user has outsourced their data to a third party, then this approach may not be feasible since the data owner may not have access to the physical infrastructure (Rahumed et al. 2011).
The use of encryption - Data is encrypted with a secret key which will be securely destroyed when the data lifecycle has ended (Geambasu et al. 2009, Diesburg and Wang 2010).

The approaches mentioned above offer some properties of assured deletion, but this is only true in a non-cloud context (i.e., when the data owner has direct access or owns the device or infrastructure which stored the data). However, in the cloud computing model, the user does not own nor have any physical access to the infrastructure. The service provider is the only one that has direct access to the infrastructure to perform such actions. Therefore, it is difficult to use these approaches to provide assured deletion in the cloud as discussed below.

Assured Deletion in the Cloud

In an attempt to offer assured deletion in the cloud, previous works (Tang et al. 2010, Rahumed et al. 2011, Tang et al. 2012, Wang et al. 2012, Mo et al. 2014a;b) have proposed different encryption-based solutions in the cloud. Sensitive data is encrypted before being outsourced to the cloud to assure deletion. Encryption keys are kept secret from the cloud provider. When data reaches the end of its lifetime, before it is deleted, encryption keys are securely destroyed, then data is deleted from the cloud. Tang et al. (2010) designed a policy-based file assured deletion system (FADE), that focuses on protecting deleted data in the cloud. Using encryption and access policies, users can revoke access to files when they wish to delete them. Each file is associated with a policy, encrypted with a data key and then a control key. The control key manages who has access to the file while the data key protects the data. When a user wishes to delete the file, the corresponding control key will be removed. Since the control key holds the data key, the contents of the file cannot be accessed, making it unreadable. To offer assured deletion of cloud backups, Rahumed et al. (2011) designed a FADE Version which is built using layered encryption. This system allows users to specify the cloud backup or the files they wish to delete. Other files and backups in the cloud remain unaffected.

While proving of deletion is more challenging than proving the existence of outsourced data, the following works have proposed various methods of proof of deletion. Xue et al.

(2018) proposed a data transfer scheme with provable data deletion. Inspired by the Merkle Hash Tree (MHT) (Merkle 1980), after data is deleted from the cloud, the cloud server follows the data deletion command by removing the corresponding notes from the MHT. After discarding the nodes, the cloud server randomly inserts some sentinel blocks where the deleted data was and then reconstructs the MHT. The reconstruction of the MHT with the new sentinel blocks generates a proof of deletion–verifying that some data has been deleted.

Using cipher-policy-based assured deletion (CPAD) scheme, Yu et al. (2018) proposed offering assured deletion and proof of deletion by using a secure symmetric encryption scheme to encrypt cloud data and then the data key of the ciphertext. The cyphertext of data and data key are stored in a fog device (i.e., a device that facilitates services between the devices and the cloud server (Tang et al. 2017)). When the deletion request is made, the fog device and the user's device can agree on a secret deletion key. Then the fog device uses the deletion key to encrypt the ciphertext related to the data stored in the cloud. Then, the fog device will send the deletion request to the cloud to remove the data. Because the ciphertext of the deleted data has changed, the user who previously could decrypt the data cannot do so anymore. To prove deletion, the user's device and communicate with the fog device using the secret deletion key.

In another attempt to proof deletion, Luo et al. (2016) proposed a scheme that disguises overwriting operations as data updating operations to delete data assuredly. When the deletion request is made, the cloud server randomly generates data and update data that is being deleted. Updating will ensure that data does no longer hold the same meaning as the previously held data. To prove data deletion, they use provable data possession protocols (PDP) (Ateniese et al. 2007) to check the integrity of the stored data. Proof of deletion is achieved if the integrity checks are different from the previous ones. Mo et al. (2014a) also adopted PDP to prove deletion. Instead of managing all data keys, the cloud user manages only one master key while data item keys are stored in the cloud. When data deletion is requested, only the data key of the requested data is deleted without the need to delete all the keys and re-encrypt all the data. To prove that data has been deleted, the user complete integrity checks (using PDP) – changes in the proof confirms data deletion.

While these studies highlight the need for assured deletion for cloud users and offer foundations for assured deletion for cloud storage, they lack empirical evidence on their usability. Also, cryptographic solutions have many shortcomings. A common problem with encryption is key management, especially when dealing with large data or many files. Users are required to manage encryption keys – storing and keeping them away from the cloud provider. Key management platforms have been recommended before but suffer the same trust issues as cloud providers. These cloud solutions face similar usability challenges that have previously been faced by email communication over the past few years (Whitten and Tygar 1999, Sheng et al. 2006, Ruoti et al. 2015). One of the first studies in usable security, Pretty Good Privacy (PGP) showed that users did not understand encryption nor the difference between public and private keys (Whitten and Tygar 1999). Encryption schemes also introduce overheads and sometimes causes bottlenecks and reduce the performance of the system (Tang et al. 2010). This previous literature suggests that assured deletion mechanisms for users should, therefore, be easy to use, less technical and not include complex security and privacy concepts. Chapter 4 and 5 examines the deletion needs and wants of users, while a detailed analysis presented in Chapter 3 discusses the challenges introduced by using encryption to assure deletion in the cloud.

The problem of the previously mentioned efforts to assure deletion is that they focus only on the customer side. They consider a situation where the cloud is public and when the provider is not trusted. These studies neglect the situation where the provider desires to assure deletion – assured deletion as a service or for ethical reasons. The current literature lacks insights on assured deletion requirements and challenges from a user perspective. In the following chapter, (Chapter 3), this thesis presents a detailed analysis of assured deletion requirements for the cloud (i.e., honest and dishonest cloud provider), and through an existing cloud platform (i.e., OpenStack) identifying and describing different assured deletion challenges that exist due to the features of the cloud.

2.2 Related Work

Table 2.2 shows a summary of the existing gaps in current literature. The following sections discuss existing literature in the following areas: cloud deletion, user preferences, information about deletion, mental models and users' conceptualization of file sharing. Each section details what existing efforts have focused on and what is missing. It will also discuss how this work fills the existing gap.

2.2.1 Usable Deletion

Deletion

While there has been some research focusing on cloud computing, user studies on cloud deletion are novel and sparse. Prior studies (Ion et al. 2011, Clark et al. 2015, Aharony 2015) have focused on users' security and privacy concerns of using the cloud. A general study on deletion conducted by Murillo et al. (2018) found that misunderstandings and unfounded expectations of deletion are limited to user interfaces. Lack of understanding of what happens

Table 2.2 Summary: Gaps identified in the existing literature

Concept	Related Research	GAP Identified
Usable Deletion	**Deletion:** Murillo et al. (2018) **Cloud:** Ion et al. (2011), Khan et al. (2018) **Social media:** Johnson et al. (2012),Almuhimedi et al. (2013), Sleeper et al. (2013), Ilia et al. (2015)	Previous work does not consider a cloud provider who desires to offer assured deletion. Usable deletion has not been considered before. There is lack of empirical evidence on users' deletion practices. Lack of systematization of assured deletion challenges.
User Privacy Preferences	**Social networks:** Gou et al. (2014), Liu et al. (2011), Misra and Such (2016; 2017) **Smartphone privacy:** Ismail et al. (2015), Hallett and Aspinall (2016), Ismail et al. (2017) **Advertising:** Melicher et al. (2016), Xu and Lee (2018) **Location:** Massa et al. (2015), Almuhimedi et al. (2015), Micinski et al. (2017) **Data sharing:** Olson et al. (2005), Mazurek et al. (2010), Sleeper et al. (2016)	No prior work considers cloud deletion preferences.
Privacy policies and Service Level Agreements	Earp et al. (2005), Pollach (2007), McDonald and Cranor (2008), Kelley et al. (2009), Ion et al. (2011), Sloan and Warner (2014)	Information about deletion in the privacy policy literature has not been considered before.
Mental models	Camp (2009b), Wash (2010) Johnson-Laird (2010), Blythe et al. (2011) **Wearable devices:** Rader and Slaker (2017) **Passwords:** Ur et al. (2016) **Experts and non-experts:** Wash (2010), Ion et al. (2015) **Multiple models:** Wash (2010), Johnson-Laird (2010)	Cloud deletion mental models are lacking in existing literature.
File sharing	Voida et al. (2006), Rader (2009), Marshall and Tang (2012), Tang et al. (2013), Voida et al. (2013), Massey et al. (2014)	Users' deletion practices (understanding, challenges and mitigations) of deleting from shared folders have not been studied in the previous literature.

on the backend may lead to unexpected results. In order to have a better understanding of deletion, users are expected to understand deletion concepts such as backend, timeliness, backup, derived information, completeness, anonymisation, deletion granularity and shared copies. However, after interviewing 36 cloud users, Ion et al. (2011) found that concepts such as timeliness or data retention are not known or well-understood among users. They also found that users did not know for how long their data is retained after they request it to be deleted. Also, users refrain from deleting old files from shared folders even when they have the rights to do so because of their lack of knowledge on how shared folders in the cloud work (Khan et al. 2018). While prior studies highlight the need for understanding deletion, none of them suggests how users can acquire such knowledge, or what users' preferences are with regards to the types of deletion they want according to what type of data. Also, there is a need to study how users delete data in settings where a file is shared. Chapter 5 of this thesis presents experimental evidence of what information about cloud deletion is essential, when and where it should be presented to cloud users.

A recent study (Khan et al. 2018), revealed a potential need for retrospective data management in the cloud. What cloud users desire to keep or delete changes over time. Khan et al. (2018) used regression models to identify user-specific and file-specific factors that would be predictive and help automate retrospective file management in the cloud. At least 83% of their participants desired to delete one of their old files either because it lacked ultimate purpose or to reduce clutter. With regards to shared folders, 55% of their participants were more likely to prefer deleting files if they were only given *view* permissions. Ion et al. (2011) found that users preferred data to be recoverable after deletion to avoid data loss after accidental deletion. The problem with the literature in this area is that there is an underlying assumption that cloud users know and understand data management in the cloud and have explicit mental models of how to use the cloud and the features offered to delete. There is no empirical evidence that cloud users can delete when they wish to, or that they understand the management of data in the cloud. However, understanding these practices would enable the development of better and usable solutions to help users delete. Chapter 4 provides empirical evidence of users' cloud deletion practices – users' cloud deletion perceptions, challenges, deletion wants, and their mental models of cloud deletion.

Deletion in Social Media

While surprisingly little work has investigated the management of unnecessary or forgotten data in the cloud, previous literature has investigated similar questions for social media and other online platforms. These domains are a useful point of comparison for cloud storage as they all allow users to share content either publicly or privately.

In social media, some aspects of deletion have been explored; for instance, there has been work focusing on understanding users' privacy concerns over social media and their challenges of deleting from such platforms (Johnson et al. 2012, Almuhimedi et al. 2013, Sleeper et al. 2013, Ilia et al. 2015). Sleeper et al. (2016) found that preserving privacy in social media is particularly multifaceted because users make changing privacy decisions based on context or situations. A later study by Johnson et al. (2012) indicated that Facebook users manage data associated with their profiles by deleting or 'untagging posts'. More than half of their participants affirmed using deletion as a way to control who has access to their posts. Some social media users deleted (i.e., unfriended) other users they did not want to share their posts with. After asking 1221 MTurk Twitter users, Sleeper et al. (2013) found that Twitter users use deletion as a coping strategy to handle regrets over posts.

In 2017, Freed et al. (2017) established that some victims of abuse delete their social accounts to disconnect with the abuser. Nonetheless, users face different challenges and conflicts when trying to delete from social media networks. For instance, when victims are part of shared social circles with their abusers (e.g., group chats), they are often faced with the decision to either delete their abusers from the group or leave the group themselves (Freed et al. 2017). Some social media accounts retain residual activities after deletion, hence providing an easy way to identify potentially sensitive information about the user (Mondal et al. 2017).

A natural way for users to protect their longitudinal privacy (i.e., management of privacy over data) is to delete old content (Mondal et al. 2017). Temporality regarding privacy is triggered whether users perceive content to be public or private. Time obscures privacy concerns, resulting in various sharing preferences, according to the specific sensitivity of the content and the context in which it was published (Ayalon and Toch 2013). The passage of time plays an important role, and users privacy preferences do change over time. Almuhimedi et al. (2013) collected 67 million deleted tweets from 292000 users and found that 700800 of those tweets were deleted within the same week while 89.1% of those were deleted within the same day of posting. Bauer et al. (2013) found users' predictions about how their privacy preferences would change over time were unrelated to actual changes over the period, implying that users cannot predict their future privacy preferences. They also found that while older social media posts were seen as less relevant, and in most cases forgotten about, some users found these posts useful for memories. Learning from this, we would expect that decisions and deletion preferences would depend heavily on the passage of time, especially if the files were automatically uploaded to the cloud and forgotten.

In other contexts, survivors of intimate partner abuse delete (e.g., harmful and humiliating content) to avoid emotional trauma. However, sometimes conflicts arise when victims are supposed to delete information from their devices. Often, victims have to keep the information (e.g., messages) to serve as critical evidence when pressing charges against an abuser (Matthews et al. 2017). When Mazurek et al. (2010) interviewed 33 Pittsburgh-based participants, they found that sharers of devices within households often use deletion to control who has access to some digital content. Since they are concerned about limiting access to their private information, they take precautionary measures (e.g., deletion) to reduce the risk of disclosure. For instance, users often delete information to prevent others from seeing it. Leon et al. (2013) found that users were more willing to share their information online when they know how long it will be kept before being deleted. Individuals were not comfortable to share their data when they knew their data would not be deleted immediately. It is clear that users delete from social media for various reasons; therefore, this work assumes that cloud deletion needs are different. It focuses on understanding the deletion preferences that may exist for cloud users.

2.2.2 User Privacy Preferences

One of the many motivations for cloud deletion is privacy (Khan et al. 2018). Prior research has extensively focused on understanding users' privacy preferences regarding different personal data and technology. Efforts have focused on online social networks (Gou et al. 2014, Liu et al. 2011, Misra and Such 2016; 2017), smartphone privacy (Ismail et al. 2015, Hallett and Aspinall 2016, Ismail et al. 2017, Abu-Salma et al. 2017), advertising (Melicher et al. 2016, Xu and Lee 2018), location (Massa et al. 2015, Almuhimedi et al. 2015, Micinski et al. 2017) and data sharing preferences (Olson et al. 2005, Mazurek et al. 2010, Sleeper et al. 2016). Results from these studies have led to various privacy mechanisms and improved user interfaces for users to control who can see their posts and avoid regrets in social media, for instance (Gou et al. 2014, Misra and Such 2016; 2017). Also, in other contexts such as online advertising users may like to state their preferences not to be tracked or select the kind of adverts they want to see, with Melicher et al. (2016) reporting that users found this beneficial and given users a sense of control. To address users' privacy concerns in smartphones, users can give various applications different access permissions (Liu et al. 2016a;b). While these studies have enhanced user control over users' data and privacy, insights on deletion preferences are missing, particularly when it comes to deleting from the cloud.

The user study presented in Chapter 5 of this thesis examines cloud users' deletion preferences. Using a participatory action research method, this work identifies how users classify cloud data and how they want it to be deleted, first in their own individual folders and then in their shared folders. This study also considers the feasibility to design deletion mechanisms that satisfy their deletion needs and desires.

2.2.3 Privacy Policies and Service Level Agreements

Today's online services use notice and choice as the paradigm for informing and seeking consent online (Sloan and Warner 2014). The notice presents the terms of "use" agreement (e.g., privacy policy) while the choice or consent signifies the acceptance of the terms of service. Users usually agree to these terms of service by either clicking the "I agree" button, or by continuing to use the service. User Agreements, Terms of Service agreements, and Privacy policies are expected to contain service provider's data practices (i.e., data collection and use) and presumably providing users with sufficient knowledge to help them make informed decisions on whether they should disclose their information or stop using the service (Kelley et al. 2009). Also, SLAs regulate the relationship between the provider and the consumer, outlining the provider's offerings and ensuring that the provider meets the requirements promised to the consumer. Nevertheless, it is common practice for free consumer cloud storage services not to offer any service guarantees (Ion et al. 2011). With regards to deletion, privacy policies are expected to explain how data is deleted, service provider's deletion practices including retention policy and recovery terms. Nonetheless, the vast amount of literature has reported on the usability problems of current privacy policies. Majority of these are long, complex and contain irrelevant information (Pollach 2007, McDonald and Cranor 2008, Kelley et al. 2009). Some are not aligned with user privacy concerns (Earp et al. 2005). Moreover, Anthonysamy et al. (2012; 2013) have found that most of these policies are usually disconnected from privacy controls and do not match service provider's data handling operations.

Moreover, understanding how the provider handles data becomes more complicated when the cloud provider relies on another (third-party) for service (e.g., Dropbox not owning cloud infrastructure but renting Amazon's infrastructure for their storage services). Research has shown that privacy policies for the "first-party" providers that users interact with are challenging enough for them to understand, but when a third-party service provider enters into the mix, the concept of effective privacy notice becomes completely unsustainable. Ion et al. (2011) suggest that cloud users do not understand these relationships; they also found that users did not know how long it took for a file to be deleted from the cloud, especially as such information is sparsely mentioned in service agreements.

Schaub et al. (2017) discussed how current privacy policies are failing to inform users and leaving them helpless. To actively help users manage their privacy, the authors suggest that control mechanisms must be relevant, actionable and understandable. They also identify four main dimensions to consider when designing to provide notice: timing, when should a notice be presented; channel, how should the notice be delivered; modality, how the information should be conveyed; and control, how choice options are integrated into the notice. In the context of cloud deletion, improving these notices will help users understand the provider's deletion practices and inform their choices with regards to deletion. The work presented in this thesis examines what information about deletion in the cloud is essential to users (modality), when should it be presented to users (timing), the channel that should be used, and integration with deletion controls.

When users want to understand how a service provider handles data or when they want to make an informed decision, or lack some skill to accomplish a task, they search for relevant information (including privacy policies) to bridge their cognitive gap. Thus, this study assumes that help-seeking behaviours involve the need for relevant information. Therefore, the research presented in this thesis seeks to find what information about deletion is essential to help users bridge their cognitive gap and considers where such information should be located for easy access. Chapter 5 shows what information about deletion is essential to users, and when it should be presented to users, and lastly, the channel that should be used to disseminate such information.

2.2.4 Mental Models

Prior literature (Dourish et al. 2004, Camp 2009b) suggests that users do not need to have a technical understanding in order to use online services. Nevertheless, when they encounter problems or get challenged using these systems, understanding of how they work may help them interpret problems, anticipate potential consequences and therefore make better-informed decisions. Current security and privacy studies (Camp 2009a, Wash 2010, Wash and Rader 2011, Johnson-Laird 2010, Blythe and Camp 2012) suggest that a critical part of users' decision making is how they see and interpret the security of existing systems. Consequently, in order to design better solutions that empower users, it is essential first to understand how they think or comprehend the systems they use.

A substantial body of research has adopted the mental model approach to understand how users comprehend the problems they encounter and the systems they use (Blythe and Camp 2012). Mental models (folk models or theories) are internal representations of how users interpret the world around them. These representations can assist in the purpose of describing, explaining, and predicting the behaviour of a system (Rouse and Morris 1986). For example, a mental model may explain a user's thought process of how an application works. Mental models are based on belief (Nielsen 2010 (accessed June 1, 2019)). Users plan and predict future actions based on their mental models (Nielsen 2019 (accessed June 1, 2019)). Moreover, these models may not always predict or result in a correct outcome. This may leave them vulnerable and failing to complete an important task successfully. For example, Wash (2010) found that home computer users were likely not to use a computer anti-virus when they had a weak or incomplete mental model of computer viruses. Mental models are dynamic and change over time, adapting to new information and features. For instance, Adams and Sasse (1999) found that as security requirements change, users tend to develop new workarounds to match those requirements.

Designers can understand users' mental models so that their products can communicate their functions through how users believe they work. A gap in understanding users' mental models may lead to failure in user interface design and leave users confused and frustrated (Nielsen 2019 (accessed June 1, 2019)). Subsequently, in other streams of literature like security (Wash 2010, Johnson-Laird 2010, Ion et al. 2015, Ur et al. 2016, Rader and Slaker 2017), mental models are well understood. Wash (2010) has studied mental models of home computer users and identified eight models that explain how users make their security decisions. Camp (2009b) has suggested that the use of mental models or security metaphors to communicate risk can yield better results. She identified five (5) different high-level metaphors that can be used to explain security advice to users better. Bravo-Lillo et al. (2013) used mental models to understand how users responded to browser warnings. Moreover, Ion et al. (2015) compared the security behaviours of computer experts and non-experts and found that they differ significantly. Ur et al. (2016) also found that users' mental models of passwords security did not match their reality.

To help users make better decisions, Camp (2009b), Wash and Rader (2011), Blythe et al. (2011), Blythe and Camp (2012) studied how these mental models can be influenced. Wash and Rader (2011) recommended promoting mental models that make or help users behave securely even when their mental models are not complete while Camp (2009b) and Blythe et al. (2011) explored the use of mental models to communicate risk.

Prior literature suggests that understanding users' mental models can help design better solutions that users find them easier to use. Nonetheless, prior work has not examined users' mental models of cloud deletion. Consequently, this thesis aims to examine users' mental models of cloud deletion. The mental model approach described in this work attempts to understand users' cloud deletion mental models; how users perceive deletion and how it affects their deletion practices. Accomplishing this task is a necessary condition for establishing what factors are essential in the design and development of deletion mechanisms for the cloud. Users' mental models may also help in identifying what information about deletion is important and should be made available to users when deleting from the cloud. Chapter 4 of this thesis presents the cloud deletion mental models that may negatively affect how users delete from the cloud.

2.2.5 Cloud File Sharing

The cloud offers new ways of sharing data across devices and users. Users are no longer forced to rely on email or pen drives to share data. With the cloud, users can access and modify shared files in the same way as accessing files on their own devices, without worrying about other collaborators with the same file. File modifications are automatically synchronised with the cloud and across collaborators' devices.

Despite the availability of file sharing features, research indicates that many cloud users still have significant misconceptions concerning how shared folders work and may not understand the implications of their actions. For example, it is not always intuitive for users to understand the difference between being a file owner or creator and being given rights to a file (Voida et al. 2013). An older study (Rader 2009) showed that users' knowledge about shared folders is biased towards the files they have added to the shared folders themselves. Rader also showed that users manage shared folder contents differently from their individually own folders, indicating some conflicting conceptual models of how to share folders. Moreover, collaborating with others sometimes reveals different assumptions about how the cloud works, leading users to develop socially negotiated practices around their use of the cloud. For instance, users find it difficult to differentiate between useful, relevant files and older outdated ones in shared folders (Rader 2009). These misunderstandings often result in them refraining from making decisions about shared files, even when they can and should. To circumvent this, they end up relegating authority to the original creator (Khan et al. 2018). Insufficient visibility of collaborators' activities has also been cited as a challenge for users when making decisions about shared content. Thus, Nebeling et al. (2015) presented MUbox, an independent collaboration tool for cloud storage to solve this problem. These authors suggest users' activity views of MUBox significantly improves users' confidence, accuracy and speed when making decisions concerning shared files in the cloud.

Tang et al. (2013) found that in many instances users' understanding of cloud storage is often shaped by their particular service provider. This sometimes includes how they solve problems or the sharing technology they replace. To help users effectively use sharing technologies, Marshall and Tang (2012) identified five (5) concepts that users need to understand. They suggest that users need to understand the concepts of personal space versus shared repositories, replication, synchronisation, triggers, and conflicts.

Some studies have focused on investigating the conflicts that arise from using cloud shared folders. More than a decade ago Voida et al. (2006) reported that users' disagreements sometimes originated from choosing the service provider, to resolve this issue, the choice of a service provider was often based on what the majority of the collaborators had access to or experience with rather than cloud offerings and cost. Other disagreements include the organisation of a shared folder or the use of emails to share other files by other participants (Massey et al. 2014). Voida et al. (2006) also suggested that the most significant challenges of using shared systems are lack of notifications and visibility of other collaborates activities.

The studies mentioned above raise various challenges experienced by users using different sharing technologies including the cloud. The research presented in this thesis considers users deletion practices from shared folders; their preferences when deleting from such folders, their challenges and the mitigation strategies they adopt. Users' perceptions, challenges and their mitigations when deleting from shared folders are presented in Chapter 4, and their shared folder deletion preferences and the essential information to help them delete from shared folders are found in Chapter 5 of this thesis. Design guidelines for shared folders are found in Chapter 6.

2.3 Summary

Research in cloud computing has focused on the security and privacy of the cloud neglecting one crucial aspect of data management – deletion. None of the present user studies have considered users' deletion practices, needs and preferences. In summary, prior research has not considered the following:

- Requirements for assured deletion from the cloud providers' and users' perspective, and whether these requirements are aligned. Previous research on assured deletion assumes a dishonest provider who has no interest in providing assured deletion, and only from the enterprises' perspective neglecting an everyday computer user. Furthermore, these efforts propose the use of encryption which limits data use and also has usability and key management issues. Also, the use of encryption only satisfies one property of assured deletion – inaccessibility of deleted data. This thesis holds that one of the key factors limiting advancement in assured deletion is that other properties of assured deletion such as timeliness, appropriate feedback, proof of deletion, deletion preferences, deletion granularity and complete deletion have not been explored. Also, existing literature does not consider an honest cloud provider that desires to offer assured deletion.

- Usable assured deletion. Previous work on cloud deletion focuses on users' security and privacy perception of the cloud. None of the previous studies on cloud computing considers usable cloud deletion as a major concern for cloud users. Despite users deleting data every day, existing research has not considered how users delete and understand cloud deletion. There is a literature gap in understanding users' perception of cloud deletion (e.g., mental models) and their practices. At present, it is not clear what motivates cloud users to delete or the challenges (if any) they encounter when attempting to delete from the cloud.

- Users' deletion preferences — how users want their data to be deleted. While all these factors are essential in the development and design of cloud deletion mechanisms, there is no empirical evidence of these practices and preferences. With regards to this, the research presented in this thesis explores the following: (1) users' deletion practices (what makes users delete, the challenges they experience, their mental models of cloud deletion, the mitigations they adopt when facing challenges, and what they want with regards to deletion), and (2) deletion preferences (what data they delete, how they want that data to be deleted, and what affects those choices).

- Information about deletion. Though information is usually used to make an informed decision, existing studies have not explored what information about deletion is essential and relevant to cloud users. It has not investigated how or where such information can be shared with users. Privacy policies have been found unsuitable to have all the information because they are long, use ambiguous language and sometimes hard to map to controls. Furthermore, it is not clear when this information should be made available to users. One of the objectives of this thesis to study and understand what type of information is crucial to users, and when and where such information can be made available. This work argues that essential information about deletion should not be restricted to privacy policies only because of the various challenges they pose on users.

Chapter 3

Assured Deletion Requirements and Challenges

The content of this chapter is based on the following publication:

Ramokapane, Kopo M., Awais Rashid, and Jose M. Such. "Assured deletion in the cloud: requirements, challenges and future directions." *Proceedings of the 2016 ACM on Cloud Computing Security Workshop.* ACM, 2016.

Previous studies have focused on assured deletion from an enterprise perspective, and assuming that the cloud provider is not interested in providing assured deletion. However, as shown in Section 1.2, assured deletion is not only important to enterprises, it is essential to both service providers and users. This chapter investigates assured deletion from the service provider's side. It provides a systematic review of assured deletion in the cloud, discussing the requirements for assured deletion, and the challenges faced by cloud service providers while attempting to offer assured deletion. The objective of this chapter is to acquaint the reader with assured deletion requirements and challenges that exist in the infrastructure, and the limitations of existing solutions. The key contributions of this chapter are as follows:

- It discusses the issue of assured deletion in the cloud from two perspectives; the dishonest provider and the honest provider, providing a unique mapping between requirements and challenges.

- It identifies critical cloud features which pose challenges to assured deletion and offers a systematic analysis of these challenges for assured deletion for both cloud tenants and providers. Thereby, demonstrating how difficult it is for providers to assure deletion.
- It provides an in-depth analysis of OpenStack (Cloud platform) feature set and the challenges that various features pose with regards to assured deletion.
- It discusses open challenges for assured deletion in the cloud, for both the tenants and the providers.

To contextualise the challenges of assured deletion in the cloud, this chapter presents two adversarial models, one involving a *dishonest cloud provider* and the other a *honest cloud provider*. Using the *dishonest cloud provider* scenario, it formulates and presents requirements for assured deletion for a cloud tenant (who wants assured deletion), and then analyses existing solutions for assuring deletion in such scenarios. It then considers the *honest provider model*, a situation where the provider wants to assure deletion and draws assured deletion requirements in such a context. Both scenarios assume the cloud tenants desire to have their data assuredly deleted from the cloud infrastructure. The difference between the two models is that in the first model (i.e., dishonest provider scenario) the cloud provider is acting as an adversary while in the second model (i.e., honest provider scenario) the provider is not an adversary and is willing to provide assured deletion as a service. These requirements were formulated using different ideas from existing work on assured deletion. Other attacks (e.g., side channel attacks (Ristenpart et al. 2009, Irazoqui et al. 2015, Liu et al. 2015) are out of our scope since there is already a substantial amount of research on such attacks. This work is significant with many benefits. For researchers, it offers an initial study on assured deletion in the cloud from both the provider and the tenants; for practitioners, it provides a comprehensive study of the existing solutions and identifying limitations and challenges in the area.

This chapter is structured as follows. Section 3.1 presents the first adversarial model (Dishonest cloud provider) and the formulated assured deletion requirements. It then presents a review of the existing solution against the formulated requirements. It concludes by presenting the limitations of the existing solutions and possible future research areas. Section 3.2 presents the second adversarial model (Honest provider). It starts by presenting assured deletion requirements for an honest provider, followed by an analysis of OpenStack (cloud infrastructure) with regards to assured deletion and then culminating in open research problems. Section 3.3 posits and discusses an initial conceptual architecture for assured deletion in the cloud and then (Section 3.4) concludes the chapter.

3.1 Dishonest Cloud Provider

> This is a situation where a cloud tenant is using a public cloud provider for storage and data processing services. The tenant outsources their data to the cloud but is suspicious and sceptical about the provider's data disposal responsibilities. It is also assumed that the tenant is aware of the risks involved in the incomplete or impartial deletion of data and does not want to be a victim of data leakage. Without losing any benefits of using the cloud or incurring any extra cost, the tenant wants to ensure that data will always remain safe even after deletion. Since the cloud is a multi-tenant environment, this scenario also assumes that the provider is not only curious about tenants' data, but it also has some other malicious tenants who are also interested in getting hold of other tenants' data. During provisioning of services, the provider's insiders (e.g., system admin) may probe freed resources or decommissioned servers for tenants' data. A malicious tenant may request for more storage resources from the cloud provider, but before writing any data to the availed resource, the malicious tenant probes the provided resource for any sensitive data that may have been left behind by a previous user.

In a situation where a tenant is the one interested in assured deletion, and the provider is dishonest, the tenant has to carry the burden of ensuring that data is inaccessible after deletion. The following assured deletion requirements are drawn using this model. These are requirements that may be considered when a cloud user does not trust the provider or desires assured deletion without the involvement of the provider.

3.1.1 Requirements for Assured Deletion

Deletion granularity: Deletion should allow the tenant to target particular data to delete while not affecting other data. Other data should not be affected by the deletion process, and it should remain useful and accessible to the tenant. The ability to only delete some data would give users more control over what to delete reducing the cost of deletion.

Usability: The deletion should not negatively affect the users' use of the cloud. The tenant should be able to delete data without affecting their daily use of the cloud. The deletion process should also not burden the user, and it should be simple to complete without accumulating any usability cost or affecting the user's productivity.

Cloud Computation: The use of data should not be affected, that is, the cloud tenant should continue to work with data (e.g., perform computations) as before without any problems. They should be able to complete all the necessary data operations, for example, sorting and searching.

Complete deletion: Assured deletion requires that the deletion should affect all copies of data associated with the deleted data including the metadata. When the tenant request data deletion, it should affect all the copies associated with the data that is being deleted. Deletion should include metadata associated with the data that is being deleted.

Timeliness: Deletion should be completed promptly; deleted data should not be accessible from the environment immediately after the deletion process has started.

From a tenant's perspective, it is always challenging to assure deletion in the cloud since there is little physical that a tenant can do – tenants do not have access to the actual physical infrastructure, therefore, they have limited options. In most cases, tenants are forced to explore other complete deletion approaches that do not require modification of the infrastructure. These solutions are mostly based on cryptographic schemes. The next section discusses encryption as a solution to guarantee deletion in the cloud. Firstly, it surveys and discusses existing encryption approaches that have been proposed in the literature to assure deletion. Secondly, it presents the challenges and the limitations of these solutions against the earlier requirements in Section 3.1.1. The last section presents the opportunities and the areas that could be explored in order to improve the existing solutions.

3.1.2 Existing Approaches

This section aims to present and discuss existing encryption solutions proposed in the literature against the assured deletion requirements outlined earlier.

To assure deletion using encryption, sensitive data is encrypted before it is outsourced to the cloud. Encryption keys are kept secret from the cloud provider. When data reaches the end of its lifetime, before it is deleted, encryption keys are securely destroyed, then data is deleted from the cloud (Geambasu et al. 2009, Reardon et al. 2014). Consequently, without encryption keys, deleted data is deemed inaccessible.

One approach which uses cryptography to assure deletion in the cloud is FADE (Tang et al. 2010). It supports policy-based assured deletion, whereby each file has its own access policy which is also associated with a control key. Before being farmed out to the cloud, files are encrypted with data keys and then control keys. Encryption keys are kept secret and managed independently by a key escrow system. Assured deletion is achieved when the associated file's access policies are revoked. This approach achieves deletion granularity because it allows tenants to select and only revoke access policies of the target file. Inaccessibility is also achieved immediately when the access policy is revoked. However, this approach does not consider a scenario where multiple versions of a file exist, which is a case in most cloud setups.

FadeVersion (Rahumed et al. 2011) aims to extend FADE by considering a case where multiple versions of a file exist. Each file version has its own encryption key which, when revoked, the version is assumed to be assuredly deleted. Like FADE, assured deletion granularity can be achieved by only revoking the access policy of the version one wants to delete. Again, inaccessibility is achieved as soon as access policies are revoked. Nevertheless, it should be noted that this approach increases the number of encryption keys for the third party to manage.

Mo et al. (2014b) explore the feasibility of assuredly deleting data using encryption without involving a third-party. Encryption keys together with sensitive data are outsourced to the cloud provider but are inaccessible to the provider. The authors claim to prevent any potential data privacy breach by using a multi-layered key structure, called Recursively Encrypted Red-black Key tree (RERK). RERK assumes the responsibility of securing data and encryption keys stored in the cloud. The cloud user still maintains a master key or metadata which helps to manipulate encryption keys and data stored by the provider. Deletion granularity is supported, but the issue of deleting multiple copies remains unaddressed.

3.1.3 Challenges and Limitations

While these solutions do satisfy some of the requirements for assured deletion, they also have various limitations. These are discussed below.

Key Management: Some enterprises and users have struggled to adopt encryption technology because of the overhead performance issues and key management issues. When using encryption, the client does not only need to deal with protecting the data but also needs to protect the encryption keys themselves. As a result, some users have adopted third-party key management systems. FADE and FADEVersion base their trust in a key escrow system to manage the encryption keys. However, if one cannot fully trust the cloud provider, should this not place the same doubt on these third-party services? These services are vulnerable to the same adversaries as the cloud providers (e.g., insider threats). Moreover, in practice, they may introduce some bottlenecks hence reducing performance. For instance, when using a third-party key manager, one has to first access the keys before they can access the data securely. Sometimes this process can delay and even affect the performance of the service.

Another challenge is deciding on the granularity of the keys, whether one key should be used for all or whether each file should have its own key. Mo et al. (2014b) and Rahumed et al. (2011) consider deletion granularity in their solutions, both solutions suggest that each file should have its own encryption key. Nonetheless, this increases the number of encryption keys one has to deal with and may be cumbersome to manage, leading to usability problems. Moreover, sharing encrypted data also becomes a problem, especially when more than one user needs to access the encrypted data. This problem is explained later in detail in the next sections. Nonetheless, key management systems are sometimes proprietary and support a limited number of cloud providers; cloud tenants may struggle to find an appropriate key management system.

Usability: Integration of encryption into systems has shown to have a negative impact on users (Whitten and Tygar 1999, Smith 2003, Furnell 2005). Research has shown that users struggle to use encryption systems and are likely to make mistakes, especially with encryption keys. It is possible that cloud users may make mistakes when deleting keys as part of assured deletion. For example, a user may delete a wrong key or forget to encrypt a file after use. Another concern is the issue of effort (i.e., the addition of extra steps into the process of completing a daily task). Extra steps often are circumvented by users when they get in the way of their daily activities (Adams and Sasse 1999). Moreover, users are not only required to understand why such solutions are in place but also, their responsibilities to make such solutions work effectively.

Limited Data Use: When data is encrypted outside the cloud, there is little to do with it in the cloud unless the keys are also shared with the cloud provider. Encryption limits data use. Current solutions do not mention how this issue will be addressed if the cloud tenant also uses the cloud to process data. Some cloud applications do not yet accommodate the use of encrypted data. When encrypted, data loses some of its properties such as the ability to be searched or sorted. A naive solution to circumvent this would be to download it first before performing such operations. Still, this can be burdensome on the client hence reducing productivity.

Though there have been some proposed solutions (Naehrig et al. 2011, Van Dijk et al. 2010, Curtmola et al. 2011) to tackle the issues mentioned above, most of them are still in their infancy and are not yet practical to be adopted in cloud applications. For example, most of the homomorphic encryption schemes that have been proposed only support a limited number of operations (Fan and Vercauteren 2012).

Data Sharing: Encryption makes sharing data difficult especially when all involved parties need to access the same data. Current cryptographic schemes depend on using a single set of secret keys, therefore, allowing access to a single user. In order to share data, users end up sharing encryption keys which can be costly during key revocations as it requires all the data to be re-encrypted (Dong et al. 2011). Moreover, sharing encryption keys leaves both the data and the keys vulnerable.

One of the benefits of cloud computing is high availability, the ability to access data regardless of location and time. However, with encryption in place, this can become a challenge as users may need to share encryption key between devices hence putting keys at risk.

Table 3.1 Summary: Open Research Areas

Emerging Solution	**Challenges**
EM1- Homomorphic Encryption	- Impractical for cloud applications. - Causes a lot of computational overhead.
EM2- Partial Homomorphic Encryption	Support limited amount of datasets (Barhamgi et al. 2016b). - Does not support all queries. - Causes a lot of computational overheads.
EM3- Searchable Encryption	- Does not support all queries. - Causes a lot of computational overheads (Van Liesdonk et al. 2010).
EM4- Trusted Computing	- Expensive - User needs to transfer secret key to the trusted hardware (Barhamgi et al. 2016b). - Compatibility with current infrastructures. - Verification and attestation (Santos et al. 2009)

3.1.4 Areas to explore - Dishonest cloud providers

Emerging solutions employ encryption for scenarios where the tenant does not fully trust the provider. However, as mentioned earlier current encryption still has many limitations. Table 3.1 summarises some of these solutions with their limitations. Research should focus more on to improving the current encryption schemes (e.g., searchable and homomorphic encryption schemes) that are being used to compute data. Development of encryption ready cloud applications might be an interesting area to explore. More work could focus on the concept of trusted computing; future trusted computing solutions could be used alongside encryption. Also, cloud providers could offer secured hardware containers within their cloud infrastructures to store data within secured areas (Barhamgi et al. 2016a).

3.2 Honest Cloud Provider

This model considers a scenario where a public cloud provider is honest but prone to accidental data leaks due to incomplete deletion. Due to escalating number of incidents of data leakages in the cloud, the provider is conscious about reputation and does not have any intentions to leak tenants' data. It is also assumed that the provider has other security mechanisms in place to protect the tenants' data from other attacks which could lead to data loss. Additionally, it is in the interest of the provider to provide confidentiality and comply with legal and standard regulations. Despite the cloud provider's good intentions, malicious tenants may randomly probe their resources for partially deleted data. Again, unless data is completely deleted from the infrastructure, malicious attackers may target decommissioned machines from the cloud provider to steal data.

As mentioned earlier, there are many reasons why a cloud provider may want to provide assured deletion as a service; it might be to comply with standards or to enhance reputation in the market. This section summarises the cloud provider's requirements for assured deletion using the honest cloud provider model.

3.2.1 Requirements for Assured Deletion

Service availability: Assured deletion should not negatively affect services offered to other tenants or the tenant who requested data deletion. Any service disruption would cost the provider.

Complete deletion (irrecoverable and inaccessible): Data should be removed from all the cloud layers and components that handle or store data. Data (including metadata) residing in temporary locations such as buffers and RAM should also be wiped out completely. Deleted data should not be accessible after deletion, either to the provider or other tenants. Data should be wiped entirely from the provider's infrastructure.

Deletion of all backup copies: The provider should be able to completely delete all backup data that is no longer needed or required by the tenants. This includes data stored locally and remotely by the provider.

Deletion granularity: The service provider should be able to assure deletion on target data (i.e., data which is to be deleted and requested by the tenant). When the tenant requests for data deletion, the provider should only be able to completely delete the requested data without affecting the tenant's other data or data belonging to other tenants. Deletion granularity reduces the cost of secure deletion operations.

Delete latency: Timeliness to guarantee deletion - this refers to the time agreed by the provider to assure deletion to a cloud tenant or the time required by any regulatory standards to assure deletion. The cloud provider should be able to completely delete data within this time.

Error handling: Deletion procedure should complete without any errors, and if errors do occur, the deletion procedure should be able to recover and complete within a reasonable time.

Proof of deletion: An honest cloud provider should be able to prove deletion to tenants, attesting that data deletion has completed successfully. For example, if a deletion procedure has completed successfully, it should return a signature (i.e., adequate feedback) which can be verified by the client.

Usability: Data deletion mechanism should be easy to use. The interface should be clean and the deletion controls should easily be accessible. Deleting from the cloud should be simple and easy to complete.

In the absence of existing literature targeting this scenario, this thesis analyses OpenStack, an existing open source cloud infrastructure. This section considers different features and characteristics of OpenStack that pose different challenges to assured deletion. It discusses these challenges against the requirements outlined earlier.

3.2.2 Case Study: OpenStack

OpenStack is a fully distributed open source infrastructure as a service (IaaS) cloud platform used for building public and private clouds infrastructures. It is used to control large pools of computing, storage, and networking resources throughout a datacentre. NASA and Rackspace were the first companies to develop and contribute to the OpenStack project, but now it has more than 30 contributing companies (OpenStack 2019 (accessed May 25, 2019)).The main core of OpenStack is fully developed in Python and has powerful APIs which allow it to be interfaced with other cloud computing platforms such as Amazon EC/S3, making it possible to offer hybrid cloud services.

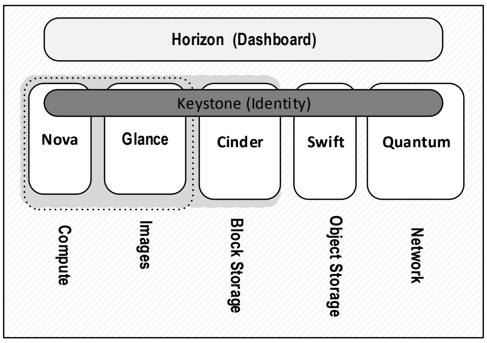


Fig. 3.1 OpenStack Core Components

OpenStack (Kilo) Key Components

OpenStack Compute (Nova): *The main component of Openstack; its main purpose is to provide virtual instances on-demand.*
Openstack Image Service (Glance): *Glance is a virtual machine image repository service which provides services for discovering, registering and retrieving virtual images in OpenStack.*
Network Service (Neutron): *Provides network connectivity between interface devices in OpenStack.*
OpenStack Object Storage (Swift): *A highly available and distributed Object Storage.*
Block Storage (Cinder): *Provides persistent block storage for virtual instances.*
Other components include:
Keystone: *Handles Authentication and authorization for OpenStack.*
Horizon: *is a web based graphical user interface for OpenStack components.*

Fig 3.1 shows the key components of OpenStack.

Table 3.2 Apparatus used for analysing OpenStack

Apparatus	**Version**
Operating System	Centos 7.1-1503
OpenStack Platform	Openstack *Kilo*
Server Hardware	Dell PowerEdge *R720*
EnCase	*v6.19*
Dell laptop	Windows 7

Experiment Setup

OpenStack *Kilo (2015.1.0)* was installed on a Dell PowerEdge server (2 Xeon cores, 384 GB RAM, 2*800GB SSD, 6 *1.2 TB) running CentOS 7.1. An EnCase v.6.19 forensic environment was also setup in a Dell laptop (8GB RAM, 500GB) running Windows 7. Table 3.2 shows all the apparatus involved.

Table 3.3 Experimental tests.

Test 1 – Deleting and recovering virtual machine instances.
1. Deploy and run a virtual machine in OpenStack.
2. Terminate and delete the virtual machine.
3. Image the cloud environment.
4. Inspect the image for virtual machine remnants after deletion.
5. Report the results.

Test 2 – Deleting and recovering data from the object storage.
1. Deploy and run a virtual machine.
2. Create and attach the object storage.
3. Store data in the object storage using the virtual machine.
4. Delete data stored in the object storage.
5. Image the cloud environment.
6. Inspect the image for data remnants from the storage after deletion.
7. Report the results.

After learning about the functionality of different OpenStack components, the subsequent step was to run two different cloud deletion tests. Test 1 involved deleting a virtual machine instance and investigating whether the virtual machine or its information are recoverable after deletion. Test 2 investigated whether data stored in the object storage could be recovered after deletion. Table 3.3 describes the steps followed during the two tests.

Experimental Challenges

Due to unclear documentation on how to install OpenStack, it took six months to have a stable environment running. The documentation lacked clear instructions on the order of the scripts that are needed to install OpenStack. If the order of the script was wrong or if one of the scripts failed, the whole installation process had to be started all over again, including the installation of the operating system. To circumvent losing more time, the researcher had to get help from other researchers who had experience working on OpenStack.

During the forensic investigations, imaging the OpenStack environment took several hours and sometimes completed with failures. Also, because of the size of the images, EnCase software often failed to scan the whole image. To avoid these challenges, the researcher installed OpenStack on a single hard drive, and the imaging was only performed on the hard drive that had OpenStack installation.

OpenStack - Assured Deletion Challenges

Like other cloud platforms, OpenStack has features and components that lead to challenges regarding assured deletion. These challenges are discussed next.

Many Components. OpenStack has a very complex architecture with many different components. Numerous services and components capture, maintain or reference tenants' sensitive data (OpenStack 2019 (accessed May 25, 2019)). Besides OpenStack storage components, tenants' sensitive data may also be found in these components: *Neutron*, *Glance*, *Keystone* and *Nova*. These data include information about the instances, storage volume data and public keys while sensitive metadata may include organisation names, internally generated private keys and users' real names. While OpenStack is expected to delete this data when required to do so, some data may remain in the platform after deletion (OpenStack 2019 (accessed May 25, 2019)). Layers, virtual and physical components may still contain data and making it difficult to locate during a deletion request. As a result, it is difficult to completely delete data from OpenStack due to the complexity of the architecture.

Virtualization Technology. In OpenStack, deletion is also subject to virtualisation. To understand this, consider a tenant launching and terminating a virtual instance in OpenStack.

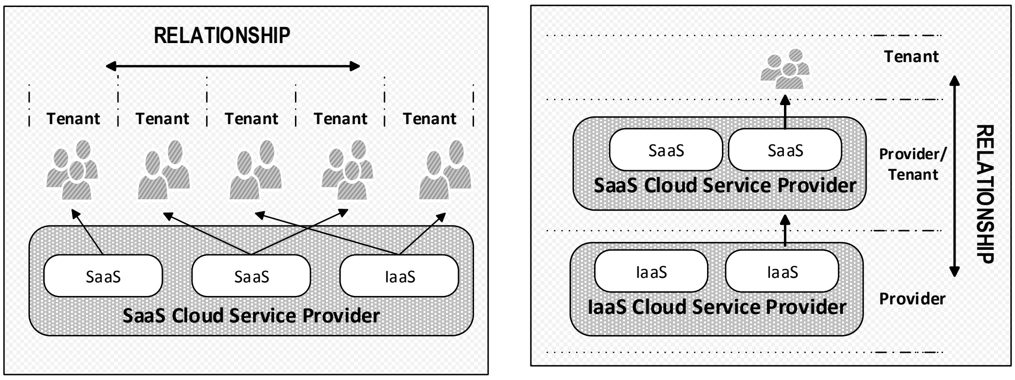
> *When a tenant launches an instance, OpenStack makes a copy of the base image from Glance to the local disk of the compute node. A new ephemeral volume is created and then attached to the running instance. For persistent writing, block storage can then be attached to the instance. When an instance is decommissioned, the metadata of the original image is updated with the new metadata from the running instance. All unsaved data is then saved to the block storage before it is detached. The previously created ephemeral storage is then deleted before memory and CPU resources are released (OpenStack 2018 (accessed May 25, 2019)).*

One of the requirements for assured deletion for the provider is the ability to make data inaccessible within an agreed time. However, the type of virtualisation in OpenStack may not always guarantee this. The destruction of images and the ephemeral storage is dependent on the hypervisor, and the compute plug-in used (Fifield et al. 2014, OpenStack 2019 (accessed May 25, 2019). Some plugins (e.g., libvirt) would not overwrite the memory block that was previously used by the running instance. It is assumed that the dirty memory extents will not be made available to other tenants, but this cannot be guaranteed. It is difficult for the provider to predict or estimate when the dirty memory would be overwritten. Also, in practice, hundreds of instances are launched and decommissioned all the time. Therefore, it is not always feasible for the provider to confirm or provide proof of deletion after every terminated instance.

Live Migration. There are two types of migration supported by OpenStack: True Live migration and Non-live migration (or simple migration). Unlike in other cloud settings where live migration is a load management feature, in OpenStack this is an operational management feature. For example, migration may be required when performing upgrades that need system reboot of the physical server. Live migrations in OpenStack take place under controlled conditions, and cloud administrators initiate them. This increases data surface area. For example, after the virtual instance is successfully copied to a new location, there might still be data left on the previous host. OpenStack currently does not have any feature to check if all data from the source host has been deleted. While the hypervisor might delete data after migration, the provider cannot assure that data and metadata of the migrated virtual machine is completely removed from the source environment (OpenStack 2019 (accessed May 25, 2019)).

Virtualized Storage. Deletion of data from *Swift* (Object storage) is a complicated process, which may not result in complete deletion. Like other virtualised storages, *Swift* does not provide direct access to raw blocks of data but rather provides blobs of data (data objects) which represents "volumes", and they are accessible through an API (Arnold 2014). These objects can be stored anywhere within the storage infrastructure or even replicated to other *Swift* clusters. Since this is virtualised, the pool of storage provided may span several physical servers or drives. This allows data to be scattered all over the storage infrastructure. Openstack uses *Swift* APIs to access data from the object storage for processing. It is difficult to assure deletion because the physical location of data is abstract and may not be known, and also, deletion methods such as overwriting may not overwrite the objects that users intend to delete but rather create new ones. Virtualisation layers may also possess some data that may be left behind since deletion is not done at all layers. Deletion of backup copies in other *Swift* clusters may not happen in time or at all. Also, some data may not be deleted since deletion is not performed at virtualisation layers.

Horizontal Multi-tenancy. Cloud computing allows multiple tenants to share the same physical infrastructure, and each tenant has their own relationship with the provider. Multi-tenancy relationships in the cloud can either be horizontal or vertical depending on the services being provided. Horizontal multi-tenancy relationship exists when a provider is offers its services directly to multiple tenants simultaneously as shown by Fig 3.2a, while



(a) Horizontal Multi-tenancy relationship (b) Vertical Multi-tenancy relationship

Fig. 3.2 Multi-tenancy relationships in the cloud

vertical multi-tenancy exists when a provider offers its services to multiple tenants who also act as service providers by leasing some resources to other tenants (See Fig 3.2b). In these kinds of setups, it can be challenging for the provider to assure deletion as tenants' data may be residing in the same physical location or even tangled together. In horizontal multi-tenancy, it is possible that when a tenant is decommissioned, their storage partition will not be securely wiped out completely before it is made available to new tenants. This may be due to a number of reasons: (1) the hypervisor used as explained earlier under virtualization technology, (2) secure deletion techniques like overwriting may not be feasible while provisioning services, (3) the secure deletion methods used may not offer deletion granularity and leave other tenants' data at risk during the deletion.

Vertical Multi-Tenancy and Third Party Backup Providers OpenStack interfaces allow it to be integrated with other cloud computing platforms either for storage or computing services. For example, OpenStack can be interfaced with Amazon S3 for storage. This would form a vertical multi-tenancy relationship. It becomes challenging to delete data that leaves OpenStack because it may be out of reach. Another challenge is that if these APIs fail then, the provider has no other way to verify the deletion in the other infrastructure. Complete deletion may require the involvement of the third party, which may add extra cost to the provider; again, it may not happen as the other provider may face the same problems highlighted in this thesis. Deletion latency could also be a problem.

Delayed Deletion. Another cloud feature that challenges the requirements for assured deletion is the delayed deletion feature. By default, most providers use a garbage collection method of deletion, this is whereby deleted data is only marked for deletion but not removed from the system immediately. This feature is enabled to protect tenants' data from accidental deletion operations or administrative errors. Sometimes, providers have to delay deletion because of legal obligations or policies. When a tenant deletes a file, the cloud system will only mark it for deletion and remove it from the tenant's interface while it remains in the cloud system. Therefore, deletion does not happen as soon as it is requested.

In OpenStack when delayed deletion is enabled, deleted accounts, images, volumes and other data are not removed from the platform upon request (Arnold 2014, Fifield et al. 2014). By default, there is a 7-day delay before an image is removed from the platform. When an image is deleted, it is marked for deletion and given an expiry date. While it is removed from the tenant's reach, it is still part of tenant's data. The removal of the image is then left for the reaper process (i.e., a periodic background process that is responsible for removing all data marked for deletion after retention time has expired (OpenStack 2019 (accessed May 25, 2019)). It is difficult to prove this deletion or estimate the time it takes to delete data from the provider completely. Data may still be accessible after deletion.

Execution Errors. OpenStack platform is not immune to failures. Its activities are vulnerable to errors. Migration, the creation of instances, volume attachments, deletion and many other processes may fail one way or another. Failures may be due to network errors, hypervisor errors, inadequate resources or even administrative errors. When an instance fails during provisioning, the service provider cannot guarantee that the tenant's data that was being processed at the time will be eradicated. The same problem exists with other components: when a component fails, the provider cannot guarantee whether data was completely removed from the component before it encountered errors. In this case, deletion assurances may be difficult to guarantee.

High Availability. According to OpenStack documentation (OpenStack 2018 (accessed May 25, 2019)), OpenStack can meet high availability requirements for its own infrastructure services by up to 99.99% uptime. Images, containers, objects, volumes and metadata can be replicated in OpenStack to allow high availability requirements. A process called *the replicator* is responsible for producing multiple copies of data and syncing all data. Through Cinder Backup API, OpenStack allows volume replication, multiple backend backup system with tiering technology and compression. This API also allows the import and the export of tenants' backup service metadata. It may be difficult for the provider to guarantee that all data copies and metadata will completely be deleted when tenants request for deletion.

Challenges that depend on deployment. Other challenges manifest on OpenStack but are dependent on deployment. These challenges include: Underlying hardware (e.g., the use of Solid-State Drives (SSDs)), Storage Tiering and Thin Provisioning, Multiple physical locations, Offline Backup, Different storage media and Third-party providers.

Table 3.4 OpenStack cloud features and their respective challenges against assured deletion requirements

OpenStack feature	Requirement Challenged	Key Challenge
Infrastructure	Complete deletion Deletion of all backup copies Deletion latency Proof of deletion	Multiple Components
Virtualization	Complete deletion Deletion of all backup copies Deletion granularity Error Handling	Virtualization Technology Virtualized Storage Underlying Hardware Virtual Instance Operations Multiple Dynamic Logical layers
Multi-tenancy	Service Availability Complete deletion Deletion granularity	Horizontal Multi-tenancy Vertical Multi-tenancy
On-Demand Elasticity	Complete deletion Deletion of all backup copies Error Handling	Live Migration Storage Tiering and Thin Provisioning Execution Errors
Backup and High Availability	Complete deletion Deletion latency Proof of deletion	Multiple Locations Offline backup Storages Different Storage Media Third-Party Providers Delayed Deletion

Different cloud features pose different challenges with regards to assured deletion. In OpenStack, these challenges include Live migrations, virtualised storage, execution errors, vertical and horizontal multi-tenancy, delayed deletion, high availability and virtual technology. Table 3.4 presents a summary of the list of cloud features and the key challenges they pose with regards to assured deletion. The second column shows the requirements impacted by these challenges while the third column shows the key challenges.

3.2.3 Areas to Explore - Honest Provider

This section presents some future research directions for assured deletion in the cloud concerning a provider wishing to provide assured deletion as a service. These research directions are classified according to the formulated requirements in (Section 3.2.1).

Service availability. Service availability is essential to cloud providers and tenants. Interruption of service for seconds could mean huge financial losses for the provider. Research could find ways of assuredly deleting data without affecting the service. For example, applying secure deletion methods (e.g., overwriting) without interrupting the service or deleting other tenants' data. Research should look into solutions which do not only rely on secure deletion methods (e.g., overwriting) which depend heavily on the properties of the underlying physical storage as physical control over infrastructure in cloud computing is no longer feasible (Cachin et al. 2013). Again, in order to assure deletion without any service disruption, data isolation mechanisms could be developed to isolate data that would require assured deletion. Restrictions could be applied to sensitive data. For example, sensitive data could be restricted to specific places in the cloud.

CloudFence (Pappas et al. 2013) and Cloudfilter (Papagiannis and Pietzuch 2012) are some of the approaches that have been proposed to restrict data movement in the cloud. Cloudfilter proposes a practical service-independent system that uses policies to restrict data movement between enterprise and cloud service provider while Cloudfence focuses on confining sensitive data to a single defined domain thus preventing the propagation of marked data to other parts of the cloud infrastructure. Researchers could leverage these approaches and use the tracking capabilities to confine sensitive data to locations where assured deletion methods could be applied in the cloud. For instance, sensitive data could be restricted to disks that allow scrubbing.

Complete deletion (irrecoverable and inaccessible). Research could focus on coming up with better ways of knowing the location of data that needs to be deleted. Data lifecycle should be tracked around the cloud in order to help providers make better decisions on which methods of assured deletion to apply.

Reardon et al. (2013), presented a survey which focused on analysing and discussing methods of secure deletion. It is stated that in order to know which secure deletion approach to use, one has to know the properties of the environment before deciding the most suitable method. Regarding the cloud, it would be ideal to track and locate data through its life cycle, that is, getting all the relevant information needed for deletion to help the provider with a better method of assuring deletion.

Deletion of metadata is another area that could be explored. To completely delete data also means getting rid of the metadata associated with data that is being deleted. These metadata do sometimes hold sensitive information which may require complete deletion. Solutions could include finding ways of deleting metadata from all places in the cloud. Diesburg et al. (2016) proposed how metadata could be deleted from NAND flash drives and hard drives.

Deletion of data could also be performed at every layer of the cloud to assure that data is completely removed. Secure deletion methods could be refined to work in these layers to guarantee deletion. Again, using provenance, data could be tracked up to the physical layer in order to capture the mapping between virtual and physical resources. This would also give cloud users some sense of transparency and show them how virtual locations and physical static server locations are linked or mapped together. Furthermore, this mapping will allow deletion done on the virtual environment to be confirmed at the physical level hence enabling proof of deletion.

In a complex system like the cloud, researchers could develop security policies for each component in the cloud rather than having policies tied only to data. As previously mentioned, data in the cloud may get trapped in different cloud components and never be assuredly deleted. New solutions can focus on developing a set of policies that could be implemented by each component to make deletion easy and more secure.

Deletion of all backup copies. To assure availability and durability in the cloud, data is replicated in the cloud several times. During deletion, it can be a challenge to guarantee that all copies are deleted. Also, data can get trapped in different cloud layers and components and be left behind during deletion. Research could focus on tracking this data and keeping a record of all the places it has been so that during deletion, such data could also be removed. Proof of Retrievability (PoR) and Provable Data Possession (PDP) are areas that could be considered to verify the number of copies in the cloud. Traditional PoRs (Bowers et al. 2009b, Juels and Kaliski Jr 2007) only demonstrate and verify that the provider has the data, but not how many copies exist in its possession. Benson et al. (2011) looked into verifying that data is present in multiple locations (e.g., multiple disks) within the cloud. Their framework complemented Bowers et al. (2011) who looked at the same verification but on a single geolocation or datacentre. Both of these approaches could prove to be vital in providing the location of data, that is, location within the datacentre and location of data concerning geographical locations. They can also be used to verify the number of copies before deletion takes place to help the provider in knowing how many copies are there to delete and verifying that the same number of copies is deleted after a deletion request. Watson et al. (2012)

also explored the possibilities of verifying that the provider is storing files at the requested location which can also be leveraged to prove deletion.

Another approach that could help to verify deletion of multiple copies of data previously held by a provider is proposed by Barsoum and Hasan (2010). This work is aimed at verifying that the provider has more than one copy of outsourced data in its possession. They extended existing PDP (Ateniese et al. 2007) that focused on a single copy of data. Under assured deletion, this work can also be used to verify whether the provider still possesses other copies after deletion.

To ensure deletion of data left behind after migration, research could focus on capturing all the necessary information involved during migration. This may include the amount of data that moved and the data that failed to move during migration. This could also give the provider an opportunity to delete all data in case of errors and actually confirm it.

Deletion granularity. Multi-tenancy is an important feature of the cloud – multiple users can share resources (e.g., storage, tables) hence reducing the cost for the provider. However, sharing of resources can lead to data being tangled and mixed up hence proving to be a challenge during deletion. As stated before, applying secure deletion mechanisms may lead to service disruption for other tenants using the cloud in case the secure deletion method deletes other tenants' data. To prevent service disruption, assured deletion granularity is needed. Completely deleting tenants' data that need to be deleted while other data remains untouched. Cachin et al. (2013) constructed a deletion scheme that aims to provide deletion granularity. The scheme maintains a collection of files and provides deletion of the files that need to be deleted. Research could focus on providing this promptly.

Not all data stored in the cloud is sensitive therefore data provenance could be explored in depth to track sensitive data and only allow data with the same level of deletion requirements or retention duration to be stored in same locations. These might be media which allow secure deletion mechanisms.

Deletion latency. Time to guarantee deletion can depend on many things, for instance, the distance between data centres or the number of copies that need to be deleted. Research could focus on these properties to develop and design better mechanisms which could assure deletion within a reasonable time. Assured deletion procedures could be timed, and the elapsed time could be used as part of the evidence in proving deletion. Research could also focus on finding out how secure deletion methods could be improved to complete promptly in cloud settings.

Error handling. Research could look into improving the current secure deletion methods which can recover during failures and still completely delete data. For instance, Diesburg et al. (2012) developed an assured deletion framework for the storage data path that could handle crashes. Such works could be extended to the cloud and incorporated with the cloud system to report any deletion crashes. Deletion mechanisms in the cloud could be improved to report failures.

Proof of deletion. Methods are needed to attest whether deletion mechanisms complete deletion successfully without errors. These could provide building blocks to proving deletion. Deletion attestations could either be done at software level or hardware level depending on deletion requirements. Backes and Bendun (2015) proposed a verification mechanism which a cloud user could use to force a cloud provider to attest deletion of data.

To conclude the discussion on the requirements and the areas to explore, Table 3.5 provides a summary of the assured deletion requirements and the research areas that may be explored to satisfy them.

Table 3.5 A mapping of assured deletion requirements and research areas that could be explored to satisfy them.

Deletion requirements	**Summary of possible areas to be explored**
Service availability	Tenants Isolation Controlled movement of data in the cloud infrastructure
Complete deletion	Use of secure deletion methods which are suitable and appropriate for the cloud. Secure deletion of metadata Deletion of data at every cloud layer Secure deletion policies for each component
Deletion of all backup copies	Keeping track of number of copies Verifying the number of copies before and after deletion Data provenance during migration
Deletion granularity	Data provenance for sensitive data. Isolation of data that needs to be securely deleted. Use of secure deletion methods which are suitable and appropriate for the cloud.
Deletion latency	Exploration of secure deletion mechanisms that can delete data promptly.
Error handling	Resilient mechanisms for secure deletion in the cloud.
Proof of Deletion	Verification and attestation of deletion in the cloud.

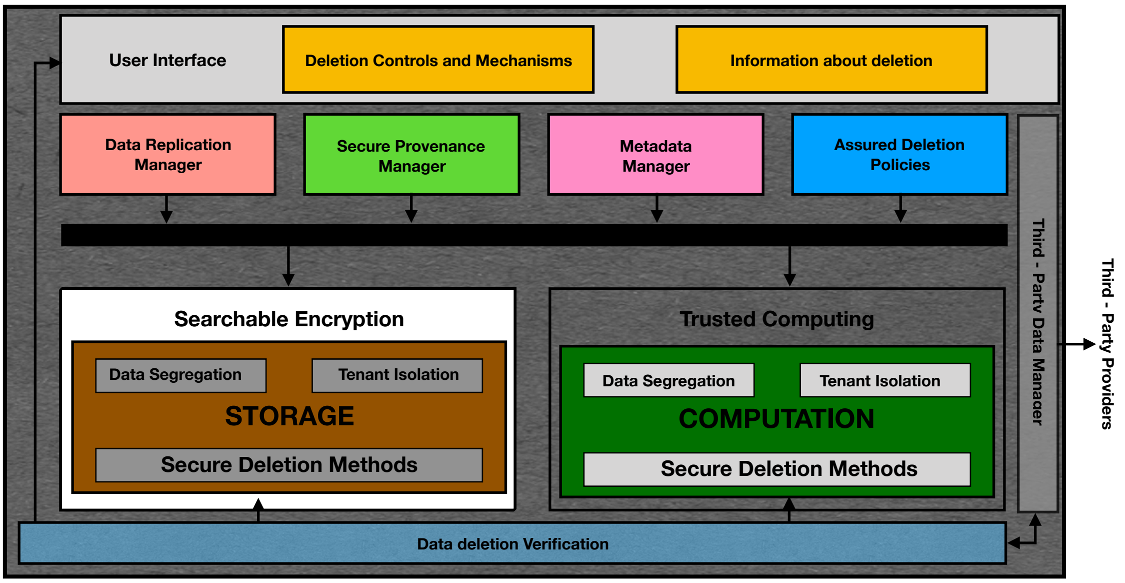


Fig. 3.3 Assured Deletion Conceptual Architecture

3.3 Assured Deletion Conceptual Architecture

To summarise the technical investigations concerning assured deletion in the cloud, Fig. 3.3 presents a conceptual architecture for assured deletion in the cloud. This architecture is based on the results of the existing literature on assured cloud deletion and the cloud infrastructure analysis conducted as part of this thesis. Various assured deletion concepts from the existing literature were identified and similar concepts were grouped. A comparison between groups was made to find out how each group related to one another. After identifying these relationships, a conceptual architecture was produced. It provides an understanding of the areas that needs attention regarding assured deletion in the cloud, showing how assured deletion components and concepts are related to one another. Overall, this architecture provides a point of reference for solutions towards assured deletion, and it also serves as a communication tool for further questions and possible directions concerning assured deletion. The conceptual architecture consists of the following main components:

- **Data replication manager**
  This is a component that is responsible for managing data replication within the cloud – keeping track of how many copies are created and need to be deleted when the deletion request is made. It will also ensure that sensitive data is only replicated according to tenants' requirements and data policies. During deletion, the replication manager will also ensure that all the other copies are considered.

- **Secure provenance manager**
  It is responsible for the tracking of all sensitive data in the cloud throughout its life cycle. This ensures that the location of all sensitive data is tracked throughout the cloud system enabling the provider to locate and delete data when required. This will ensure that data trapped between the cloud layers is tracked and deleted when data deletion is requested.
- **Metadata manager**
  This component is responsible for the management of information about data, together with the provenance and replication manager, it ensures that all metadata is tracked and deleted along with the data when deletion is requested.
- **Assured deletion policy module**
  This module ensures that all data deletion policies are enforced in the cloud system. It will ensure that replication and provenance policies are considered during deletion. It will also guarantee that user deletion requirements (e.g., type of deletion) are fulfilled.
- **Third-party data manager**
  This is a component that is responsible for the management of third-party activities in the cloud, ensuring that third-party services respect deletion policies (enforced by the policy module). For example, making sure that data marked sensitive does not leave the cloud system where secure deletion cannot be guaranteed.
- **Deletion verifier**
  This is a trusted module that is responsible for verifying deletion in the cloud. This component works with other modules to verify that data has been deleted. To confirm data has been deleted, a tenant poses a challenge which the cloud is expected to solve. This component will respond to the tenant's challenge with a signature which can be verified by the user as proof of deletion. The interface will also provide adequate feedback to users.
- **Trusted computing module**
  This is responsible for executing and processing sensitive data with strict requirements. This will allow the use of secure deletion methods to ensure that sensitive data is securely deleted when the request is made. Tenants will be isolated to prevent data tangling together.
- **Encryption module**
  This module will ensure that encrypted data can still be processed in its encrypted state. As homomorphic encryption schemes develop and become more usable, this module will ensure that more encrypted data is processed within the cloud.
- **User interface** This is the component that is responsible for connecting the user and the cloud system. It contains all the necessary controls to allow users to input their deletion requirements and desires. The cloud provider shares information and gives feedback through the interface.

Significant efforts have considered some areas of this architecture such as cloud provenance (Lu et al. 2010, Zhang et al. 2011a, Abbadi et al. 2011, Pasquier et al. 2016), tenant isolation (Ristenpart et al. 2009, Zhang et al. 2011b, Benson et al. 2011, Zhang et al. 2012), trusted computing platform (Krautheim 2009, Krutz and Vines 2010, Shen et al. 2010) and encryption (Bellare et al. 2007, Geambasu et al. 2009, Reardon et al. 2014, Mo et al. 2014b).

3.4 Summary

This chapter has focused on understanding deletion from the cloud providers' side, it has investigated the cloud infrastructure and highlighted different requirements that the cloud providers need to consider in order to assure deletion. Moreover, it has revealed how salient features of the cloud pose various challenges to assured deletion. With the help of using two threat models – honest and dishonest provider, a conceptual model is proposed. Achieving this architecture would ensure that data stored and processed in the cloud is accounted for when deletion is requested. One aspect of this model that has not been considered to date is the user interface (usability of data deletion in the cloud), despite the interface being what users use to interact with the cloud. Currently, there is no evidence on whether the current cloud deletion mechanisms meet the deletion needs of users, and also user requirements for assured deletion are missing.

The remainder of the work presented in this thesis aims to investigate assured deletion with regards to users. As previously shown in Chapter 2, there is a gap in the cloud literature on the issue of usability of deletion in the cloud though computers users are the main users of the cloud. The next chapter, Chapter 4, focuses on understanding cloud users' deletion practices while Chapter 5 studies deletion preferences and information about deletion. Chapter 6 uses the findings of these chapter to develop design principles for assured deletion.

Chapter 4

Users' Cloud Deletion Practices and Coping Strategies

The content of this chapter is based on the following publication:

> Ramokapane, Kopo M., Awais Rashid, and Jose M. Such. "I feel stupid I can't delete...": A Study of Users' Cloud Deletion Practices and Coping Strategies. *Proceedings of the Thirteenth Symposium on Usable Privacy and Security*, 2017.

This part of the thesis aims to examine cloud users' deletion practices, it explores several questions: (1) what are the key motivating factors that underpin cloud users' need to delete? (2) Do they find current cloud deletion mechanisms usable and, if not, what are the factors underpinning users' failure to delete? (3) What are the coping strategies that users deploy to work around these challenges, and (4) What do users want concerning usable deletion in the cloud? The key contributions of this chapter are as follows:

- It identifies four key drivers that motivate users to delete from the cloud.
- It reveals why users fail to delete from the cloud, highlighting what causes and contributes to such failures.
- It uncovers different coping strategies adopted by cloud users to address their deletion challenges, discussing the consequences of such strategies with regards to their motivations to delete.
- It demonstrates that some of the challenges that are faced by users are due to their mental models.
- It reveals what cloud users want regarding deletion from the cloud.

As previously explained in Chapter 2 and 3, deletion of cloud data is a concern for most organisations and users and research has suggested encryption-based solutions to alleviate this problem. However, as previously mentioned these approaches assume that users have sufficient understanding of data management in the cloud — they have explicit mental models of how data is deleted from the cloud. It is also assumed that users can accomplish the task of deletion through either the features offered by cloud providers or using more sophisticated assured deletion mechanisms such as encryption-based solutions. Prior studies have not explored the usability of data deletion in the cloud has not been explored, and users' understanding and challenges of deleting from the cloud are still to be studied. For instance, at present, cloud users are allowed to access and delete from the cloud through mobile apps, web interfaces and from their computers. Nonetheless, deleting from these platforms requires different mental models, and this can be a challenge for users as most of them assume these platforms work the same way and expect the same results. Through an exploratory study using semi-structured interviews (Bryman 2015), this work explores what motivates users to delete from the cloud, their success/failures, and their coping strategies with regards to deletion in the cloud.

This chapter is structured as follows. Section 4.1 describes the methodological approach and the demographics of the participants. The findings of this chapter, current user deletion practices, challenges, coping strategies, and what users want are presented in Section 4.2. Section 4.3 discusses the relationship between users' motivations to delete, their coping strategies and mental models.

4.1 Methodology

A qualitative inquiry (i.e., 26 semi-structured interviews) was conducted into how cloud users understand and think about deletion in the cloud. These interviews were carried out between November and December 2016.

4.1.1 Ethical considerations

Lancaster University Ethics Committee approved this study before any research activities began. Each interview session began with obtaining written consent from the participant to take part and record the interview sessions. The information sheet and consent form can be found in Appendix A.

4.1.2 Participants/Sampling methodology

Participants were recruited through Lancaster University's existing professional networks, word of mouth and posters in Lancaster University and city centre. Interested participants

Table 4.1 Summary: Interview Demographics.

	No. of participants
Gender	
Male	12
Female	14
Age	
18 - 20	3
21 - 25	8
26 - 30	6
31 - 40	7
45 +	2
Educational Background	
High school/College course	5
Bachelors	9
Masters	5
PhD	6
Preferred not to say	1
Employment status	
Unemployed/Retired	1
Full time	12
Part-time	3
Student	10
Cloud services	
Dropbox	15
iCloud	9
OneDrive	6
Google Drive	17
Box	17
Microsoft Office 365	15
Google Docs	14
OneNote	2
Cloud Access	
Smartphone	26
Tablet	13
Desktop	23
Laptop	25

were invited to complete an online form which contained a set of questions designed to screen participants (UMASS 2015 (accessed June 1, 2019)) that could be invited for one-to-one interviews.

Other than for balancing demographics, the screener also focused on asking participants about the cloud services they were currently using and the devices they used to connect to such services. Respondents were also asked about their activities in the cloud, that is, whether they have ever deleted, shared or uploaded and downloaded data from the cloud.

Throughout three weeks, a total of 48 responses were received. From these, 16 were males, 28 were students (3 doing Master's degrees while 12 were pursuing a PhD), 18 had some form of employment and two were unemployed or retired. The majority (30) were between the ages of 21 and 30 years old. Table A.1 in Appendix A shows the demographics of all the users who responded to the adverts.

All 48 stated that they used smartphones to connect to the cloud while 43 also used their laptops to connect to cloud services. 97% of the respondents stated having shared, uploaded and deleted data from their cloud services. The screener questionnaire aimed to identify a group of about 20 to 25 participants for one-on-one interviews, a number that was enough to reach a saturation point concerning the emerging codes (i.e., the point in a qualitative research when there is enough data to ensure the research questions can be answered (Bowen 2008)). For maximum variation, this study purposefully aimed at respondents from a wide variety of backgrounds, ages, education and socio-economic classes. More importantly, this study targeted respondents who use more than one cloud service, preferably a storage service (e.g., Dropbox) and data processing service (e.g., Google Docs). Preference was also given to those participants who had mentioned that they had uploaded, shared and deleted data from these services. Interviewing respondents from a wide range of backgrounds was an excellent opportunity to capture different perceptions of deletion in the cloud and identify common patterns.

In the end, 26 out of the 48 respondents to the preliminary screener were invited by email to participate in the interview. The sample included 14 women and 12 men, between 18 and 50 years old. For a 30 to 45 minutes interview, participants were compensated £7.87 for their time and effort. Table 4.1 summarises the demographics information of the participants.

4.1.3 Interviews

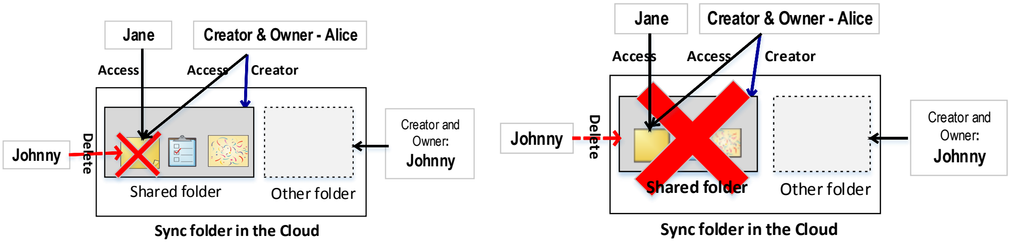
Interviews were conducted in different places to meet participants' needs and requirements (e.g., at a participant workplace because of their work schedule). The vast majority of interviews (25) were conducted in-person, while a single one was conducted via Skype because the participant was in another town.

At the beginning of each interview session, participants were given a participant sheet which explained what the study was all about and consent forms to take part in the study. To guide each interview session, a semi-structured interview protocol was adopted. This protocol gave the researcher the flexibility to make up follow up questions (and probe participants for more information) during the session without any restrictions (Portigal 2013). In addition to semi-structured approach, a reflective questioning technique (Roulston 2010) was also adopted. The reflective technique allowed participants to reflect their actions and decisions aloud hence not directing them to a conclusion. The reflective technique also gave the researcher the opportunity to explore a participant's knowledge, skills, experiences, attitudes, beliefs, and values.

The first set of questions focused on the general use of the cloud, the researcher asked participants how they use cloud services on a day-to-day basis and their reasons behind using and choosing a particular service. The second set of questions set to find out what users deleted, and the reasons behind deleting. These questions also included questions on the problems they encountered while attempting to delete from the cloud. Some of the questions included scenarios (*see below*) that required participants to think and make decisions on how to delete data. All interviews were audio recorded using a secure audio recorder and stored securely. The audio recordings and the transcriptions were not accessible to anyone other than the researcher and the transcribers who had to fill the Non-Disclosure Agreement(see Appendix A.6).

Scenarios

As part of understanding users' perception of deletion, two scenarios were used, and participants were asked to describe what would happen when deleting a file under each scenario. By doing this, the researcher gave the interviewees the opportunity to apply their understanding of the cloud and deletion. These scenarios represent typical deletion tasks associated with cloud storage services. Each scenario was contextualised based on the information the interviewee provided and then narrated to the participant. For example, if the interviewee mentioned that they regularly used Dropbox to share photos, then the scenario would involve Dropbox and sharing of photos. The reasons behind selecting these scenarios were to use an example that users were familiar with or once experienced when using the cloud.



(a) Deleting a shared file

(b) Deleting a shared folder

Fig. 4.1 Deletion scenarios: (a) We asked a user "Johnny" what would happen if he deleted a file shared between him and his friends "Jane and Alice" created by "Alice", and (b) we asked what would happen if Johnny attempted to delete a shared folder instead of the file.

Scenario 1: Deleting from a shared folder. This scenario (shown in Fig. 4.1a) asked participants what would happen if they deleted a file from a shared folder created by their colleague or friend. In this scenario, Alice has created and shared a folder with both Jane and Johnny. Johnny is running out of space but decides that the only file he can delete is the one in this shared folder because he has finished using it. However, without first contacting Jane and Alice, Johnny has to decide whether to delete the file or not. Would Johnny be able to delete this file? If he deletes this file, what would happen?

Ground truths - Deleting from a shared folders

The time when this study took place, the following were possible.
Dropbox: *All the members of the shared folder can delete any item from the folder regardless of the author or creator. The item will be removed from all the members' individual accounts.*
Google-Drive: *Item creator or author can specify which members of the shared folder can delete. The deleted item is removed from all the other members' individual accounts.*
Box: *By default, all the members of the shared folder can delete any item from the folder. However, all the other members of the folder will receive a notification alerting them which item has been deleted and by who. The item will be removed from all the member's individual accounts to the deleted item's folder.*
OneDrive: *Users given the permission to edit can delete items from the shared folder. The deleted item is then removed from all the shared folder members to the owner or creator of the folder.*

Scenario 2: Deleting a shared folder. The second scenario (shown in Fig. 4.1b) asked users what would happen if they tried to delete a shared folder created by their colleague or friend. In this scenario, Alice has created a project folder and shared it with both Jane and Johnny. After the project has been completed, Johnny thinks he has no use for all the files in the shared folder. Johnny goes to his laptop and deletes the shared folder from his sync folder. Would Johnny be able to delete the folder? If he deletes it, what would happen? Will Jane and Alice still have access to the files in the shared folder or the folder will disappear from both of their accounts? Alternatively, will it only disappear from Johnny's sync folder?

Ground truth - Deleting a shared folder

The time when this study took place, the following were possible.
Dropbox: *Members who delete the shared folder do not delete it from everyone else account, they are just leaving the folder. The folder will remain in other members' individual accounts.*
Google-Drive: *The creator or owner of the folder is the only one who can delete the shared folder. Other members of the folder can only request to be removed from the folder.*
Box: *Members of the shared folder cannot delete the shared folder; the owners are the only ones who can delete a shared folder. When an owner deletes a shared folder, it is removed from all other members' account.*
OneDrive: *Members can delete a shared folder from their accounts, but this deletion does not affect other members of the shared folder. The owner of the folder is the only one who can delete the shared folder and it is removed from all the other members' accounts.*

4.1.4 Grounded Theory Analysis

To explore users' deletion practices, it was essential to adopt a methodology that is suitable for generating a theory or understanding an area that has not been explored deeply, rather than testing a theory. Consequently, the Grounded theory method was selected as it met these requirements – user cloud deletion practices have not been studied before. The grounded theory analysis originated from the work of Glaser and Strauss (Glaser et al. 1967, Glaser and Strauss 2009). This study follows *Straussian grounded theory* analysis (Stol et al. 2016).

Transcription and coding began immediately after the first five interviews. It was essential to start coding immediately because identifying interesting categories and themes early on gives the researcher the opportunity to explore some areas further. Sometimes, it can also help to identify which questions need to be changed, removed or explored further (Charmaz 2006; 2014, Glaser and Strauss 2009) during subsequent interviews.

Interview scripts were analysed through several iterative stages of open, axial and selective coding and constant comparisons of codes (Ryan and Bernard 2003, Glaser and Strauss 2009, Charmaz 2014). Using NVIVO[1], the researcher, went through each transcript line-by-line and developed the first descriptive open codes. Several codes about how cloud users use the cloud and how they delete began to emerge. The first coding process resulted in 120 unique codes. To verify the codes, after the main coder had coded the first two scripts, the second researcher independently coded two other scripts resulting in a subset of codes from the primary coder and the codebook was modified accordingly (MacQueen et al. 1998). The

[1] http://www.qsrinternational.com/what-is-nvivo

second coder later randomly selected three more transcripts to code and then the inter-coder reliability agreement between the two researchers was calculated.

The second phase of coding involved identifying patterns, connections, and relationships between the codes identified by the earlier coding process. To complete this, the researcher grouped similar or related codes to form categories (or concepts) and in some cases expand the codes themselves to make categories. As different groups continued to emerge, the researcher began to compare the groups against each other looking for more further connections between them. The interview process continued alongside the coding process. Coding and memo writing until no more new codes were emerging. New codes stopped emerging after 13 interviews scripts were analysed – regarding grounded theory analysis this was a saturation point. In this phase, memos were used to describe codes, events, behaviours, and record emerging questions during the study. Following this, the categories and themes were ordered and further grouped into more broad and abstract groupings.

In the last round of data collection, few questions were added to the interview script based on the groupings and the questions that emerged from memos made during analysis. For example, in the second phase of data analysis, two participants mentioned they would want their photos to be deleted entirely, so the subsequent participants were asked to state what data they would want to see completely deleted from the cloud.

The last phase of coding involved selective coding, where further transcripts were analysed in an attempt to identify a linking core category that describes the underlying phenomenon in the observed and interpreted behaviour. This iteration gave the researcher the chance to engage more with the study, understanding what the users were saying and doing with regards to deletion.

4.1.5 Validity of grounded theory

Ground theory as a methodology has some limitations, and these may include methodological errors, limited generalizability or lack of repeatability (Hussein et al. 2014, Ahmed and Haag 2016). Researchers using grounded theory may sometimes restrict themselves to purposeful sampling rather theoretical sampling and result in having a lack of conceptual depth (Charmaz 2014). To help circumvent this, coding continued even after the theoretical saturation was reached. The issue of generalizability is less discussed in qualitative research because the aim of qualitative method is to provide a rich contextualised understanding of a phenomenon (L Berg 2001). Generalizability in qualitative research is not only attained by confirmatory studies but can also be attained through systematic replication, which leads to confirmatory evidence (Polit and Beck 2010). To increase reliability and account for coding biases, a second researcher was invited to code other transcripts. Then both

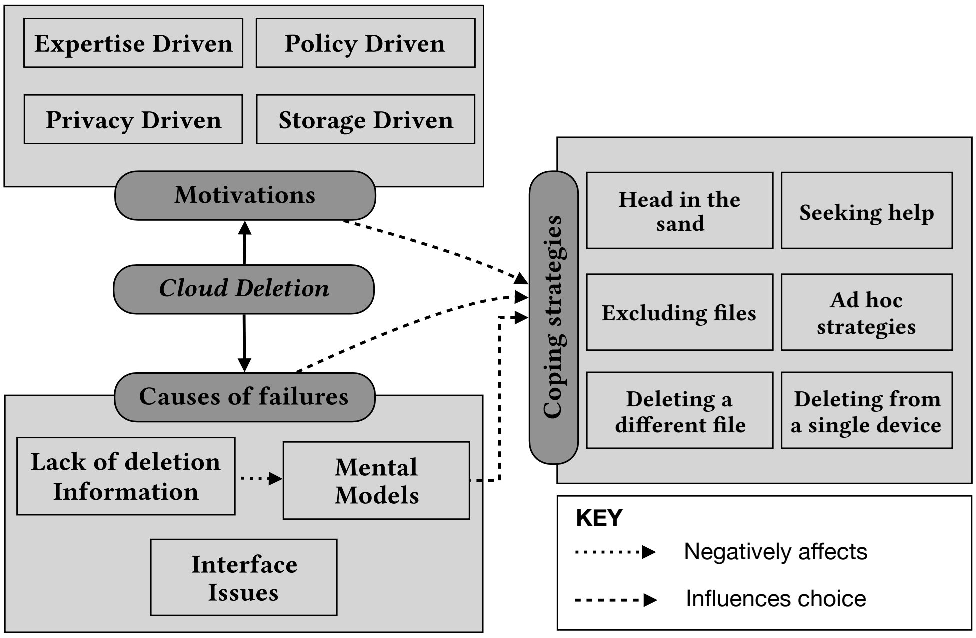


Fig. 4.2 Summary of the key findings

researchers discussed codes and finalised the codebook. The inter-coder agreement using Cohen's kappa coefficient was *0.86*. A value above *0.75* is considered a good level of coding agreement (Fleiss et al. 2013, Neuendorf 2016).

4.2 Key Findings

Fig. 4.2 presents an overview of the key findings which are summarised below.

(1) What makes them delete?

Users' motivation to delete falls into four major categories: privacy-driven, policy-driven, expertise-driven and storage-driven.

(2) What causes deletion failures?

Not everyone can delete when they want to. Poorly designed user interfaces do not merely cause failure to delete in the cloud but rather this can be attributed to a lot of other factors which may include incomplete understanding of the cloud deletion and lack of sufficient information on deletion.

(3) How do users cope with deletion failures?

Users develop and choose a coping strategy based on their motivation to delete or the cause of their failure to delete. For example, users whose intention to delete is privacy-driven favour strategies that will remove the file from the cloud or will stop uploading the files they perceive to be essential or confidential. Whereas, users whose reason to delete is to gain more storage space, will not mind cloud hopping to gain additional storage.

(4) What do users want?

Users desire four key characteristics with respect to deletion in the cloud: transparency in deletion, deletion to be complete, control over deletion, and help service to support deletion tasks.

The next section details each of these findings.

4.2.1 What motivates users to delete?

Before understanding why cloud users could not delete, it was important first to understand what motivates them to delete or the situations in which users want to delete. Users' motivations to delete were: privacy-, expertise-, policy- or storage-driven.

Privacy-driven Motivations.

Users' concerns about online information, the level of trust the user has towards the provider, or the perceived negative consequences of not deleting a file from the cloud are often driving factors for deletion.

Lack of trust in provider. Users with high privacy concerns towards cloud providers usually desire to delete. Participants revealed that they delete because they fear that their data may fall into the wrong hands.

For instance, P4 said,

> *''It's just me being cautious that this, from my understanding and it's not that good, Dropbox is something like an online database like or a storage space, so I prefer just to be on the safe side to delete everything so that nobody else can access to these files apart from me."*

P4 later on continued,

> *''...I don't want anything bad to happen or Dropbox being hacked...I have interviews with children, so I send recordings of interviews with children so just to be on the safe side once she receives everything, I delete everything."*

To avoid future conflicts. Participants deleted to avoid future conflicts which may be unearthed by the data they have in the cloud.

> *''...I no longer want the pictures of my lover to be accessible to anybody else I'd want them gone from the servers forever because my next future [partner] might discover them ... " P11*

> *‘‘I get rid of things that could come back to haunt me. . . ” P10*

To forget. Users deleting in order to forget. Participants who automatically save their photos in the cloud revealed to us that they commonly delete photos from the cloud in order to forget about them, e.g., photos which are perceived to be embarrassing or contain an unpleasant memory.

> *‘‘. . . I am always deleting pictures and stuff that I don’t need to remember." P20*

> *‘‘Some unhappy memories maybe, I would want them [to] disappear forever, and I don’t want to see them again.” P17*

Expertise-driven Motivations.

These are factors that are motivated by the level of understanding a user has or their ability to delete successfully.

Self-efficacy. Users’ desire to delete is heavily influenced by their confidence in their ability to complete the deletion task successfully. Participants who had enough knowledge or skill to delete tended to decide to delete whenever they wanted and execute it immediately. However, those who struggled to delete showed less interest in wanting to do so.

Deleting after unintentional use. As highlighted in other cloud studies (Clark et al. 2015), first coding resulted in the category “unintentional use”– using the cloud unintentionally. Several participants (5/27) deleted from the cloud because they first used it without knowing. At the time of this study, it was common to find cloud-based applications already installed on smart devices and computers. However, at first, most users are not aware that some of these cloud applications will automatically sign them into these services and start saving data to the cloud. Upon realisation, most users’ response was to delete the data as soon as possible. Participants *rushed* to delete because they were not sure how their data got there in the first place and, because they were not sure whether their data was public and, hence, visible to everyone. One participant who did not realise she was using OneDrive for two years noted:

> *‘‘When I first found out I had it I tried to delete all the photos because I got scared... and I managed to. . . , I deleted the files, and I got really confused when I first opened it because I was like well how did these files end up here? I never put them there but obviously, I’d whacked them in my phone, and that’s where it’d automatically saved to. So, I deleted them." P18*

Policy-driven Motivations.

Users are driven to delete due to *extrinsic* policies, e.g., organisational security policies to which they must adhere, as well as *intrinsic* ones, e.g., the perceived value of information held in a file.

Organizational policy compliance. Compliance with organisational security and information sharing policies is often a driving factor for deletion. Some participants mentioned that their work policies required them to manage data securely which included deletion. However, interviewees also revealed that they continued to use public clouds – despite this being in violation of organisational policies – because they were convenient and easy to use.

> *''... I use [Box] because it was recommended and we were told that we couldn't store research work on Dropbox ... Sometimes I use Dropbox ... most people use it, but if I use Dropbox I delete..." P9*

Perceived value. Users' decision to delete was also influenced by the usefulness, sensitivity and value of a file. When users perceived a file to be confidential or more sensitive or valuable, they would want to delete it immediately after using it. Also, if users considered a file to be no longer useful or needed, they would consider it a good candidate for deletion.

> *''... It's just normal deleting. I read the paper, and if it's no use for me, then I delete it, that's it... There was a document we had, it was not good for our job, we just deleted [it]." P8*

Storage-driven Motivations.

Users are also driven by storage size and the need to organise their data systematically.

Storage size. The most prominent repertoire was deleting in order to free some space. Instead of buying more storage cloud users use deletion to reclaim used space. The less space a user has, the more motivated they will be to delete.

> *''I didn't have a lot of space, so I had to take out some pictures and videos, so I deleted them from iCloud." P16*

> *''[When my] space was limited I would actually go through and prune out what's important what's not." P11*

Tidying up. Sometimes cloud users delete to keep things tidy. When users have the skill and the knowledge to delete, they will sometimes take time out to tidy things up from their cloud accounts.

4.2.2 Why do users fail to delete?

While the exact details of users' failure to delete varied, the results of this study suggests a range of common factors that lead to deletion failure. These factors include insufficient deletion information, incorrect understanding of cloud deletion, and user interface designs.

Insufficient information

Insufficient information on deletion contributed to a lack of understanding of the deletion process. Though participants were not asked questions about information on deletion, participants stated that such information is hardly available. It also emerged that service provider advertisements had information on the benefits of using the cloud (e.g., storage size) but none on deletion.

> *''Nothing like that is made very clear when you sign up. Maybe if you read through the gazillion terms and conditions, you would find out, but there's nothing obvious that I've come across anyway that says, 'After this amount of period of time this will actually be deleted.' So, surely that should be one of the first things that they tell people." P19*

Users sometimes find deletion messages challenging to understand and not providing them with relevant information that is pertinent to the deletion task.

> *''Sometimes on iCloud it does not allow you to delete like if you are trying to delete something, it says that if you delete it will mess up everything else, but on Google Drive and on Dropbox, I've never found anything like that. On iCloud sometimes it does not allow you to delete some stuff." P24*

While some information on deletion is sometimes made available in the terms and conditions of the cloud services, interviews showed that users do not read the terms and conditions. Therefore, they do not have a full understanding of how the provider performs data deletion. Concepts such as retention period are not well-understood among users. Some of the participants (7/27) noted that, while they did not have problems with deleting, there was insufficient information on whether their deletion is permanent or not.

Mental Models

Users have been known to come up with ideas when they are insufficiently educated on an issue but have to complete a task. However, most of these models are inaccurate and, in most cases, lead to wrong results (Wash 2010). Unsurprisingly, most cloud users with incomplete

Table 4.2 Common beliefs of cloud deletion

Popular Models	Uncommon Models
- Sync folders are not the cloud - Saving and deleting work the same way - The cloud within an app - Borrowed deletion models - Shared folder: Deletion is one sided	- Shared folder: Deletion affects all members - Providers don't delete

mental models of data management in the cloud reported not being able to delete. Users' incomplete mental models did not just lead to deletion failure but also to wrong expectations.

Data collected during this research showed that most cloud users have less or minimal understanding of the cloud or deletion. For example, some participants(7/27) presented limited knowledge on how the folder synchronisation works or that they can access the cloud independent of apps that consume or used the cloud as a storage. Regarding deletion, they did not understand the concept of *Deleted folder* (i.e., where deleted data goes), retention period, and different levels of deletion. Using these themes, several beliefs that existed among the participants concerning deletion and the cloud were identified. Table 4.2 shows which beliefs were shared among participants and which ones were not.

Sync folders are not the cloud.

To automatically save data in the cloud, users have to install synchronisation software or applications in their devices. During installation, a synchronisation application will create a local folder in the user's computer which will be linked with the cloud service. The purpose of the synchronisation application is to detect all the changes (e.g., adding or removing files) in the folder and update user's contents in the cloud. The sync program or application ensures that user's files in the cloud and in the local device are up to date. For this to work, a user is required to be signed into their cloud account through the application. While most users use sync folders, the interviews suggest that most of them do not understand how this sync folder works. Some participants (3/27) deleted from the sync folder while their computers were offline but expected the files to be deleted instantly from the cloud. Also, another group did not understand that deleting from this folder would also mean deletion from the cloud.

> *''So once I put [my] files in Box sync folder, and it uploaded automatically but I [was] wondering if deleting my files in that folder would also delete from the cloud as well... but then I just deleted those files from my folder and then logged into my box [account] and found [that the] files in the cloud were deleted as well. I was not happy to see that." P24*

Saving and deleting work the same way.

Some users (9/27) wrongly concluded that deletion in the cloud worked the same way as saving a file. Though this is correct to some extent, this is true when using a sync folder in a computer. When using a sync folder, every file placed in that folder will automatically be uploaded and saved in the cloud. When a file is deleted from a sync folder, it will be removed from the user space visible to the user but placed in a *Deleted folder* in the cloud. Depending on the cloud service being used, the local files may be moved to the *Recycle bin* or *Trash can*. However, this may not apply in a situation where a user connects to the cloud through another app (e.g., camera app to backup photos). Deleting a photo from the camera app may not necessarily delete it from the cloud. However, this study found that some participants (4/27) expected files to be deleted from the cloud when they deleted them through other apps since they automatically get saved to the cloud using some apps.

One frustrated participant said,

> *''I used to think that it[deletion] was kind of automatic that if I deleted from the phone that it would like [delete] because the fact that I save the photos, that it's saved in iCloud I think that if I delete it, it will delete itself from my iCloud as well." P14*

The cloud within an app.

The use of applications to consume cloud services has left many users not knowing that they can access the cloud independent from these applications. They believed that their cloud storage was part of the applications they were using. This group of users does not delete from the cloud because they do not know that their data may still be in the cloud and they can access the cloud to delete. Users also do not know that some applications may backup data automatically to the cloud.

One participant was surprised when asked if she had ever directly logged into their iCloud and deleted some files:

> *''I [have] never come to the conclusion that I could actually go there and delete" P14*

Borrowed deletion models.

Some users (4/27) transferred their beliefs of deletion from other online services such as online social networks to the cloud. When asked about how they would delete a file from a shared folder in the cloud, one Dropbox user responded saying:

> *''. . . I think I will ask my friend to delete it. And if they don't then I can't do anything apart from untagging myself. I think it's quite a similar policy like in Instagram or Facebook when you want to delete it, it always gives you the options either contact your friend to have them delete the photo, or you just unfriend them." P14*

Shared folder: Deletion is one sided.

The concept of deleting from a shared folder is a challenging and confusing one for most users. Some participants (3/27) believed that when one deletes from a shared folder, the deletion will only remove a file from their accounts but that particular file would still be available for other members of the shared folder. To these users, deletion from a shared folder is one-sided. They believed that when a file is uploaded to a shared folder in the cloud, the cloud will create multiple copies of the file for all the members of the shared folder. Thus, assuming that when they delete from a shared folder, they are only deleting their copy and not deleting from everyone. One participant likened deleting from a shared folder to deleting from Whatsapp messenger:

> *"... [If she deletes it] it wouldn't get deleted from my side but it will obviously just get deleted from her side [because] she doesn't want it. Like if you send a WhatsApp message. So normally it wouldn't get deleted from my end, it would just get deleted from her end." P12*

Shared folder: Deletion affects all members.

Some participants (15/27) reported that they knew that deleting from a shared folder may remove the file from all other members' accounts as well. Nevertheless, this caused a conflict within the user who no longer needs a shared file within a shared folder and wishes to delete it. Users find it difficult to decide whether to delete or not. They believed that there was no alternative way of deleting a file from a shared folder while other members of the folder are still in need of that file. Users who possess this model prefer to leave the file in the shared folder undeleted to be on the safe side.

> *"... when I do this it's always after my transcriber has used the material and sent me the transcriptions back, so I always think it is safe to delete them now because in my head I'm thinking... if I delete them she won't be able to see them, so I wait for her to finish the job, and then I delete them." P4*

Deletion is permanent and instant. Some participants (5/27) had a very under-developed model of deletion in the cloud. These subjects did not think about deletion in any depth other than that it was instant and permanent. The fact that a file disappears from their sight was enough for them to conclude so. These users were unconcerned about their deleted data and believed they were safe after deletion. The interviews further suggested that this belief affected their view on recovery from the cloud; according to these users, data recovery from the cloud is not possible. They believed that the moment one clicks 'Delete' a file will be deleted from the cloud.

There seemed to be a conflict within the individuals who had this model. On the one hand, they believed deletion was permanent because they could not recover deleted data in the cloud. On the other hand, they also believed that deletion in the cloud could not be permanent since the cloud is an online service. This conflict was caused by the belief that online services do not delete data; therefore, cloud being an online service would also not delete data. A handful of participants, however, did not fall into these conflicting models. They suggested deletion in the cloud was not permanent and even constructed attack models for deletion:

> *‘‘...you delete something, but they still keep a copy of it and then some can hack in and get your information. I think because they keep a copy of everything, so I think after deleting they still keep a copy,... it means that somewhere they keep information that could be retrieved later, but whether that information is kept confidentially or [if] it could be hacked, people hacking in and getting other people's pictures and then blackmailing them and stuff." P24*

Providers don't delete. A group of participants believed cloud providers do not delete for advertisement and research purposes. They held the view that there is a "secondary" storage where deleted data is stored, but users are not allowed to access it. Although they reported high privacy concerns, they also exhibited some defeatism:

> *‘‘I never read the T&Cs I don't really know if it's deleted forever. It's probably still stored somewhere, but I don't have immediate access to it." P11*

> *‘‘I think the provider might not want that[deletion] to happen because they keep the data and then they want to use it for marketing and, you know, different purposes." P24*

However, these participants were not concerned about their data being used for adverts because they believed that as long as they could achieve what they wanted to do, this was acceptable.

> *‘‘I'm not so much concerned with that [deletion] as to the underlying reasons, and the drivers for the business are more important to me than their terms and conditions." P26*

4.2.3 User interface issues

As expected and reported in other studies, interface issues negatively affect users' deletion process. Poorly designed user interfaces caused much distress to some users which resulted in them losing interest in deleting or left them frustrated. Users are affected by screen sizes, type of interface (i.e., whether it is a web or mobile application), and how the deletion process is completed in that application. This finding also suggests that when users find it difficult to use an interface to delete, they are unlikely to attempt to delete using the same interface in the future. A small number of participants who access the cloud through mobile phones reported that sometimes they do not know where to go in their mobile phone interfaces in order to delete from the cloud.

Effort.

Some cloud features (e.g., autosaving) affected users' deletion process. Users who have auto synchronisation enabled ended up not deleting from the cloud because much effort may be required from them to delete all the unwanted files synchronised to their cloud storage. One participant who had this feature turned on informed us that sometimes their smartphone accidentally takes photos while in their pocket leaving them with a lot of unwanted photos. It required a lot of effort and time to delete such photos from their phone and then from the cloud, as a result sometimes they chose not to delete from the cloud.

> *"... often with mobiles these days if the camera goes off in your pocket, which it often does, you can end up with all these blank photographs. Of course, they go into OneDrive, so you look at your OneDrive, and you want to clear all that. But sometimes it can take ages." P23*

Buggy software.

Buggy applications and unresponsive interfaces left some users not being able to delete. Some users (4/27) reported that there are times when they try to delete and the app or web interface would not respond resulting in them abandoning the process. Less satisfying mobile apps and unresponsive web interfaces resulted in users having less desire to delete. For example, some users reported that for them to delete, they have to try it a couple of times before the operation successfully completes. Users found this annoying and preferred not to try deleting when they wanted.

4.2.4 Coping mechanisms

Regarding coping strategies, further data analysis revealed that users had developed different coping mechanisms to address or mitigate their failure to delete from the cloud. These mechanisms were ranging from ignoring deletion altogether to changing cloud providers through to seeking help from others and ad-hoc strategies. The next section discusses various coping mechanisms employed by participants when they encounter challenges when deleting.

Head in the sand

Most participants who could not delete preferred to leave their files undeleted in the cloud. There are four reasons why users settled on this strategy: (i) They perceive this method to be easy, and quick—it does not require them to put in any effort. Users who felt deletion could be burdensome preferred this method. (ii) When a user has sufficient storage left on their account, they are highly likely to leave the files in the cloud. (iii) When users perceive the file to be harmless or non-confidential. (iv) Trust in the cloud provider: when users trust the service provider, they are more inclined to leave data in the cloud.

Cloud hopping

Those with high privacy concerns or low-levels of trust in providers often opted to stop using certain cloud services or changed their service providers. Although this may be considered an extreme measure, users weighed the benefits of using a cloud service against the cost of their undeleted data being exposed. Others noted that they changed providers because of running out of storage space with their current provider. Interestingly, none of the participants who changed providers deleted or deactivated their accounts with the previous provider.

Excluding certain files from the cloud

Some participants (2/27) reported that they explicitly decide on what goes into the cloud before they upload to the cloud. By excluding potentially sensitive or confidential documents and sharing them by offline means, such users believed they were safe, and they did not have to worry about undeleted information. This approach is common among users who have high privacy concerns and low trust towards providers. However, this strategy may open up other threats regarding data exposure, e.g., through lost removable (potentially unencrypted) media—using USB sticks to share sensitive files.

Deleting from one device

Some participants (3/27) reported that they only deleted from devices they were more comfortable with and were confident would yield expected results. For example, opting to delete from their computers (sync folder) than to delete from the web interface or mobile applications.

> *‘‘Yes, I have the app on my phone, but I rarely use it, my app. I have downloaded the app on my desktop. So, I delete from there instead . . . I mostly delete it from the desktop because I found it difficult to delete it on the phone." P24*

This strategy does delete the file from the cloud; however, this may lead to delays in deleting a file because the user may not always have the device they are comfortable with all the time.

Seeking help

We found that participants ask for help from their friends, family and colleagues (Nthala and Flechais 2017) if they think it is urgent and important that a file should be deleted. Others revealed that they would search online for solutions, for example, from tutorials, forums and blogs. The type of help sought depends on users' motivation to delete. For example, when the motivation to delete is due to high privacy concerns or trust issues, then the user will not hesitate to ask their social network for help. However, when the file to be deleted is sensitive (e.g., explicit photos) or confidential, users have a likelihood of not seeking help from other individuals with the fear of being exposed or shamed. They would opt for looking for solutions online. This strategy leads to high chances of deleting from the cloud and participants who reported they knew how to delete revealed this is how they learnt about deletion.

Deleting a different file

When participants are deleting to free up space, and they fail to delete a certain file, they will instead choose a different file to delete than the original one. One participant explained that sometimes they get a warning not to delete a file, but because they do not fully understand the consequences of deleting that file, they will instead look for a different file to delete.

> *''. . . I will not delete that one, I will try to find something else to delete instead of it, to get more space otherwise, I can't. So, at times I will remove a different file because I have not [yet] found a correct solution on how to get over it." P24*

Ad-hoc strategies

Participants adopt ad hoc strategies when they encounter challenges. Some participants (3/27) revealed that they did not have a well-defined method of overcoming the failure to delete. They reported that they would try to find the best possible solution that fits and suitable for that moment in time. Some participants (2/27) reported that if they cannot delete a file, but they feel it is important to delete such a file from the cloud, they would try deleting it from all their devices including the web interface. One participant revealed that when they are in need of more storage and they cannot delete they will buy more storage.

4.2.5 What deletion experience do users want?

The previous section discussed different coping mechanisms employed by users when they struggle to delete. This section focuses on the themes and categories that emerged as to what cloud users want. Four key themes were identified across participants.

Transparency in deletion

Participants want providers to be more transparent about the deletion of data; they wanted information on how their data is deleted to be made freely available. Participants who struggled to delete suggested that providers should provide tutorials on how to delete and that deletion information should be made clear when they first sign up for services.

However, those who could delete were not so much interested in such information but rather in knowing more about how deletion is done in the cloud. Others suggested that they should receive notifications when data is completely removed from the cloud. That is, they want guarantees that their data is completely gone from the provider. This was popular among participants with privacy-driven deletion practices.

Complete deletion

Early in this study, the assumption that cloud deletion was complete and instant emerged. To explore this further, participants were asked what deletion meant to them. Most users defined deletion as “getting rid of” or “destroying data”. Most users further revealed that data could not be recovered after deletion. This implied that cloud deletion is complete to them. Despite this group of cloud users, participants who had better knowledge of the cloud and deletion stated that they desired complete deletion (or permanent deletion), implying that the current deletion process is not complete:

> *‘‘I suppose all data should be completely deleted. Once you press ’delete’, delete should mean delete, so then you don’t have that sort of a grey area as to what’s sensitive, what’s not sensitive. Delete should mean delete, I think." P6*

> *‘‘The moment I delete something from my iCloud, or my laptop, I want it to be deleted completely. I feel like once I [have] said I don’t want this on my laptop again, or I don’t want it on my phone, I would rather have it deleted everywhere, complete.” P3*

After informing the participants who thought cloud deletion was complete that it is possible that data may still be in the cloud after deletion (Tang et al. 2010), most of them responded with shock to this revelation while others reported they had always thought it might not be deleted.

> *''I've always had it in the back of my mind that what you delete does not completely go. I didn't know like it's almost impossible to delete... " P3.*

While most participants (17/27) reported they would want complete deletion, this was not true universally. Some, understandably, wanted to have the opportunity to recover deleted data especially data deleted by mistake.

With regards to complete deletion, participants exhibited the following beliefs: (1) data perceived essential or confidential should be completely deleted, (2) Data belonging to other users or data that contains identifiable information should be completely deleted, and (3) only law-abiding citizens should be allowed to completely delete things from the cloud.

With regards to the final point, users who perceived themselves as harmless and law-abiding citizens did not mind if their data was not completely deleted. However, they reported that if the data did not belong to them or contained other users' identifiable information, then they would want it completely gone. Some participants (7/27) believed that complete deletion would enable law breaking citizens to commit and hide their crimes in the cloud. Despite this, other users reasoned that they would still choose permanent deletion because it is their data and no one has the right to access it after deletion. With regards to recovery, such users reported they would change the way they work and just be careful when deleting data. Participants who belong to this group were users who had high privacy concerns about the cloud and would rather lose data because of mistakes than have it undeleted in the cloud.

Contact Point

During the interviews, there are participants who reported that it was hard for them to get verifiable information on using the cloud. They suggested that a service dedicated to resolving their cloud queries would be useful especially when they cannot ask anyone. One participant put it this way:

> *''[First thing is,] I don't know whom to call. If I want my data back I don't know who to call, whom to contact because I don't think they have any helpline or service like this where a customer can call and say, 'I deleted by mistake, send it back to me,' or I don't know whether the provider has access to retrieve particular data of a customer." P24*

Control over deletion

The analysis performed in this research suggests that users feel the need for control over deletion in the cloud. They wanted to be the ones who decide when a file should be permanently deleted or held for potential recovery. It also shows that users want control over what is synced over to the cloud so that they will not have to delete.

> *‘‘Because if I have deleted something, I am saying, ‘I don’t need it anymore,’ or, I don’t want evidence of it anymore,’ then surely it should be deleted completely because I no longer have the use for it. Who’s meant to still have use of what I’ve deleted?” P22*

4.3 Discussion

Table 4.3 A summary of motivations to delete and coping strategies

Motivation to delete	Coping Strategy
Privacy-driven	Seeking help Deleting from single device Excluding certain files from the cloud Cloud hopping Ad hoc strategies
Expertise-driven	Head in the sand Ad hoc strategies Seeking help Excluding certain files from the cloud Cloud hopping
Policy-driven	Ad hoc strategies Deleting from single device Seeking help Excluding certain files from the cloud
Storage-driven	Cloud hopping Delete a different file Deleting from a single device Ad hoc strategies

4.3.1 Deletion motivations and coping strategies

Users’ choices and development of coping strategies are dependent on context, time and their motivation to delete. Participants who mentioned having experienced challenges while deleting were also asked how they responded to such challenges. Table 4.3 shows a mapping between users’ motivations to delete and the coping strategies they may adopt when facing challenges.

Motivation: Privacy-driven

Users whose motivation to delete is privacy-driven are usually more likely to seek help in deleting or have a higher chance of employing some ad hoc methods to try to delete from the cloud. Ad hoc strategies do not always guarantee data will be deleted from the cloud.

When struggling to delete some of these users may opt to delete from the device they are most comfortable with or confident that they will manage to delete data using it. This choice is usually based on users' past experience; the user chooses it because it has worked for them before.

Motivation: Expertise-driven

Expertise-driven users often resort to ad hoc strategies when they cannot delete. If ad hoc strategies do not work, those with high self-confidence would either decide to leave the file in the cloud or hop to another provider. Such users do not usually ask for help because of their self-belief. Users with less skill and low confidence in using the cloud are likely to leave the file undeleted but not change the provider. They will only change the provider if they are confident of using the platforms/interfaces from the new provider because they do not want to go through the process of having to relearn how to use the new cloud provider. Expertise-driven users may also resort to excluding sensitive files when using the cloud to avoid the anxiety associated with not being able to delete the file from the cloud.

Motivation: Policy-driven

Users whose motivation to delete is policy-driven usually fear the consequences of having that data not deleted in the public cloud. They usually adopt ad hoc strategies as their first coping mechanism, and if they still cannot delete from the cloud, they would then attempt to delete from the device they are confident in using. They would finally ask for help if everything they tried has failed. Nonetheless, prior failure to get data deleted causes them to exclude valuable or work-related files from the cloud.

Motivation: Storage-driven

Users who delete for storage reasons adopt strategies such as deleting a different file, deleting from a single device, cloud hopping and some ad hoc strategies. Deleting a different file may temporarily create space, but as the number of files that the user cannot delete increases, the user eventually runs out of space. Deleting from a single device and some ad hoc strategies may yield results since files get deleted. However, other ad hoc strategies like buying more storage cost the user but helps them fulfil their initial goal without the need to delete. The results of cloud hopping are temporary; it only works until the user fills out all the new storage provided. In general, a cloud-hopping strategy does not scale as users are unlikely to keep changing providers regularly.

4.3.2 Mental models and coping strategies

A potential connection between users' mental models and coping strategies was observed. Table 4.4 summarises how users' mental models and their coping strategies are linked. This is a mapping between participants' beliefs and what they commonly do when faced with challenges while deleting.

Table 4.4 A summary of users' mental models and their coping strategies

Mental models	**Coping Strategy**
The cloud within an app	Seeking help Deleting from single device Excluding certain files from the cloud Cloud hopping Ad hoc strategies Head in the sand
Borrowed mental models	Head in the sand Ad hoc strategies Seeking help Cloud hopping
Sync folders are not the cloud	Ad hoc strategies Deleting from single device Seeking help
Providers don't delete	Cloud hopping Excluding certain files from the cloud
Shared folder: Deletion is one sided	Head in the sand
Shared folder: Deletion affects all members	Head in the sand Seeking help
Deleting and saving work the same way	Seeking help Cloud hopping
Deletion is permanent and instant	Head in the sand

The cloud within an app.

Users who believe the cloud is inaccessible are likely to seek help in order to delete since they do not believe they can actually log on and delete. Others may just try to delete using the single device that they believe is connected to the cloud, which may not be as effective as discussed earlier. Participants may choose to cloud hop in search for a cloud that they believe they can access and delete data from, or they may leave files undeleted. Some privacy-driven users with this belief may opt to exclude files from the cloud, only storing files they perceive to be not confidential.

Borrowed mental models.

Upon realising that their deletion understanding is different from cloud deletion, users may adopt a head in the sand approach, change to a new provider, seek help, or try other ad hoc strategies. They may adopt a head in the sand approach because they recognise the mismatch and unexpected outcome. Such users may hop to another cloud where such beliefs may yield expected results, and they may finally seek help after trying some ad hoc strategies.

Sync folders are not the cloud.

Users who believe sync folders are not part of the cloud may adopt ad hoc strategies to delete from the cloud. Participants may seek help to delete while others may delete from web interface instead of the sync folder.

Providers don't delete.

These users are likely to change providers or choose to exclude files they perceive to be important from the cloud. Although changing a provider does not solve their deletion issues, they believe the new provider will provide a certain assurance of deletion.

Shared folder: deletion models.

Interestingly, though shared folder beliefs are different, in both cases, users employ the same coping strategy: head in the sand. Users who perceive deletion to be one-sided may choose to leave the files undeleted believing that deleting would not delete the files from other members of the shared folder, while, those who perceive that deletion affects all the members of the shared folder may choose to ignore the file because they do not want to delete files which other users are still using. Users who believe that deletion affects all members of the shared folder often seek help to confirm whether they could delete from the shared folder.

Deleting is permanent and instant.

The main challenge faced by users with this model is decision making when it comes to deciding whether a file should be deleted or not. They will believe that it will get deleted forever and instantly. As a result, these users often resort to leaving files undeleted in the cloud with the belief that they might need them again.

Deleting and saving work the same way.

Users who possess this model tend to seek help when encountering problems with deletion. Some may also choose to delete from the device where this beliefs accurately applies, hence leading to successful outcomes. In this case, users rely on recalling previous successful deletion experiences.

4.3.3 Design implications

This study revealed a major gap in users' understanding of how the cloud and deletion work. Although the responsibility to delete lies between the providers and the users, this study pointed out that cloud users want more transparency regarding cloud deletion policies. Information on deletion should be clear and easily made available for users especially about how data is disposed of after use. Users would also benefit from cloud providers making deletion mechanisms easy to understand and accessible.

Regarding help, cloud users would like to have the option to contact someone directly concerning their deletion problems. This implies that some cloud interfaces provide users with not enough feedback or complicated information which is hard to understand. Something akin to Deletion Service Points would help users resolve their problems quickly. With regards to user interfaces, improving user feedback (e.g., notifications during deletion) would inform users of the results of their actions, therefore, influencing or improving their weak understanding of cloud deletion.

Another possible avenue for improving the current situation is improving users' understanding and awareness-building. This study found that users possess different mental models at the same time, of which most are incomplete and lead to failure to delete. This may suggest that these differences make user education a challenging task hence such education should be customised. Also, since not all incomplete mental models lead to failure to delete, user education should focus on maturing mental models that are weak or those that lead to failure to delete.

4.3.4 Limitations

This study is a qualitative inquiry – based on a sample size of 26 participants. This sample is sufficient for such a study and saturation in grounded theory was reached by 13 transcripts. As such one can be confident that the motivations, failures, coping strategies and desires discussed in Section 4.2 are grounded in the data from the study. To account for coding biases, a second researcher verified the codes emerging from the grounded theory analysis.

However, the participants who reported that they used the cloud for editing documents could not distinguish between Microsoft Office Online, Office 365 and Office 2016. This may have influenced their judgment and perception of deletion from the cloud. At the same time, it further reflects the inaccuracy of users' mental models with regards to the cloud. Future studies ought to explore the users' mental models of the cloud in general and their impact on various user interactions with the cloud with regards to security and privacy.

4.4 Summary

This chapter presented the details of the user study which was conducted to understand users' cloud deletion practices. It has identified what motivates users to delete from the cloud, the challenges, and the mitigations that they adapt to ease their problems. Cloud users' deletion mental models that hinder users from deleting are also highlighted in this chapter and lack of relevant deletion information that could help them delete properly. More importantly, this chapter has shown the need to offer deletion mechanism that can meet users' different deletion needs. For instance, a user may delete to preserve their privacy, or to tidy up their account – manage old data. These different motivations for deletion may require different ways of deleting, for instance, deleting to preserve one's privacy may require that deleted data be removed entirely from one's cloud account, while deleting just to tidy up one's account may not need deletion to be complete (e.g., moving data to the 'trash' folder may be enough). However, as previously mentioned, prior work does not consider these different needs, it assumes all data is deleted the way. In the following chapter, Chapter 5, deletion preferences are studied in depth — to understand what deletion preferences exist and what affect users' choices concerning these preferences.

This chapter has also highlighted the deletion challenges that users face when deleting from the cloud. Users failure to delete leads to construction or development of coping mechanisms to address the problem. Users develop these strategies if they believe that it is essential that data be deleted from the cloud. However, information on deletion affects how users construct deletion mental models. The lack of information on deletion leads to the construction of incomplete or inaccurate mental models which eventually leads to a failure to delete. These mental models have a direct impact on the choices and the development of coping strategies. Incomplete or inaccurate mental models may lead to the development of strategies which do not delete data from the cloud, or strategies which only solve the problem temporarily or bring up additional problems. To alleviate these challenges, Chapter 5 focuses on information about cloud deletion, it aims to understand what information about deletion is important to users, when and where such information should be made available for users.

Chapter 5

Cloud Deletion Preferences and Information Requirements

Note: *The content of this chapter is based on a paper which is currently under review.*

The previous chapter has shown that users have various deletion needs but currently face many challenges with regards to deletion. Earlier research has also shown that current cloud services lack mechanisms that meet deletion needs of users (Khan et al. 2018, Murillo et al. 2018). However, before proceeding and building such mechanisms, it is better to understand the patterns of preferences in deleting different types of data from the cloud by various users. It is essential to know what type of data requires what type of deletion, and in what situations. Only then we can build mechanisms that will meet users' deletion needs. Which factors affect these preferences, or do they differ when one is deleting from individual folders and shared folders? Also, are there any other external factors, all these can reduce complexity or inform what the cloud systems need in order to provide better deletion mechanisms that meet users' needs. In addition to understanding users' deletion preferences, there is a need to provide better information on cloud deletion. The findings of Chapter 4 highlighted the absence of information on cloud deletion that can inform users while deleting or help them recover from deletion challenges. It is also necessary to understand what information on deletion is needed and when can this information be made available to users.

This chapter explores users' cloud deletion preferences and information that may support their deletion needs in the cloud. Drawing from participatory action research, the work presented in this chapter seeks to answer the following key questions: (1) what do users want, their deletion preferences, and (2) how do they want to be informed about deletion?

To address the first research question, this chapter aims to understand the following: (1) how do cloud users classify data stored in the cloud, what kinds of data do they treat similar and differently with regards to sensitivity and importance? (2) How do they want these data to be deleted? What preferences can be identified from this, considering (a) deletion under individual context and (b) social context – shared folders? Are these preferences consistent? (3) Is it feasible to design deletion mechanisms that satisfy these preferences?

To address the second research question, this chapter will examine the following: (1) what information about deletion do users consider crucial, (2) when do users want this information to be presented to them, and (3) where do users prefer to find such information.

In summary, the key contributions of this chapter are as follows:

- It identifies three groups that cloud users commonly use to classify cloud data and discusses how these classifications are made under individual and social context. The groups that users classify data under are: (1) *essential and sensitive*, (2) *less important and less sensitive*, and (3) *important but less sensitive*. Essential and sensitive – this is a group of data that users consider to be important or useful and private. Less important and Less sensitive – a group of data that users consider having less value, easy to produce and less private, while important but less sensitive data is the data users consider useful and hard to produce, but they are happy with other people knowing or seeing it.
- It provides an empirical exploration of cloud deletion preferences through a participatory research exercise, thereby showing that users' deletion preferences are: (1) not always aligned with how they classify data, (2) often changing as users' life, file utility and deletion needs change, and (3) complex and multi-dimensional, depend on the user's context of deletion – personal, social and external factors.
- It elicits what information users consider essential with regards to deletion, the time at which they prefer to be provided with such information and the communication channels they prefer to have such information. It demonstrates that satisfying cloud users' deletion needs require more than just one solution. Consequently, it explores possible solutions to the problem.

This chapter is structured as follows. Section 5.1 describes the methodological approach and the demographics of the participants. The findings of this chapter are presented in Section 5.2, 5.3, and 5.4, while Section 5.5 discusses the implications of this study.

5.1 Methodology

To investigate deletion preferences, 20 active cloud users (divided into four groups of 5 users) were invited to take part in three participatory action research (PAR) activities (Bergold and Thomas 2012, Weber et al. 2015). The work presented in this thesis uses and takes advantage of one important aspect of this method — include the ultimate arbiters of the systems, that is, those who daily use the system in their lives. It involves participation and action from a group of users who are affected by the same problem and act together to tackle it. As a collaborative research methodology, it offers researchers the opportunity to co-develop or investigate with users. It stresses the lived experiences, social changes and construction knowledge of users, which can be useful for solving their everyday challenges (Bennett 2004). As a result, PAR discovers and develops solutions that are viable and useful to them. Until we become familiar with and prioritise users' concerns and cloud deletion preferences, there is little that technology can be improved to help users achieve their goals. Taking part in this study involved completing a pre-study survey, and then three tasks as shown in Fig. 5.1.

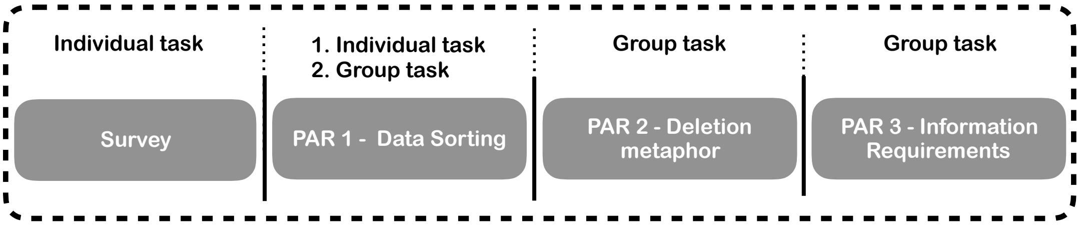


Fig. 5.1 Study methodology. The survey was taken prior to attending the study sessions. Participants first classified data types, then completed the task based exercises.

Pre-study Survey

Before attending a session for task-based exercises, participants were asked to complete an online pre-survey that obtained demographics details and assessed their perception, cloud deletion preferences and practices. This survey contained six demographics questions and twenty-four questions about cloud deletion. It took an average of 20 minutes to complete. A copy of this survey can be found in Appendix B.2.

PAR1 – Data sorting

To explore the deletion preferences of users, a new sorting task using multiple data types was developed. The purpose of this experiment was to understand how users classify cloud data and investigate whether these classifications had any influence on their deletion

Table 5.1 Summary: List of data types used in the data sorting activity.

Data types			
Medical report/ information	Music videos	Old birthday video	Research papers
Rifle licence	Honeymoon photos	Genetic information	Facebook downloaded data
Immigration documents	WhatsApp backup	Family photos	Research data
Personal information	Meme videos	Job application letter	Biometric card copy
Biometric data	Meme images	Legal documents	Business contracts
Passport copy	3GB wildlife video	E-books	Friends photos
Old bank statements	Children photos (Family)	Pet care information	4MB video clip

preferences. We selected 25 data types (see Table 5.1) for this task. These are types of data (e.g., Whatsapp backup) that may be stored in the cloud and were from our pre-study survey and prior research (King and Raja 2012, Srinivas (last accessed Sep 11, 2018, Lee et al. 2016, Henry (Last accessed Sep 14, 2018, WhatsAppInc. (accessed June 11, 2019)). These were data types that users were already familiar with and mixed them with other data that are considered important and private in a social context.

This thesis adopted a free-sorting technique (Blanchard and Banerji 2016) for this experiment. For this task, participants (as individuals first and then as a group) were asked to categorize data according to how they perceive it so that similar data types are gathered together. Categories were not predetermined; participants were not told what kind or how many groups they are to expect. Participants were simply requested to categorize data in a way that made sense to them or how they wish. The objective of this task is to visualize and identify users' intuitive categories, to see what categories are created based on how users perceive these data. Consequently, to understand what data they consider important (utility) and sensitive (privacy) and if these would have influence in their deletion preferences. This activity took individuals around 5 min and 13 min as a group.

By letting participants to cluster these types of data into different categories as they perceived, this thesis hoped to expose the range of variability that may exist from data stored in the cloud. This thesis predicted that there would be some overlaps between the groups (e.g., data considered sensitive and important) which may later affect how they would want such data to be deleted. The researcher also anticipated more categories during individual sorting than in group sorting because groups tend to summarize categories (Blanchard and Banerji 2016). These observations are critical to our understanding of users' deletion preferences because they carry evidence on whether how users perceive data has any influence on how we want our data to be deleted.

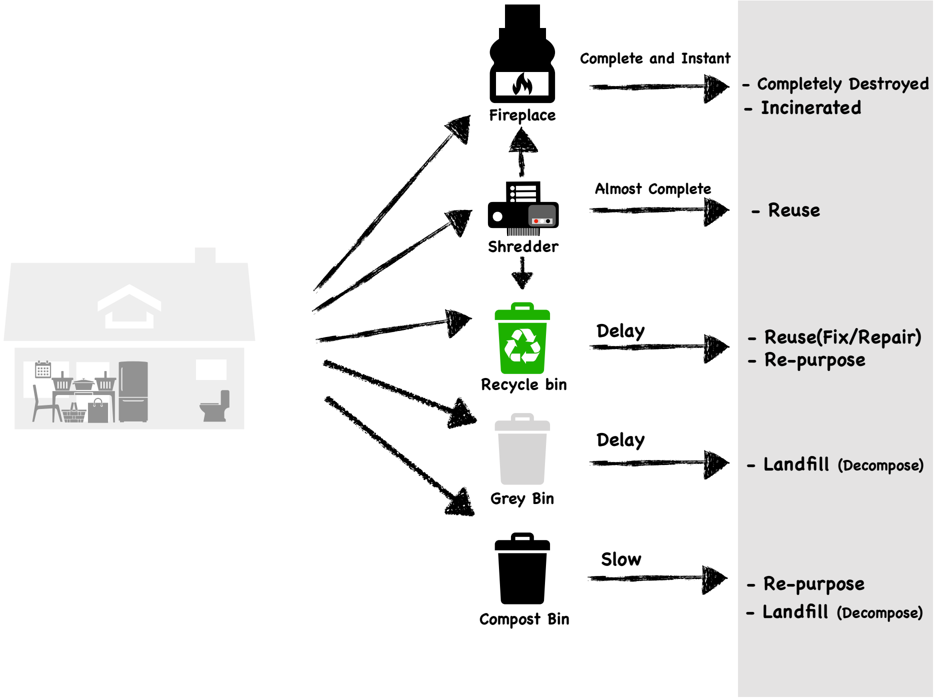


Fig. 5.2 Deletion metaphor task: Participants completing the metaphor task.

PAR2 – Deletion Metaphor

In order to help elicit cloud deletion preferences, this task included a garbage collection metaphor activity. This metaphor was used in the study to minimise participants introducing their metaphors and leading to inconclusive results since users can possess various mental models (*see Mental models 4.2.2*), some of which may be incomplete. Using a single metaphor would help control the study environment and yield better results. The use of metaphor is a well-known HCI method to help users think about digital objects as they would think about real-world objects to increase their familiarity with them (Dix et al. 2003, Blackwell 2006). This activity was divided into two parts.

In the first part, the metaphor depicted how household waste is managed. A diagram of a house and five empty boxes depicting different ways of managing waste was drawn on an A3 size paper (see Fig. 5.2). The boxes had a picture of a *fireplace*, *shredder*, *green bin*, *grey bin*, and *compost bin*. Next to each picture was a text describing properties of each box – how waste is disposed into each box. This included eight paper labels for waste. The labels were: *old bank statement*, *confidential letter*, *Fizzy or Soda can*, *newspaper*, *milk carton*, *rotten apples*, *candy wrappers*, and *old working computer keyboard*. Using these props, participants were asked to place each type of waste in the bin that they found appropriate for that type of waste. This metaphor was chosen because it somehow depicted deletion "destruction of unwanted material" and participants could easily understand and relate to it. Starting with this metaphor also helped participants to settle down during the study.

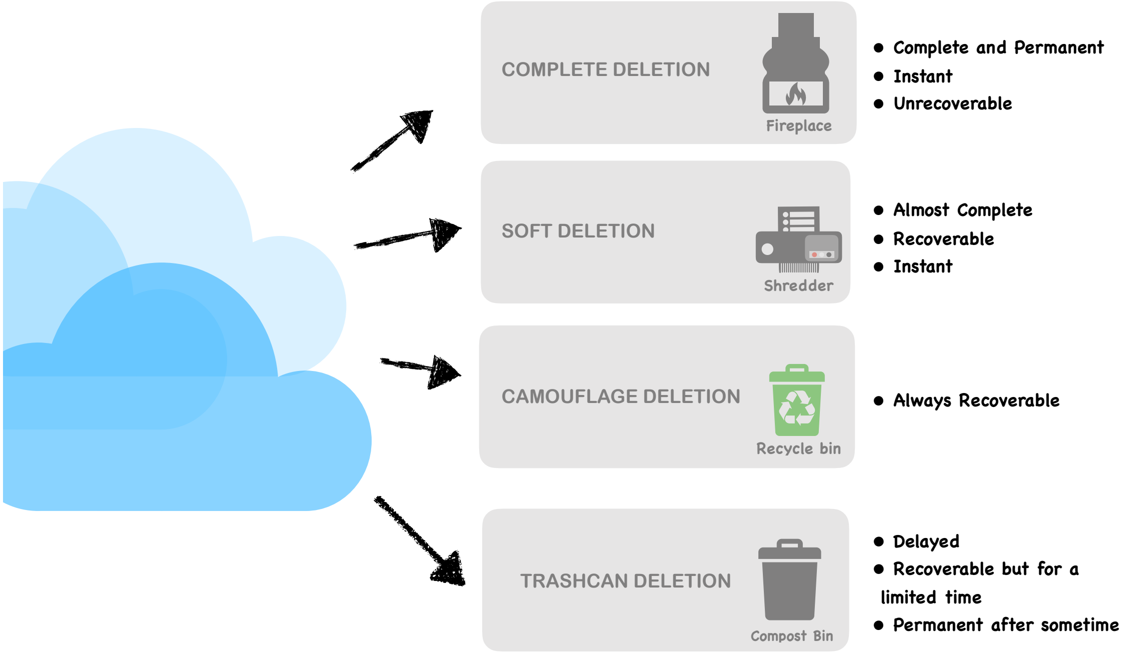


Fig. 5.3 Deletion preferences activity

The second part of this activity was tailored around the first metaphor; it used the main concepts to design a deletion preference activity which required participants to sort out how they would delete cloud data. Instead of having boxes depicting just different bins, the boxes now depicted different types of deletion and their properties (see Fig. 5.3). This task considered four main types of deletion or properties according to previous literature (Reardon et al. 2013, Murillo et al. 2018). For easy understanding, these types of deletion were named as follows:

1. Complete deletion – deletion that removes all copies of data permanently from the cloud.
2. Soft deletion or partial deletion – nominal deletion where some parts of the deleted data can still be recovered.
3. Camouflage deletion – deletion that removes data from the users' locality but can always be recovered.
4. Trashcan deletion – deletion that allows recovery for some time before data is completely erased.

Just like the waste task, this included twenty-nine data labels (from the data sorting task) for participants to group among the boxes. Each group of participants was asked to sort the provided data labels by how they would want such data to be deleted in two different

Table 5.2 List of information used in the study together with the concepts considered.

Concept Considered	Information
Backend, Shared copies	In whose account does data in shared folders reside
Backend	How data is deleted
Backend	How data is deleted from a "sync folder" or "cloud folder" in my computer
Backend	How much storage size you have left
Shared copies	How shared content is deleted from a shared folder in the cloud
User Interface	How data is deleted from the cloud using a smartphone
Shared copies	How does shared folder work
Backend	How data is stored in the cloud
User Interface	How is data deleted from a web interface
Accountability	Who has access to deleted data in the cloud
Backend	Whether data is deleted completely
Accountability	Who has access to all data stored in the cloud
Shared copies	Who has the rights to delete from a shared folder
Anonymisation, Backend	What happens to deleted data
Backend	What happens when I delete my cloud account
Shared copies	The number of copies of data stored in the cloud
Backend	The location where data is stored
Time	The period of time it takes to completely delete from recycle bin
Time	The period of time it takes to completely delete from the cloud
Time	How long it takes for all copies to be deleted from the cloud
Time	How long it takes to completely delete a cloud account
Shared copies	How copies are deleted
Backend, Data Recovery	Data recovery after data has been deleted from 'deleted folder' or 'trash folder' in the cloud

situations, firstly, when deleting from a personal account, and secondly, deleting from shared folders.

PAR3 – Information Requirements

The last activity concerned information about cloud deletion. There were twenty-three (23) labels containing information related to the cloud and deletion. This information covered deletion concepts suggested by prior research (Murillo et al. 2018) such as time, shared folders, copies, back-end and the User Interface (UI). Participants were asked to categorize these in three ways: (1) the order of importance with regards to deletion (modality), (2) the point (time) when they would prefer to see the information, and (3) where they would prefer to find that information(channel). Table 5.2 shows the list of the information used in the study.

Unlike *PAR 1* where participants were given not categories at the beginning, in this activity, participants were given some categories to use. Participants were given these categories to simplify and reduce task time. However, they were not restricted to these categories and were informed they could create other categories if they need to. Regarding modality, participants were given three categories: the most important info, less important info and neutral category. In terms of time, participants could choose between before signing up for cloud service or deleting, during usage or deletion, and after deleting. With regards to the communication channels, they had privacy policies, blog pages, dialogs (including popups), adverts, FAQs and main account page. These channels were chosen because they are commonly used for disseminating cloud information.

Table 5.3 Summary: Study Demographics.

Code	Group	Gender	Age	Employment	Accounts	Cloud Services
P1	A	Female	21 - 25	Student	2 - 3	Dropbox, iCloud, Google drive
P4	A	Male	31 - 35	full time	4 - 5	Dropbox, Box, Google drive, OneDrive
P11	A	Female	31 - 35	full time	2 - 3	Google drive
P17	A	Male	18 - 20	Student	2 - 3	Dropbox, Google drive, OneDrive
P18	A	Male	31 - 35	full time	6 +	Dropbox, iCloud, Google drive, OneDrive
P2	B	Male	31 - 35	PhD Student	2 - 3	Google Drive Box
P5	B	Female	31 - 35	full time	2 - 3	Dropbox, Google drive
P7	B	Female	18 - 20	Student	1	Google drive
P13	B	Male	26 - 30	full time	2 - 3	iCloud, Google drive
P19	B	Female	41 - 45	full time	4 - 5	iCloud, Google drive, Dropbox
P14	C	Male	26 - 30	full time	2 - 3	iCloud, Google drive, OneDrive
P3	C	Male	26 - 30	PhD Student	2 - 3	iCloud, Google drive
P8	C	Male	21 - 25	full time	2 - 3	Google drive, OneDrive
P9	C	Female	18 - 20	Student	1	Google drive
P20	C	Female	31 - 35	full time	2 - 3	Google drive, OneDrive
P10	D	Male	26 - 30	part time	2 - 3	Dropbox, iCloud, Google drive
P6	D	Female	26 - 30	Unemployed	2 - 3	Dropbox, Google drive
P12	D	Female	26 - 30	PhD Student	1	Google drive
P15	D	Female	31 - 35	Student	4 - 5	Google drive, OneDrive
P16	D	Male	36 - 40	full time	2 - 3	Google drive, Box, Amazon Cloud Drive

5.1.1 Recruitment, Ethics and Data collection

After obtaining an ethics clearance, participants were recruited through social media, word of mouth, and advertisements around the University of Bristol and the city centre. Interested respondents were encouraged to complete a screening form hosted online. The purpose of this questionnaire was to identify active cloud users who were 18 or older meeting three or more of the following: having deleted from the cloud through more than one device or interface, having more than one cloud account, sharing some folders, having experienced some challenges when deleting, interested in cloud deletion, and being able to attend the participatory study. The aim of having participants who manage more than one account or deleted from different devices was to increase the chances of having varying discussions concerning users' experiences regarding usability of cloud deletion.

Sixty-five (65) participants (40% identified as male) responded to the adverts and completed the screening questionnaire. Sixty stated that they have deleted in the cloud, 17 of whom experienced some challenges while deleting, 76.9% sharing folders, 46.2 % had more than one account, and only three participants stated they could not attend to do the study.

In the end, 20 (10 females and 10 males) participants were invited to take part in the study. These were divided into four equal groups. In order to improve and encourage deep discussions within each group, gender, employment, number of cloud accounts and the cloud service participants used were considered during group creation. The goal of this was to ensure diversity within each group, i.e., different participants with different experiences of using and deleting from the cloud. For example, each group contained at least three non-student participants decreasing the chances of having participants with identical use of the cloud in the same group. Table 5.3 lists the demographics of all the participants.

The study sessions were on Tuesdays and Thursdays of the first two weeks of July 2018. Written consent was required to record audio and take pictures (i.e., pictures of the props without showing the faces of the participants) during the sessions. Participants who took part in the participatory study sessions received compensation worth £5.00 for their time.

5.1.2 Data Analysis

The data were examined using thematic analysis (Braun and Clarke 2006). This is a method for identifying, analysing and reporting patterns emerging from the data. One of the benefits of using thematic analysis is its flexibility. It allows researchers to determine broad patterns that are rich and detailed, nonetheless providing an elaborate description of data (Nowell et al. 2017).

This study followed the six steps of analysis suggested by Braun and Clarke (2006). To familiarise themselves with the data, the lead researcher conducted the participatory sessions and transcribed the audio scripts from the participatory sessions. Then, the lead and the second researcher independently coded all the data from the first group session and generated codebooks. Both researchers met and discussed the codes, refining disputed codes. Disputes were found to have been caused by the two researchers interpreting some codes differently. A single codebook was then generated. After compiling the codebook, the two researchers independently coded the transcript from the second group using the compiled codebook. Cohen's Kappa coefficient agreement was found to be 0.72, showing a high degree of agreement between the two researchers. The lead researcher coded the last remaining scripts.

After the initial coding was complete, the process of identifying themes started. Both researchers searched for potential themes and grouped similar codes. Then the groups were refined and named. This was an iterative process which at times involved going back to the scripts to understand the context. Some groups were further broken down to allow better definitions and naming. Researchers identified and reported on interesting themes. These were themes that are related to what influenced deletion preferences.

As part of the analysis and with the help of photos from each session, the lead researcher recorded each group's data sorting patterns and deletion preferences in the form of tables. Then, a comparison between deletion preferences and data type sorting was made. This comparison aimed at finding if there was any relationship between how participants sorted the data and how they wish it could be deleted. The first comparison process was focused on how each individual group sorted the data and their deletion preferences while the second comparison was between the groups (i.e., identifying whether there is a relationship between the way each group categorised data and how they wish it could be deleted–deletion preferences).

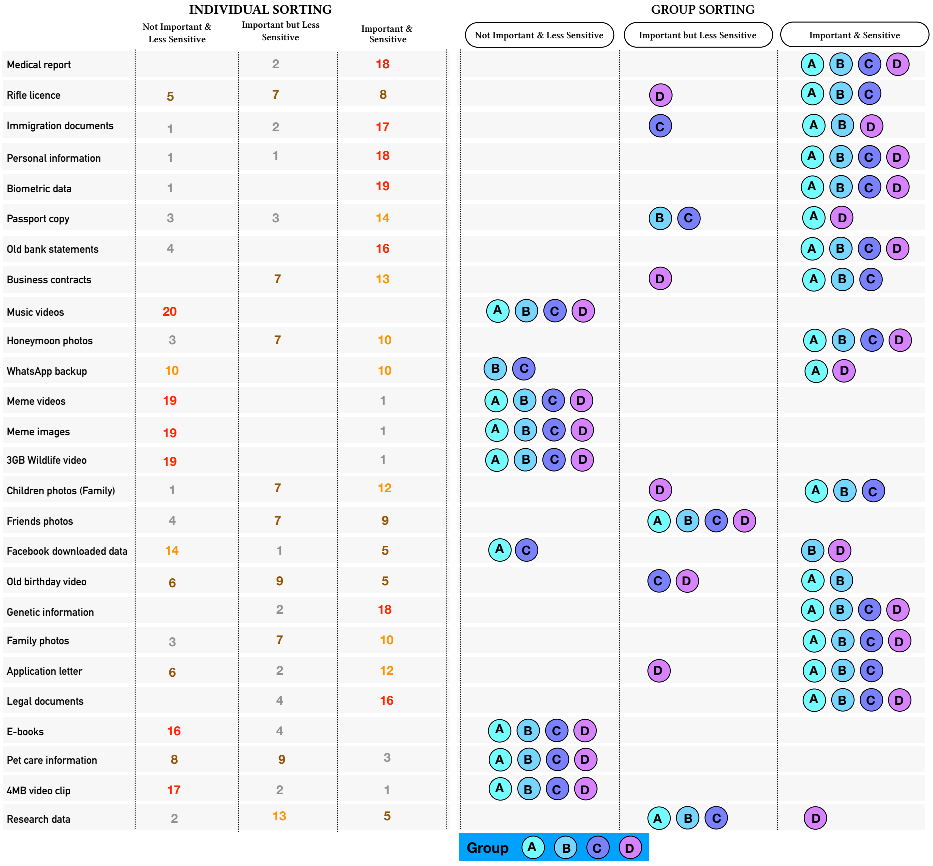
5.1.3 Data Validity

To increase data reliability, the researchers used photos which were taken during the participatory sessions to verify some of the conclusions from the analysis. Triangulation increases the validity of the data through cross verification from two or more sources (Jonsen and Jehn 2009).

The following three sections of this chapter will present the results of all three activities. First, it will focus on the results from PAR1 (the data sorting activity), and then PAR2 (the deletion preference activity) followed by PAR3 (information requirements activity).

5.2 Data sorting Activities

Participants classify their cloud data differently, creating various groups. These groups include: *Sensitive*, *Personal*, *Important*, *Less Important and Less Sensitive*, *Miscellaneous*, and *Important and Sensitive*. Other participants highlighted that important data is the data that they consider useful (utility), hard to get, and which they do not want to lose. *Sensitive* or *Personal* data is the data they consider private and which can be used to identify them. However, for analysis and consistency, similar or related groups were grouped and recorded as one. In the end, only three groups remained: *essential and sensitive*, *Less Important*



	INDIVIDUAL SORTING			GROUP SORTING		
	Not Important & Less Sensitive	Important but Less Sensitive	Important & Sensitive	Not Important & Less Sensitive	Important but Less Sensitive	Important & Sensitive
Medical report		2	18			A B C D
Rifle licence	5	7	8		D	A B C
Immigration documents	1	2	17		C	A B D
Personal information	1	1	18			A B C D
Biometric data	1		19			A B C D
Passport copy	3	3	14		B C	A D
Old bank statements	4		16			A B C D
Business contracts		7	13		D	A B C
Music videos	20			A B C D		
Honeymoon photos	3	7	10			A B C D
WhatsApp backup	10		10	B C		A D
Meme videos	19		1	A B C D		
Meme images	19		1	A B C D		
3GB Wildlife video	19		1	A B C D		
Children photos (Family)	1	7	12		D	A B C
Friends photos	4	7	9		A B C D	
Facebook downloaded data	14	1	5	A C		B D
Old birthday video	6	9	5		C D	A B
Genetic information		2	18			A B C D
Family photos	3	7	10			A B C D
Application letter	6	2	12		D	A B C
Legal documents		4	16			A B C D
E-books	16	4		A B C D		
Pet care information	8	9	3	A B C D		
4MB video clip	17	2	1	A B C D		
Research data	2	13	5		A B C	D

Fig. 5.4 Final data groupings by individuals and groups

and Less Sensitive, and *Important but Less Sensitive*. Fig. 5.4 shows the results of the classifications, individually and in groups.

Important and sensitive: This group contained data that participants classified as private and did not want shared with unauthorised users.
Less important and less sensitive: This is data that is considered less useful, easy to reproduce and less private.
Important but less sensitive: This includes data that participants considered useful and hard to reproduce, but they were happy with other users knowing about it.

5.2.1 Individual Sorting Activity

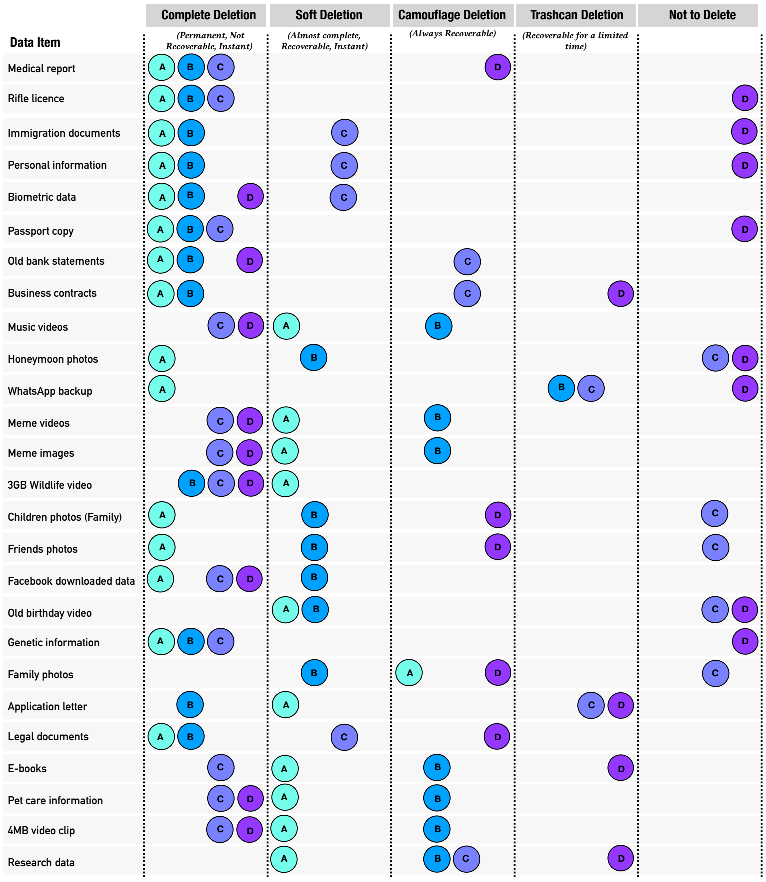
During individual data classification, participants generally categorised data into four or five groups. The most common groups were: personal, sensitive and important, not important and miscellaneous. Other groups included work-related, entertainment, family, and less sensitive and less important. Similar groups (according to their descriptions) were joined together to form three groups described earlier (i.e., *Important and Sensitive*, *Less important and Less sensitive* and *Important but Less sensitive*). Individually, participants had fewer data classified as *Important and Sensitive*, and it was data that related to them personally (e.g., a copy of passport) or involved their families (e.g., family photos). Data related to work was classified as *Important but Less Sensitive*. Music videos were commonly classified as *Less Important and Less Sensitive*.

5.2.2 Group Sorting Activity

There were some noticeable differences between how users categorised their data individually and when they are part of a group. The number of groups generated to categorise data decreased and were generally three or four. After grouping similar groups, the most common groups were *Important and Sensitive* and *Less important and Less Sensitive*. The *Important but Less sensitive* group had fewer items. During groups sorting, multi-media related data except "old birthday video" were all regarded as *Less important and Less Sensitive*. All groups regarded data that is related to "work" such as research data as *Important but Less Sensitive*. The "photos of friends" data type was classified as *Important but Less Sensitive* by all groups.

5.2.3 Individual vs Social Context

Comparing the results between individual and group sorting, the number of data types classified as *Important and Sensitive* increased in the group sorting than it was during individual sorting. Furthermore, most data classified as *Important but Less Sensitive* during individual sorting ended up in *Important and Sensitive* in group settings. The differences in the results may be because, in group settings, users discussed different risks concerning each type of data. For instance, participants discussed different threats that could affect such data or how such data can merely be misused (e.g., such data being used to impersonate the owner). These discussions may have impacted individual users' perception concerning some of the data because all the data types which were discussed in this manner were generally moved to the sensitive and important group. For instance, during individual data sorting task, 50% of the participants classified *WhatsApp backup data* as *Not important and Less sensitive* but during the groups where the risk was discussed participants agreed that such data should be classified as *Sensitive and Important* because WhatsApp data may contain personal and private information.



Data Item	Complete Deletion *(Permanent, Not Recoverable, Instant)*	Soft Deletion *(Almost complete, Recoverable, Instant)*	Camouflage Deletion *(Always Recoverable)*	Trashcan Deletion *(Recoverable for a limited time)*	Not to Delete
Medical report	A B C		D		
Rifle licence	A B C				D
Immigration documents	A B	C			D
Personal information	A B	C			D
Biometric data	A B D	C			
Passport copy	A B C				D
Old bank statements	A B D		C		
Business contracts	A B		C	D	
Music videos	C D	A	B		
Honeymoon photos	A	B			C D
WhatsApp backup	A			B C	D
Meme videos	C D	A	B		
Meme images	C D	A	B		
3GB Wildlife video	B C D	A			
Children photos (Family)	A	B	D		C
Friends photos	A	B	D		C
Facebook downloaded data	A C D	B			
Old birthday video		A B			C D
Genetic information	A B C				D
Family photos		B	A D		C
Application letter	B	A		C D	
Legal documents	A B	C	D		
E-books	C	A	B	D	
Pet care information	C D	A	B		
4MB video clip	C D	A	B		
Research data		A	B C	D	

Fig. 5.5 *PAR2* activity results. Deletion preferences for each data type by the four groups.

5.3 Deletion Preferences Activity

This section discusses the findings of *PAR2*. In summary, these results suggest that deletion preferences are: (1) not always aligned with how participants classified data, (2) different and vary across individuals and groups, (3) often changing (as their cloud usage, file utility, sharing status, social life, storage, and privacy needs change), and (4) complex and multi-dimensional, dependent on deletion context – individual, social and external context. Fig. 5.5 shows the results from the deletion preference activity. The following sections discuss these aspects in detail.

5.3.1 Mind the gap

Using the results from participants' data sorting tasks (*see Section 4.2.1*); it was found that deletion preferences do not always align with how data is classified. Intuitively, one may assume that deletion preferences will align with how they grouped data. For instance, one may assume that data considered private or personal will always be deleted permanently. However, this was the opposite – there are instances when participants would prefer to have that data recoverable or never consider it for deletion.

Group D preferred complete deletion over the data they considered *Not Important and Less sensitive* except when it was *old bank statements*, *biometric data* or *Facebook download data*. This group decided that every other data (i.e., data they classified as *Important and Sensitive*, and *Important and Less Sensitive*) should either ***Not be deleted*** or ***Always be recoverable***. However, unlike group D, group A and B never considered not deleting some files. These differences may have been influenced by the different deletion scenarios and other factors which are discussed later in this chapter. For instance, when deleting *honeymoon photos* (classified as *Sensitive and Important* by all groups), Group A argued that the thought of deleting this kind of photos is highly likely to arise from conflicts (e.g., divorce), or incident where someone would want to avoid trauma hence they would prefer ***Complete and Permanent deletion***, while Group C and D did not consider such context.

There may also be an assumption that data not considered *Important or Less Sensitive* would not require ***Complete and Permanent deletion*** since such type of deletion may come at a cost. Despite this, there were cases where ***Complete and Permanent deletion*** for such data was preferred. For instance, Group C and D preferred ***Complete and Permanent deletion*** for *meme* data which was classified by all groups as *Not Important and Less Sensitive*. They reasoned that such data is unnecessary (i.e., not significant) and may not be needed in the future. However, this is not consistent, Group C did not always want to ***Complete and Permanent deletion*** for data they considered *Unnecessary or Less Important*.

5.3.2 Multiple preferences across individuals and groups

Cloud deletion preferences vary across individuals and groups. These differences could stem from the fact that different users have diverse deletion motivations. Users motivation to delete are privacy-, expertise-, policy- and storage-driven. There is a chance that some participants associate some deletion types with motivations to delete (e.g., some participants may associate deleting highly sensitive to complete deletion). However, since these motivations differ (*see Section 4.2.1*), this may also lead to different preferences.

During this activity, it was also observed that deletion preferences may differ for some data despite it being classified similarly by all the groups. In some cases, these preferences do not overlap. For instance, during the sorting task all the groups classified *friend's photo* as *Important but Less Sensitive*, however, during the deletion preferences task, each group decided to have this type of data deleted differently. This was, however, different when deleting *family photos*, where Group A and D opted for ***Always Recoverable*** while Group B and C chose ***Almost complete*** and ***Not to delete*** respectively, despite all the groups having said the type of data is *Important and Sensitive*. This behaviour suggests that users do not want to lose their *Important and Sensitive* data forever.

This activity also showed that preferences differ across individuals. While this study did not specifically test for individual deletion preferences, some participants expressed that in some cases they would have some data deleted differently than what the group has decided. In some cases, three (3) members of a group would all have different opinions. It is also easy to assume that since each group discussed various threats models towards each type of data in the earlier activity, group members would readily agree on the type of deletion concerning that particular data type. However, this was not the case, and it prompted further discussions on the type of deletion. Participants deciding to have further discussions on the type of deletion shows that users' deletion preferences may always differ despite users' having the same information about data security threats.

5.3.3 Preferences change over time

Deletion preferences change over time, in the same way privacy preferences are known to change (Ayalon and Toch 2013, Khan et al. 2018). During group discussions, participants stated that notable changes in their lives or the use of the cloud might impact their deletion preferences. Participants listed: cloud usage, file utility, social life, storage needs, and privacy needs as factors that could change their preferences.

Cloud usage

Participants mentioned that their use for the cloud is always changing. Some accounts start as personal accounts only but end up later being used for other purposes such as storing work-related data. They mentioned that this in many cases may change how they want such data to be deleted.

File utility

Participant P3, a PhD student from Group C, asserted that at that moment he could not delete his research data, but after ten years he is required by the university policy to delete such data completely. He further noted that after the ten years has elapsed, the data may not have the same importance as before.

Social life

From participants' discussions, when life-change happens, users are likely to change their deletion preferences in their shared folders. For example, when the strength of their offline social links. When these links become weaker or stronger, they may affect how they delete some files, e.g., stronger links may lead to a less destructible type of deletion than weaker social links.

Storage size and needs

Some users' preferences change when they need more storage especially for those that use services that count deleted items as part of user's storage quota. Participants stated that when deleting from such services, one may prefer complete deletion than soft deletion which may leave such data as part of the account.

Privacy needs

Other participants stressed that as their privacy needs change or when data becomes irrelevant their deletion preferences for such data also change. For instance, when private or personal information change, they may no longer be required to entirely or permanently delete files containing such information.

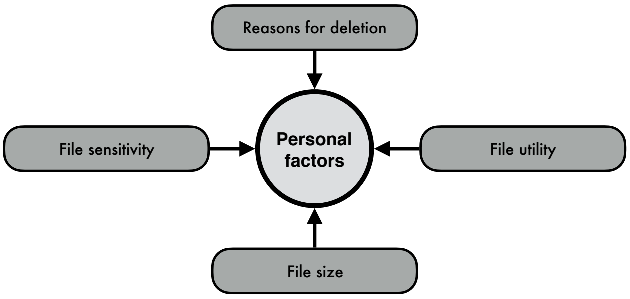


Fig. 5.6 When deleting from their own personal folders, users' deletion preferences are influenced by these factors.

5.3.4 Preferences are complex and multi-dimensional

Another interesting finding is that users' ideal deletion preferences are complex and multi-dimensional. They depend on the user's deletion context – personal, social and external. Personal context includes reason for deletion, the perceived importance of a file (i.e., file utility), perceived sensitivity of the file, and the size of the file (See Fig. 5.6). Social context (i.e., deleting from shared folders) includes the type of the relationship between the members of the shared folder, the number of members, trust, and censorship and authorship. External context includes the service provider a user has subscribed to and the platform type.

These factors are discussed below.

Personal Context

A) Reason for deletion

We found that users consider deletion reasons when choosing their preferred deletion method. For instance, when deleting to prevent others from seeing or having access to a file (e.g., medical report), users desire a deletion process that is instant, complete and allows no recovery. However, when the reason for deletion is to tidy the account (e.g., deleting meme and music videos which they do not consider *Important or Sensitive*), they preferred soft deletion citing that completeness is not essential. However, some users stated that non-essential data might require complete deletion since that data may never be necessary for the future. When users are deleting to free space, they prefer instant and soft deletion when the deleted files do not count towards the storage quota, but complete deletion when they count.

B) File sensitivity

Participants preferred complete deletion for files they consider private and would be unhappy if unauthorised users see them. During the group activities, three groups decided that files like *medical reports*, *passport* and *licence copies*, *biometric data*, and *genetic information* should be deleted entirely without an option to recover. However, one group suggested that they would always want to recover *medical reports* and would not delete *genetic information*, *passport* and *licence copies* as that information is useful and may be hard to get. Participants who had strong feelings about their privacy (e.g., Group A) preferred complete deletion with no recovery for everything they delete.

C) File utility

Participants do not favour deleting files they consider essential (i.e., useful or hard to get), however, if they do so, they want deletion with an option to recover. For instance, during the deletion activity, participants favoured not deleting *honeymoon photos* and *old birthday videos*. Nonetheless, groups suggested that the value of honeymoon photos could change over time hence leading to users desiring deletion that is complete and allows no recovery. Files considered less important such as *meme images* and *videos* attracted complete (2 groups) and soft deletion (2 groups).

D) File size

In some cases, users may consider the file size when deleting, but not on its own; users consider it together with other factors such as its content and the reason for deletion. Smaller files merited soft deletion over complete deletion when they are not considered to contain any private or personal information. However, participants preferred complete deletion for bigger files like *3GB random videos*.

Social Context - (Shared folders)

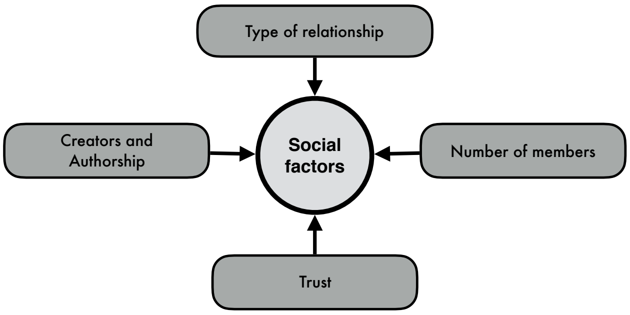


Fig. 5.7 When deleting from a shared folder, participants consider several factors before deciding which type of deletion they desire.

During deletion preferences activity, participants were asked if their deletion preferences would change if they were deleting from shared folders. This was found to be true in some cases, but more importantly, when users delete from shared folders their preferences are affected by the type of relationship between members of the shared folder, the total number of users involved, authorship, and trust. Figure 5.7 shows factors that are considered when deleting from shared folders.

A) Type of relationship

When deleting from shared folders, participants stated that the type of relationship they have with other members plays a significant role in how they delete from shared folders. For instance, if a folder is shared among users who are close (e.g., family members), they would prefer deletion that allows recovery because deletion from such folders is frequent and such groups may include non-tech savvy users. This preference allows them to restore the most valuable files. One member of group A stated that if there is no recovery in family shared folders, some users will delete embarrassing photos of themselves (e.g., while there were children) and such memories may never be recovered. Deletion preferences also change when a folder is shared among colleagues. Similar to authors (Voida et al. 2006, Capra et al. 2014), users prefer not to delete from work-related shared folders, stating that some members of the group might still use the file. When a file is to be deleted from such folders, they desire complete deletion with no recovery. This usually happens when all the members of the folder have authorised such activity. All four groups agreed they would not want to delete legal documents and contracts when shared between colleagues unless they have proof of backups.

B) Number of members

In shared folders, when a smaller number (e.g., less than 5 people) of users share a folder, the folder members tend to allow complete deletion, while they favour deletion that allows recovery in larger groups. For instance, some participants reported that they may delete from folders they share with their friends without contacting them first. This is usually tricky when deleting from a folder shared by a larger group.

C) Trust

Some participants stated that deleting from shared folders depend on the trust that exists among the members of the group. In particular, two forms of trust that are considered in shared folders are (1) trust that a member of the shared folder has the technical skills to perform a technical task correctly in the cloud —confidence in the ability of others, and (2) trust that one person has for another person —type of relation (interpersonal). For groups where members of the shared folder are not trusted to have enough technical skills (e.g., non-tech savvy users), the type of deletion preferred is the one that allows recovery even if it is for a specific time to avoid deletion mistakes. This is also found to be true for files considered important; users would prefer such files to be recoverable. In contrast, the groups suggested that tech-savvy users should be allowed to completely delete files, for instance, when files are irrelevant or have been uploaded by mistake. Regarding interpersonal trust, these results suggest that when interpersonal trust is high (e.g., a couple sharing a folder), users are comfortable with complete deletion. However, there are instances where both aspects of trust are considered together when deleting a file. For instance, tech-savvy partners are more often trusted with complete deletion than non-tech savvy ones.

D) Creators and Authorship

This work also discovered that participants consider who created the shared folder or uploaded/created a file when deciding whether, when and how to delete. All the four groups that took part in the study agreed that they would consider complete deletion when deleting a file that they have uploaded or created themselves. However, they would prefer deletion with an option to recover when deleting a file uploaded by someone else. This behaviour suggests users prefer complete deletion when they are confident about the file they are deleting from the shared folder. For instance, users suggested that they would completely delete their passport copy with no option to recover if they are deleting it from a shared family folder. This may be because they could re-create the copies of their passports.

External Context

These are factors that are beyond file properties and shared folders. They stem from the type of service provider an individual is using, the device or interface one is deleting from and the mental model of cloud deletion a user has.

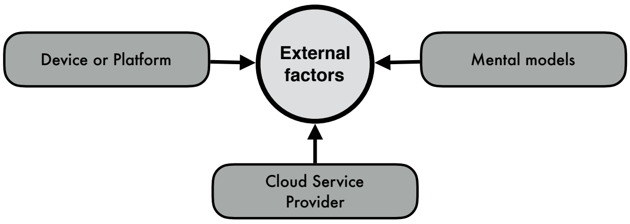


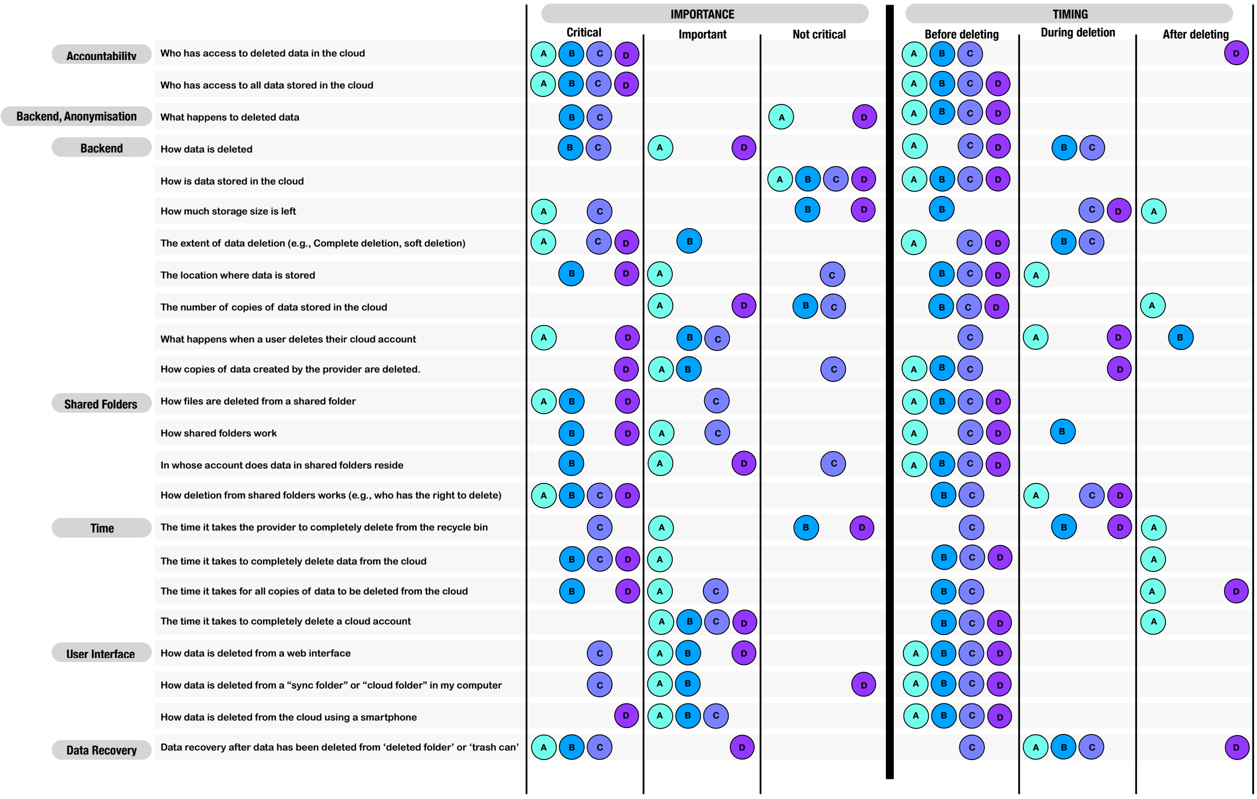
Fig. 5.8 When deleting from their own personal folders, users' deletion preferences are influenced by these factors.

A) Cloud service provider

Some users' deletion preferences are influenced by the service provider they are using. For instance, one participant from group A explained that when deleting to create more space, they would target bigger files (e.g., *3GB video file*), and request for complete deletion rather than have the file moved to deleted item folder or trash can. They reasoned that in some cloud services deleted folder counts towards their account storage quota. Some participants explained that when using a personal service provider for work-related matters, they would desire complete deletion. It is also evident that some preferences depend on the trust that a person has on a provider. For instance, one participant reported that they would want complete deletion when deleting from Dropbox because they do not use it for sharing personal information.

B) Device, platform or interface

Group B and C discussed how the device or platform they use to access the cloud influenced their choice of deletion. They argued that they would not want complete deletion when using a mobile application because one can more easily make mistakes when deleting from a mobile phone than when using a browser or deleting from a sync folder in a computer. Others reported that soft deletion from a computer is better because one can effortlessly recover the files from the trash can or recycle bin if there is a need.



		IMPORTANCE			TIMING		
		Critical	Important	Not critical	Before deleting	During deletion	After deleting
Accountability	Who has access to deleted data in the cloud	A B C D			A B C		D
	Who has access to all data stored in the cloud	A B C D			A B C D		
Backend, Anonymisation	What happens to deleted data	B C		A D	A B C D		
Backend	How data is deleted	B C	A D		A C D	B C	
	How is data stored in the cloud			A B C D	A B C D		
	How much storage size is left	A C		B D	B	C D	A
	The extent of data deletion (e.g., Complete deletion, soft deletion)	A C D	B		A C D	B C	
	The location where data is stored	B D	A	C	B C D	A	
	The number of copies of data stored in the cloud		A D	B C	B C D		A
	What happens when a user deletes their cloud account	A D	B C		C	A D	B
	How copies of data created by the provider are deleted.	D	A B	C	A B C	D	
Shared Folders	How files are deleted from a shared folder	A B D	C		A B C D		
	How shared folders work	B D	A C		A C D	B	
	In whose account does data in shared folders reside	B	A D	C	A B C D		
	How deletion from shared folders works (e.g., who has the right to delete)	A B C D			B C	A C D	
Time	The time it takes the provider to completely delete from the recycle bin	C	A	B D	C	B D	A
	The time it takes to completely delete data from the cloud	B C D	A		B C D		A
	The time it takes for all copies of data to be deleted from the cloud	B D	A C		B C		A D
	The time it takes to completely delete a cloud account		A B C D		B C D		A
User Interface	How data is deleted from a web interface	C	A B D		A B C D		
	How data is deleted from a “sync folder” or “cloud folder” in my computer	C	A B	D	A B C D		
	How data is deleted from the cloud using a smartphone	D	A B C		A B C D		
Data Recovery	Data recovery after data has been deleted from ‘deleted folder’ or ‘trash can’	A B C	D		C	A B C	D

Fig. 5.9 Deletion information preferences: Importance versus Timing

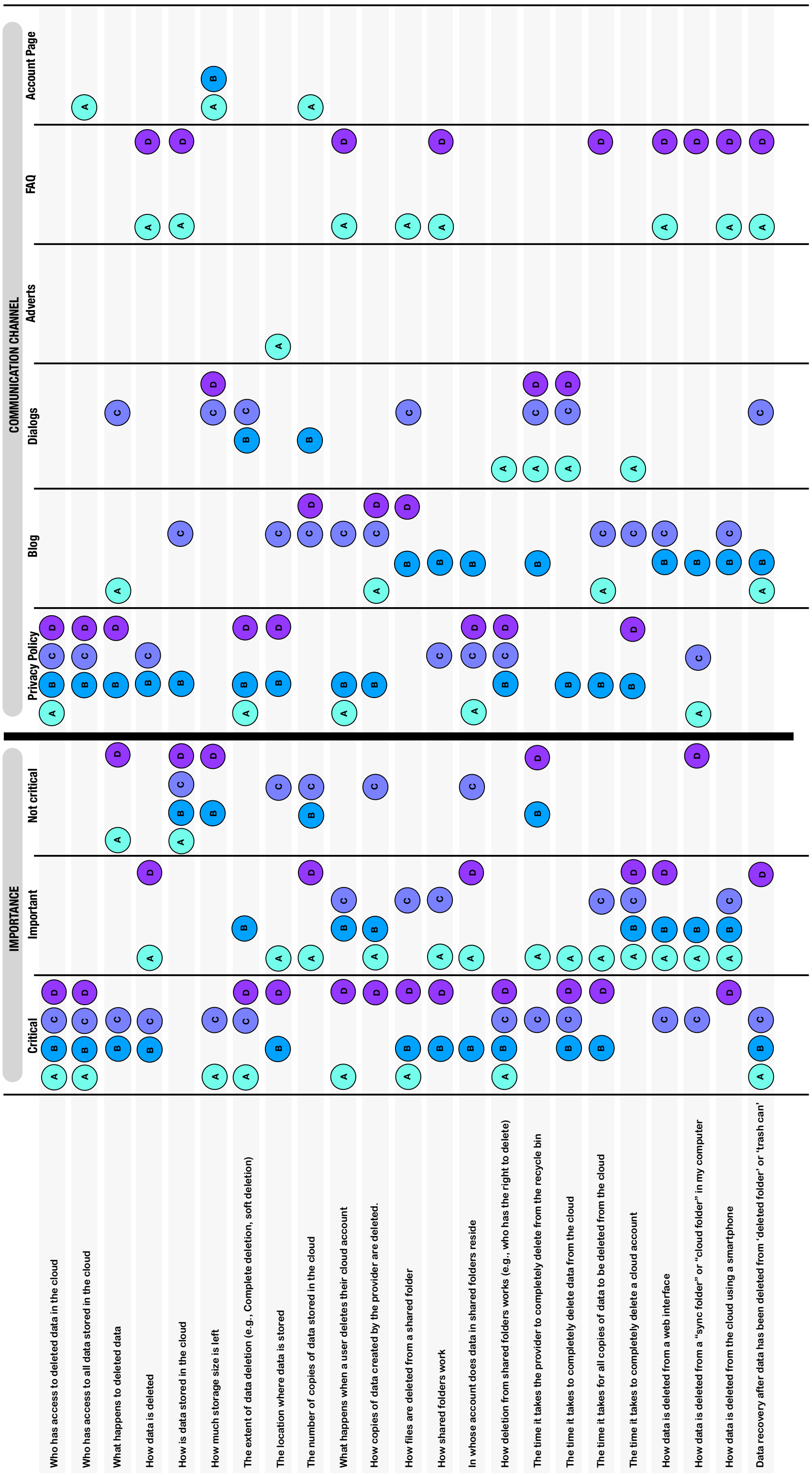


Fig. 5.10 Deletion information preferences: Importance versus Communication Channel

5.4 Deletion Information Requirements

This section discusses the findings of *PAR3*. It first presents the results of the activities and then the lessons from the activity.

5.4.1 Activity Overview

Figure 5.9 and 5.10 shows the overall results of *PAR 3*; Fig. 5.9 shows information type importance against timing while Fig. 5.10 shows information type importance against the communication channel used. The following sections summarises these results.

Group A

During information preference task, group A had 13/23 items classified as important information to know regarding deletion while only 2/23 items were regarded as information that is not critical. Regarding the time when should such information be made available to them, Group A categorised types of information as follows. They preferred 11/23 information types to be made available before they use the cloud service or before deleting. Then, 6/23 information types were grouped under during and after deletion. Concerning communication channel, 8/23 items were placed under privacy policies, 5/23 items under privacy policies, 4/23 items under blogs and dialogs, 3/23 items under account page and single item under adverts.

Group B

Group B classified 14/23 info types as critical, 5/23 as important and 4/23 as information not important to know concerning cloud deletion. With regards to timing, they preferred knowing 14/23 info types before deleting or using the cloud service and 7/23 info types while deleting. A single item was put under after deletion. Group B preferred 12 info types to be made available in privacy policies, 8/23 in blogs, 2/23 in dialogs and 1/23 in the account home page. They preferred no information about deletion in adverts and FAQs. Group B preferred the most critical information (9/12) to be available in privacy policies.

Group C

Group C categorised 12/23 as critical information, 6/23 as important and 5/23 as not critical information. Regarding when the information should be made available, they decided 13/23 info types were useful to know before deleting or signing up to use the service. They also stated that some of the information was necessary to know before and during deletion. As a result, 10/23 info types overlapped between before and during. They did not put any info types in the after-deletion group. Group C preferred 7/23 items to be made available in the privacy policies while blogs and dialogs had 8 items each. Most of the info types (6/12) considered critical were grouped under dialogs while least 1/12 was under blog page.

Group D

Group D grouped 12/23 items as critical information, 5/23 as not critical and the rest as just important. Out of 12 items considered critical, Group D suggested that 7 of those info types were important or necessary to know before signing up or attempting to delete. They preferred 2/12 to be made available after deleting and 3/12 during deletion. Group D grouped 8/23 items under privacy policies, and 5/8 of those were from the group of critical information. The rest of the critical information was classified as follows: 4 items under FAQs, 2 items in blogs and 1 item under pop-up dialogs. Most information (9/23) were grouped under FAQ while blogs and pop-dialogs had 3 items each.

To summarise, participants preferred having information about deletion through different communication channels. However, they preferred and expected to have more information about deletion in the privacy policies (*32 info items*) than on adverts (*1 item*). Only group A preferred some information (*1/23*) to be made available in the advertisement (i.e., information about the location where data is stored). Group A reasoned that such important information is important concerning their decision to use a particular cloud service provider. In contrast,

Table 5.4 A summary of how each group classified information about deletion with regards to the communication channel and the time they would prefer to have such information.

Group	**Importance**	**Communication Channel**						**Timing**		
		Privacy	Blogs	Dialogs	Ads	FAQ	Acc. Page	Before	During	After
Group A	Critical	3	1	1	-	3	2	4	3	1
	Important	2	2	3	1	4	1	5	3	5
	Not critical	-	1	-	-	1	-	2	-	-
Group B	Critical	9	4	1	-	-	-	10	4	-
	Important	2	3	-	-	-	-	2	2	1
	Not critical	1	1	1	-	-	1	2	2	-
Group C	Critical	5	1	6	-	-	-	12	10	-
	Important	1	3	2	-	-	-	6	-	-
	Not critical	1	4	-	-	-	-	5	-	-
Group D	Critical	5	2	1	-	4	-	7	3	2
	Important	2	1	-	-	3	-	4	1	1
	Not critical	1	-	2	-	2	-	2	3	-

other groups stated that they preferred privacy policies because they are used to having valuable information on privacy policies and having important information on adverts may lead to many people missing the information. Moreover, participants (all groups) expected to find information about who has access to their stored or deleted data in privacy policies but preferred to have information about how the deletion process is completed in both privacy policies, FAQ and blog pages.

Participants preferred to have more information about deletion before signing up or deleting (*A = 11, B =14, C=23, D=13*) from the cloud than during and after deleting. This is mainly the information they considered critical. However, there are some instances where groups preferred to know less critical information before deleting, this occurred 23 times (*A=4, B=4, C=10, D=5*).

Most groups preferred to be given information that they consider critical such as who has the right to delete from a shared folder and whether the data will be deleted completely during deletion. With regards to information about deleting, groups stated that they would want to know how much storage is left, how long their data would be removed entirely from the cloud and who has access to their deleted information.

Regarding retention time, participants preferred to know about retention time before signing up (*Group B, C, D*) and immediately after deleting (*Group A*). They reasoned this would influence what they store in the cloud and being reminded of the retention time after deletion will make them aware of how long their data will still be in the cloud.

Before signing up or deleting, participants (i.e., all the groups) stated that they preferred having more information about who and how deleted data is handled. During the process of deletion, participants stated that it was important for them to have information that will guide them in making decisions concerning deletion (e.g., whether data will be deleted completely). After deletion, they prefer to have information about the consequences of their action (e.g., what happens to deleted data).

5.4.2 Lessons from the PAR 3

In summary, these results suggests that (1) cloud users are not aware of what deletion information is available in the cloud, (2) most information about deletion is not relevant to users' needs, (3) deletion information is not available when users need it (not provided at the right time), and (4) essential information is limited to privacy policies.

Limited information about deletion

This study revealed that most cloud users are not aware of what information about cloud deletion is available or where they could find it. For instance, many participants discussed how they were not sure whether some of the information (e.g., how long it takes to remove all copies of data from the cloud) we used during our study was indeed available. They reported not having seen some of the information we used in the study but hoped it is available. One such example is information about data recovery after deleting from the trash can or deleted item folder. Participants were also observed to have assumptions about where some information could be found, for instance, most participants assumed that most information is contained in the privacy policies.

Precise and relevant information

While participants appreciate and acknowledge the scarcity and the importance of information about deletion, they explained that too much information can be overwhelming and may cause users not to be interested. They suggested that providers should only share information that is essential and needed for making informed decisions about deletion. During the first task of this activity, all four groups deemed information about how data is stored in the cloud not necessary for deleting. However, all the groups noted that information about how deletion works in shared folders is relevant and vital information.

Timing

Participants want relevant information presented to them at the right time — the point of time when they need it, suggesting that some information should be available before they sign up for service, during usage, and some while after deleting. This study indicates that some users prefer seeing information about ‘how’ and ‘who’ handles data before they sign up for a service — not just in privacy policies. For instance, all four groups suggested that information about who has access to their cloud data should be made available before signing up for a service. During the process of deleting, participants wanted to see information that will assist their decisions while deleting (e.g., a clear message informing them how long before a file will be completely removed from the cloud). They also prefer to be updated about their actions immediately after they have deleted from the cloud, for instance, being informed about the status or the consequences of their action like, how long the file will remain in the deleted folder. Moreover, the group task revealed that participants desire to have any information related to time (e.g., the time it takes to delete from recycle bin completely) after the user has deleted.

Beyond privacy policies

While privacy policies usually have more information about the service, this study reveals that users do not always expect or want to have all the information in privacy policies. The activity on deletion information (PAR3) revealed that participants prefer that some information should be made available in other places such as provider's blog posts, account dashboard, adverts, frequently asked questions (FAQs), and pop-ups. Participants desire to have information on how deletion is achieved or completed (technical information) in blog pages and FAQs. For example, all four groups suggested that information on how to delete using a web interface and a mobile application should be in blogs and FAQs. Participants desired to have information about 'who' and 'what' to be made available in privacy policies while they preferred to have information about the time a deletion process take to complete through other channels like alerts (e.g., the time it takes to delete from the cloud completely). Fourteen (14) respondents to the questionnaire stated that they would expect relevant information to be easily available on their account page and notices, while five (5) said they hope such information to be made available in the privacy policies and FAQs.

5.5 Design Implications

Based on the findings above, the following sections below discuss the implications of these results and also propose several guidelines for developers of cloud storage platforms for improved controls and information for deletion.

5.5.1 Deletion Controls

Informed by these findings, the next-generation deletion controls or cloud interfaces should have the following properties:

Intelligent and Personalized

Just like sharing preferences (Olson et al. 2005, Mazurek et al. 2010, Sleeper et al. 2016), deletion preferences are dependent on many factors like file attributes, relationships between owners, and mental models. The results above reveal that preferences are ever changing depending on the context and time. These findings suggest that deletion needs cannot be generalised and there is no one-size-fits-all solution for meeting the deletion preferences of users.

It is also clear that users preferred complete deletion for more files than just the ones they considered private. For instance, sometimes users preferred complete deletion for files they considered less important or less sensitive.

Intuitively one may assume users would prefer complete deletion for data perceived essential and private. However, users consider other factors such as the importance of the data and how easy it is to get or reproduce. These mismatches suggest that even privacy concerned users may not always choose the deletion mechanism that one would expect to reflect their concerns.

Since the deletion preferences are many, different, and dependent on numerous factors, this calls for intelligent deletion controls that can adapt and cater to different users. Researchers may use AI techniques (such as machine learning) to investigate what deletion controls can better meet users' deletion needs and actively assist them in their deletion tasks. AI can also be used to learn about users' data classifications, for instance, automatically clustering data together into types of data (as users were observed doing themselves) with similar deletion preferences. Moreover, learning about the deletion preferences concerning particular data for individuals may help to establish suitable default settings for them.

Interface-aware

Users use different interfaces (i.e., web interfaces, mobile apps, and sync folders) to access cloud storage depending on the need, sharing audience, cost and accessibility. When developing a deletion mechanism, designers should account for these factors. For instance, users may be willing to incur the cost of setting their deletion preferences, or deleting when using the web interface rather than the mobile device. Furthermore, the previous chapter (Section 4.2.4) has shown that users switch between devices when facing challenges when deleting, but relied on one device for simply deleting. Designers should, therefore, consider whether one interface can contain all the necessary features required to offer deletion preferences or whether some features may be excluded. And, also consider how this can be communicated to users.

Layered

This study shows that the level of detail of deletion preferences is different across users, some users would be interested in having fine-grained controls while others may find them demanding. Therefore, designers should develop deletion mechanisms that are layered; offering simple fine-grained deletion controls for those who are interested in defining their preferences and coarse-grained ones for those who might find the task demanding or wish to have broader deletion preferences. Fine-grained controls could focus on file properties and sharing context. They should account for costs and other use cases of the cloud such as shared folders and files, devices synced with the cloud (e.g., complete deletion for all files deleted through the web interface). Coarse-grained controls may include default settings that are clearly communicated to users.

Retrospective

The above results suggest that deletion preferences may change over time. This may mean users might have certain requirements, or wish to delete old data or data that have not been modified over a long period. Prior work of Khan et al. (2018) has explored retrospective data management in the cloud and found that 83% of participants wanted to delete at least one of their old files. However, this study reveals that users may have preferences on how such data may be deleted. This highlights the need for retrospective deletion mechanisms that give users the chance to specify how data should be deleted when it has become old. Users could be allowed to specify how old data should be handled by the cloud provider – state when and how old files should be deleted. For instance, notify the user of files older than a year and request action on how to delete such files.

Multiuser-aware

In the context of shared folders, participants prefer not deleting files in shared folders unless they have created or authored them. Designers should therefore consider implementing deletion controls that take into account the multi-user nature of shared folders, so that members of a shared folder can agree for a folder or file to be deleted. For instance, a user may specify during file upload whether others can delete the file. This solution would reduce conflicts that exist within shared folders regarding the decision whether to delete or not and what type of deletion (e.g. recoverable or not) should be used. However, data sorting task showed that individuals preferences adapted to social context easily, and that does not remove the possibility of intentional or accidental conflicts.

5.5.2 Deletion Information

Relevant information at the right time

Results from a study by Murillo et al. (2018) suggest that information about deletion should cover six topics: back-end, time, backup, derived information, anonymisation and shared copies. However, the authors do not study where such information could be presented to users. This work shows that this information should not only be constrained to privacy policies but be made available through other channels like blogs, FAQ, and dialogues. This study also confirms that communicating these concepts to users at the right time may improve understanding of deletion – influencing mental models. For instance, explaining retention period (i.e., time) immediately after users delete may make them realise that data is not entirely deleted but still available in the cloud. Likewise, concepts such as anonymisation may be included in adverts to signal to users that the deletion process may start with anonymisation.

It is also clear from the study that users seek help (i.e., look for information on deletion) in order to delete (i.e., accomplish a task) but not to improve their knowledge or know where such information is located. Therefore, information about deletion should be divided into two groups: primary information and secondary information. Primary information (e.g., retention period after deletion) should contain information that is essential for deletion and is needed by users to make decisions on deletion while secondary information (e.g., how copies created by the cloud provider are deleted) should focus on information that is not contextualised but may be necessary for users. Critical primary information should be easily accessible when deleting while secondary information should be made available to users when they need or request for it. These findings suggest that research should not only focus on improving privacy related information on policies only but should study the impact of information on other channels.

Deletion status and summaries

Users prefer to be informed about the status of their deletion. This suggests that they want assurance of their actions. One way of informing users about the consequences of their actions is to add timers to their cloud trash cans or deleted folders. Each deleted file could have a timer showing when it is going to be removed from the bin. This would enable users to see which files are about to expire and which ones they can still recover. For instance, a user deleting for privacy reasons will be able to see that their file has not been deleted entirely, and time left before it is removed entirely from the cloud.

Users would also like to have deletion summaries. A hypothetical question concerning deletion log in the survey received much positive feedback. Out of the 20 participants, only two stated that they would not wish to receive a summarised report about their data deletion. Nine (9) stated that they would wish to receive these summaries monthly. This feature would allow users to see their deletion records and help them audit their accounts, allowing them to reverse some of their decisions, for instance, a user may have a chance to recover a file from their trash can before its retention period elapses. Developers should produce these summaries and enable users to choose how frequent they would want to receive them or make them available on-demand.

5.5.3 Limitations

One disadvantage of participatory action research is domination. It is possible that one or two participants may have dominated the group and overruled others. Domination may have limited the ability of others to express their views freely. To mitigate this, the researcher conducting the fieldwork stepped into the discussions and encouraged quiet group members to share their views.

This study was exploratory and mainly qualitative in nature. While the sample was varied and roughly balanced across different demographic variables like gender, education and employment, other variables like age were varied but less balanced (18-45). Additional studies with a larger and more diverse cloud user population should be considered, though getting older participants may be challenging as they tend to engage less with new technologies (Olson et al. 2011). In addition, the hypothesis raised over the preferences on deletion and information about deletion should be quantitatively tested in a subsequent confirmatory study.

During some exercises, few participants showed signs that might have led to task exhaustion. To mitigate this and the negative impact this might have on the results, e.g., participants rushing through the tasks, breaks were introduced in between the tasks.

5.5.4 Summary

This chapter has provided empirical evidence on users' cloud deletion preferences and deletion information requirements. Firstly, it has shown different deletion preferences that users desire when deleting from the cloud. Secondly, it has presented different factors that users consider when deciding which type of deletion to employ. These factors range from file utility to the type of relationship and trust that exists between users sharing a folder in the cloud. Thirdly, with regards to deletion information, this chapter has shown how important timing, location and content are when it comes to providing information about deletion. It also examines how dissemination of information about deletion can be improved to better inform users when deleting from the cloud. The next chapter uses the findings of this chapter and the previous chapters to inform deletion requirements and design principles for assured deletion in the cloud. it focuses on how deletion mechanisms can be improved to meet these preferences.

Chapter 6

Conceptual Framework for the Design of Usable Assured Deletion in the Cloud

The chapter aims to bring together the findings of the studies presented in Chapter 3, 4 and 5 of this work. Based on these findings, this chapter first proposes a conceptual framework for the design of usable assured deletion and then presents user requirements and principles for usable assured deletion. The theoretical contribution of this framework lies in its ontological approach, making *usable assured deletion* a constitutive concept of using cloud storage in a safer and usable manner. This framework aims to clarify concepts related to assured deletion and to identify vital interrelations among them. The user requirements presented herein support the proposed conceptual framework, however, they do not serve as a solution but are instead building blocks towards a more complex solution while the principles may be used as a guide when implementing cloud systems that satisfy the formulated user requirements.

Chapter 2 this work has shown that existing literature provides neither a coherent view of cloud deletion practices and needs of users nor a roadmap towards mechanisms that satisfy them. The main problem at hand is that cloud users' deletion needs, and challenges are not well understood. The understanding of assured deletion in the cloud is, therefore of critical importance for the design and implementation of cloud systems that are capable of meeting the deletion needs of users. Moreover, an improved theoretical understanding of usable assured deletion allows for research conclusions that are more relevant to practitioners. This chapter aims to provide this understanding. In summary, the key contributions of this chapter are as follows:

- A conceptual framework for the design of usable assured deletion in the cloud.
- Six (6) assured deletion user requirements, and seven (7) design principles for usable assured deletion.

This chapter is structured as follows. Section 6.1 discusses the methodology. while Section 6.2 presents the proposed conceptual model for the design of usable assured deletion. Section 6.3.1 presents the user requirements for cloud deletion, and a comparison between these requirements and the cloud service provider requirements presented in Chapter 3. Section 6.3.3 presents the seven principles of usable assured deletion. Section 6.4 concludes this chapter.

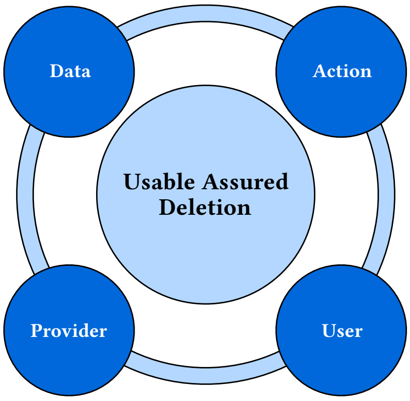


Fig. 6.1 Four main concepts for the conceptual framework for the design of usable assured deletion

6.1 Methodology

The conceptual framework was developed using a conceptual framework analysis approach (Jabareen 2009). Conceptual framework analysis aims to generate and trace the phenomenon's major concepts, which together establish its theoretical framework. The eight (8) stages of the conceptual framework analysis include: (1) Mapping selected data sources; (2) Extensive reading and categorising of the selected data; (3) identifying and naming concepts; (4) deconstructing and categorising concepts, (5) Integrating concepts; (6) Synthesis, resynthesise, and making it all make sense; (7) validating the conceptual framework, and (8) rethinking the framework.

During the first phase of the conceptual framework analysis approach, the findings of the studies presented in Chapter 3, 4 and 5 of this thesis were reviewed. In the second phase, key ideas and themes from all the three chapters were identified and categorised by each chapter and order of importance. The next step was to identify concepts from the ideas and themes categorised in the second phase. As part of the fourth phase, each concept was deconstructed, to identify its main attributes and role. These concepts were then categorised according to their features, scope and role. In phase 5, concepts with similarities were identified and grouped. At this phase, the number of concepts reduced significantly; four main concepts were identified (see Fig. 6.1) and 12 sub-concepts which supported them. Then, a synthesis of the identified concepts was developed to produce a theoretical framework. This process was repeated several times to produce a framework that made sense. To validate the framework, the primary researcher presented and discussed the conceptual framework with other researchers. Using this feedback, the conceptual framework was reviewed and refined (combing some of the concepts) to make sense.

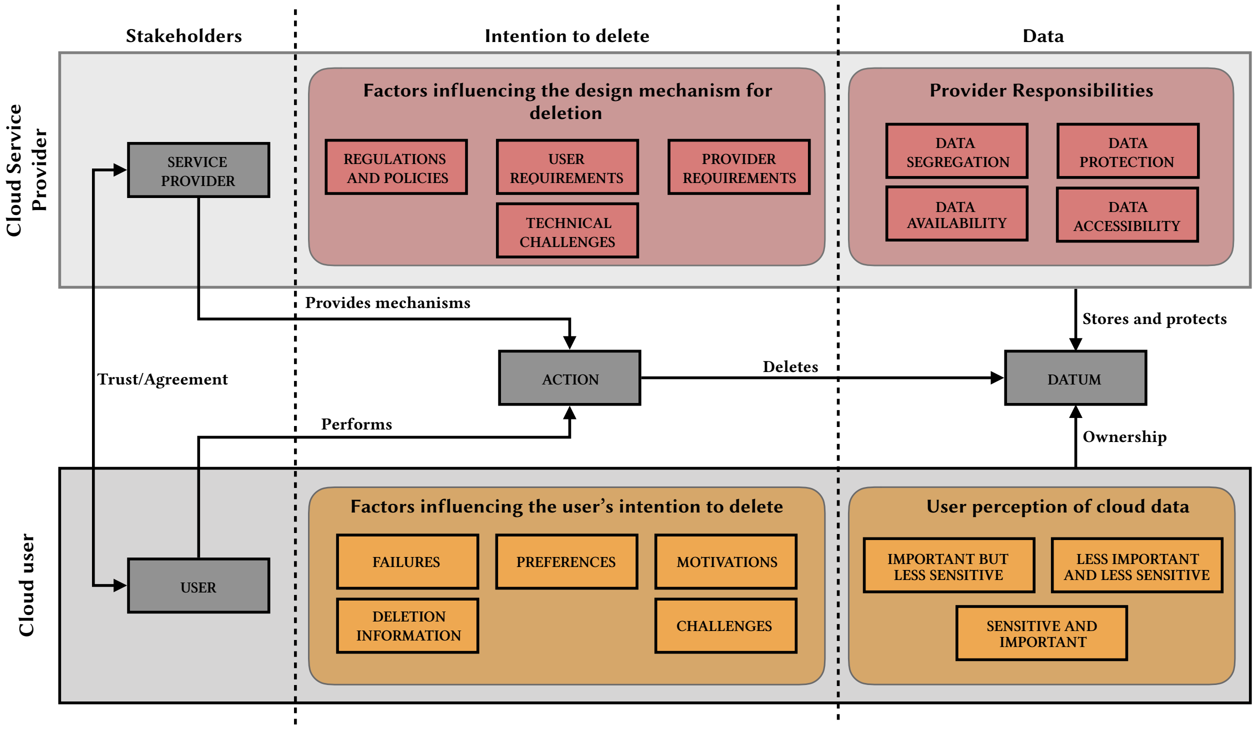


Fig. 6.2 Conceptual Framework for the design of Usable Assured Cloud Deletion

6.2 Conceptual Framework

The proposed conceptual framework of usable assured deletion is composed of four interrelated concepts (i.e., User, Provider, Action and Data) of cloud assured deletion, where each concept has a unique role in the framework. The user interacts with the service provider to delete "action" their data from the cloud. Figure 6.2 presents a detailed conceptual framework for the design of usable assured deletion, showing how these concepts are related to each other. An overview of this framework is provided below.

In order to design usable deletion mechanisms for the cloud, it is important to understand the stakeholders and their role within the cloud ecosystem. The work presented in this thesis has shown that cloud *users* and *service providers* (user and provider as defined in the introduction 1.7) play a vital role in how usable assured deletion can be offered. Chapter 3 has looked at assured deletion from the provider's perspective, showing that assured deletion is important for providers for the following reasons: to comply with regulations and to meet the deletion needs of users. From the users' perspective, assured deletion offers them trust that their data is well managed and safe with the provider. As the main stakeholders in the cloud ecosystem, the decisions, intentions, roles and requirements of both parties influence how usable deletion mechanisms should be designed. The following paragraphs explain the role of users and service providers in the framework.

The concept of ***user*** is one of the pragmatic foundations of usable assured deletion framework. It represents actors who store data in the cloud and have an interest in deleting data from the cloud in a *usable* way. Users trust that the provider will store and protect their data. Most importantly, the provider will delete their data when they request to have such data deleted. Users have to perform an action through the mechanisms provided by the service provider to delete. They are the primary consumers of the mechanisms provided by the cloud service provider. Therefore, it is essential to understand their deletion practices and preferences, that is, their motivations to delete, how they delete, the challenges they face while deleting and how they prefer their data to be deleted. Chapter 4 and 5 of this work bring these practices into perspective. These chapters find that users' motivation to delete are storage-, policy-, privacy- and expertise driven. They prefer destructive methods of deletion (e.g., instant permanent deletion) over soft deletion (e.g., deletion that allows recovery) when deleting data that they consider sensitive and can easily reproduce. Given that the current mechanisms for deletion work by initially moving deleted data to the *Deleted Items Folder* or *Trash can* before being completely removed from the platform, new designs could give users the choice to choose how data is deleted when they are deleting.

Service provider as a concept represents actors who are responsible for storing or managing users' data. Users and service providers have an agreement that both parties wish to honour. Service providers usually agree to store and protect users' data while users may agree to use the provider's services for free in exchange for advertisements . In this framework, service providers offer users with storage facilities and the mechanisms (i.e., applications and interfaces) to delete data when they no longer have use for it. Providers also determine how data will be deleted (i.e., type of deletion) from their platform. When users request for data to be deleted, providers receive and handle these requests, and they are expected to honour these requests by deleting data from their platform.

Service providers offer users with mechanisms (i.e., applications and interfaces) for deletion and they also determine how such data is deleted (i.e., type of deletion). When users request for data to be deleted, providers receive and handle the request by deleting data from their platform. The findings presented in Chapter 5.4 revealed that information importance, the channel of communication and the timing of delivering information about deletion is important as it may influence users deletion practices and decisions. These factors should be considered when deciding where to have information about cloud deletion. Understanding where users search or expect to find information about deletion may help providers improve their help mechanisms. Chapter 4.2.2 revealed the struggles that users face while deleting and how they recover from such challenges. It is important to understand what communication channels users use when seeking help information about cloud deletion. The findings in Chapter 5.4 give insights on what information users consider essential and where the service providers may make it available for users.

The concept ***data*** represents the information that users store in the cloud storage provided by the service provider. Users own the data, and the service providers store these data. Data stored in the cloud may include information that users perceive *Important and Sensitive*, *Important but less sensitive* and *Less Important and Less Sensitive*. Chapter 5 gives insights on how users' perception of cloud data may influence their deletion preferences. The OpenStack analysis conducted as part of this thesis revealed that data stored in the cloud goes through a lot of virtualised layers which may make it difficult to reach for deletion. This may introduce a lot of challenges for the provider who is seeking to completely delete data and comply with the law (technical assured deletion challenges discussed in Chapter 3.2.1). Data provenance reveals the need for deletion mechanisms that reach different layers of the cloud storage architecture and provide deletion. Usable deletion mechanisms should reach or be able to remove data from all the places it is stored or have been.

Action represents the *deletion operation* performed by users when they desire to have their data removed from the cloud. Users make these requests through deletion mechanisms provided by the service provider. A deletion request is complete when the provider has deleted the data that the user requested to be deleted. It is essential that the cloud provider aims to meet the deletion requirements (discussed in Chapter 3) and the user requirements and principles (discussed in Section 6.3.1 of this chapter). Satisfying these requirements will help providers comply with regulations as well as meeting users' deletion needs. With regards to users, user deletion practices and preferences are important and should be given most attention. The user study presented in Chapter 5 revealed different user deletion preferences. Assured deletion mechanisms should aim to provide users with the ability to exercise their deletion preferences when deleting from the cloud. Moreover, different motivations for deletion as discussed in Chapter 4.2.1 may merit different deletion type. The differences in users' deletion preferences while deleting from individually owned folders and shared folders show the complexity of providing and meeting the deletion needs of cloud users. Designs should consider these factors. Participants of the study presented in the Chapter 4.2.3 mentioned various user interfaces related issue they face when attempting to delete from different platforms, e.g., mobile applications. Ensuring that users can perform deletion without facing software failures is vital for the design of usable assured deletion mechanisms. Information about deletion also plays a vital role concerning the action, it could provide users with sufficient information on the consequences of their deletion action.

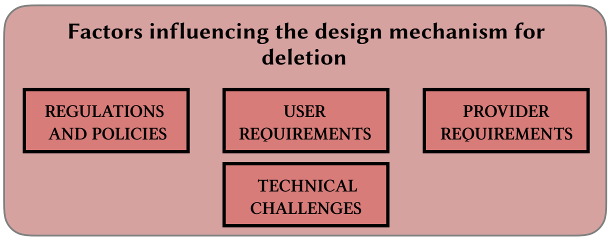


Fig. 6.3 Factors that influences the design mechanisms for assured deletion

The following section expands on the factors that need to be considered when implementing mechanisms that offer assured deletion (see Fig. 6.3). It specifically discusses the user requirements and principles for usable assured deletion which have not been previously discussed in this thesis. These requirements were formulated using the findings of Chapter 4 and 5 of this thesis. The relationship between these factors (i.e., regulations, provider requirements and technical challenges) have been discussed extensively in Chapter 3, while the following section (Section 6.3) focuses on the user requirements and how they relate to other factors. Table 3.4 in particular maps the key cloud features, technical challenges and the requirements from both an honest provider and dishonest provider scenarios. It shows what key cloud features pose challenges to assured deletion and the provider requirements challenged. In Section 6.3.1, after formulating the user requirements, a comparison between the user requirements and the provider requirements is made. This comparison aims to highlight which requirements are similar and which ones are peculiar. Fig. 6.4 summarises this comparison. In section 6.3.3, this thesis presents the principles of usable assured deletion; each principle is mapped to its relevant technical constraints (previously discussed in Chapter 3) that poses challenges to usable assured deletion. This mapping is to show how technical challenges may hinder the satisfaction of the user requirements formulated below. Lastly, to elaborate more on the relationship between the user requirements and principles, Table 6.7 provides a mapping of the user requirements and the relevant principles that should be considered in order to satisfy them.

6.3 User Requirements and Principles for Usable Assured Deletion

When providing mechanisms for assured deletion, the provider must aim to meet and satisfy both their deletion requirements and users' deletion requirements. Chapter 3 of this thesis only formulated the requirements for providers but did not include user requirements. This chapter presents the user deletion requirements and principles for usable assured deletion. These are requirements that need to be satisfied by the provider in order to offer usable assured deletion in the cloud, and the principles act as guidelines that should be followed when implementing the principles. Chapter 4 and 5 highlighted valuable insights about deletion practices and preferences of cloud users. Using these insights, this chapter proposes a set of requirements and principles for the implementation of usable assured deletion mechanism.

6.3.1 User Requirements for Usable Assured Deletion

Complete deletion: The cloud system should enable users to delete data from the cloud entirely. Data deemed deleted should not be recoverable after the retention period has elapsed. (*Supporting Evidence:* ***Section 3.2****,* ***Section 4.2.1****,* ***Section 5.3.1****, and* ***Section 5.3.2***)

Relevant, informative and immediate information and feedback: The cloud system should provide users with information and feedback that is relevant, easy to read, informative and available at the appropriate time. This information should be easily accessible through the system and feedback should be simple and informative to users. (*Supporting Evidence:* ***Section 4.2.5*** *and* ***Section 5.4***)

Choose type of deletion: Users should be able to choose the type of deletion they desire, and for how long deleted data should be kept in the cloud system before it is removed entirely. (*Supporting Evidence:* ***Section 4.2.1*** *and* ***Section 5.3.1***)

Help: The cloud system should offer clear information on how to complete and recover from deletion failures. Users should not be required to leave the system for help. A help guide should be easily accessible and clear to follow. (*Supporting Evidence:* ***Section 4.2.2****,* ***Section 4.2.4****,* ***Section 4.2.5****, and* ***Section 5.4***)

Error handling: The system should allow users to recover from mistakes and errors. This means that users should have the ability to return the system to a normal working state after mistakes or system errors. (*Supporting Evidence:* ***Section 3.2***, ***Section 4.2.3*** *and* ***Section 4.2.4***)

Usable interface and controls: The cloud system should have an easy to use deletion controls. Deleting from the cloud should be simple and easy to complete.(*Supporting Evidence:* ***Section 3.2***, ***Section 4.2.2***, ***Section 4.2.3*** *and* ***Section 5.3.3***)

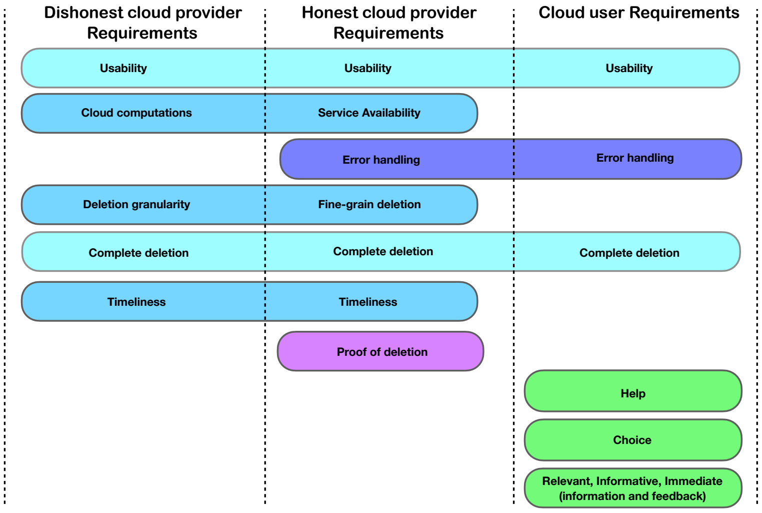


Fig. 6.4 Requirements mapping.

6.3.2 Service Provider Requirements vs User Requirements for Assured Deletion

The following section discusses and compares the above user requirements and the cloud provider requirements identified in Chapter 3. Figure 6.4 provides a summary of how all these requirements are related to each other. This summary was created by comparing all the formulated cloud service provider requirements from Chapter 3 and the user requirements mentioned above. After comparing these requirements, similar requirements or requirements aiming to achieve a similar goal were synthesised together. This resulted in figure 6.4 above.

Cloud Service Provider and User Requirements

Usability and *complete deletion* were common between all sets of requirements. This finding suggests that in all circumstances and regardless of the scenario, the cloud system must be usable and offer complete deletion.

Usability: The system should offer usable deletion, and deletion tasks, user interface and navigation should be easy to learn and complete. Users should perceive the system as easy to use.

Complete deletion: The system should completely delete data upon request. Deleted data should not be accessible after deletion has completed.

Honest Cloud Provider and User Requirements

Focusing on the honest cloud provider requirements and user requirements, it is clear that, in addition to above-mentioned requirements, the ability of the system to handle system and user errors is essential for users and service providers willing to offer assured deletion.

Error handling: Both users and the system are likely to experience errors, so the system that is designed to assure deletion should be able to handle and recover from errors. Users should be able to get the system back to a normal state if an error occurs during deletion.

Dishonest Service Provider and Honest Service Provider

In order for users to attain assured deletion in the cloud despite the intentions of the dishonest or honest provider, the following requirements must be satisfied: *service availability*, *deletion granularity* and *timeliness*.

Service Availability: Deletion tasks should not affect service delivery. The cloud system should continue to offer service uninterruptedly. Users should continue to have the ability to complete computation in the cloud.

Deletion granularity: For both providers, honest and dishonest providers, it is essential that the system only deletes data that is requested for deletion without affecting other data stored in the system.

Timeliness: Deletion task should complete swiftly or within a reasonable time. Deleted data should not be accessible after this time has elapsed.

The next section presents the principles for assured deletion. These principles are based on the discussion of the requirements from both the user and the provider side. Table 6.7 concludes this chapter showing the mapping between requirements and principles.

6.3.3 Principles of Usable Assured Deletion

These design principles seek to address some important aspects of usability and assured deletion. They focus on promoting consistency of deletion controls, availability of choice while deleting, relevant and informative information including feedback, minimisation of errors, and allowing users to understand what is possible when deleting. By introducing standard principles, users can concentrate on using the system and achieving their desired output without the need to understand the technical aspects of the system (Norman 2013).

The following principles should be considered when implementing the user requirements. They are adapted from Nelson's 10 usability heuristics for user interface design (Nielsen and Molich 1990, Nielsen 1994).

Principle 1: Offer informative and immediate feedback. (*See Table 6.1*)

Principle 2: Allow users to choose. (*See Table 6.2*)

Principle 3: Transparent. (*See Table 6.3*)

Principle 4: Consistent controls. (*See Table 6.4*)

Principle 5: Sufficient and relevant information. (*See Table 6.5*)

Principle 6: Reduce errors. (*See Table 6.6*)

Table 6.1 - Principle 1: Offer informative and immediate feedback.

Name: Offer informative and immediate feedback.	
Intent	To ensure that users get sufficient, useful, immediate and clear feedback when deleting from the cloud.
Motivation	Chapter 4 revealed that users sometimes fail to delete because they do not understand deletion or the consequences of their actions. Users stated that some messages are difficult to understand, or not pertinent to cloud deletion. Some users also reveal that it is not always clear whether the deletion task has completed successfully or what it will impact. The user study about deletion information requirements (*see Chapter 5*) also revealed that users desire relevant feedback at the right time in order to reduce actions which may lead to errors. Furthermore, appropriate feedback would inform the mental models of users, helping them to better understand the consequences of their actions and thus reducing deletion mistakes.
Description	For every deletion task or action that the user undertakes, the cloud system should provide feedback. This should be prompt and sufficient to accurately define the current state of the cloud system and the consequences of the users' actions. Frequent deletion actions require the feedback to be minimal and modest, while the feedback for infrequent action should contain substantial information. Successful deletion tasks should provide immediate feedback to indicate completion while incomplete ones should instantly notify users of incompleteness. Furthermore, where a deletion task affects more than one user, the system should be able to notify other relevant users immediately.
Example	Users should be explicitly informed that when they delete from a shared folder, the deletion will affect all the members of the shared folder.
Infrastructure Constraints	With different backups (usually deployed to offer availability), it is possible that some feedback that concerns time, (for instance, feedback on the exact time complete deletion would end) might be delayed. However, it should be possible to inform users exactly which parts of the deletion process has been completed and which ones have failed. The complexity of the infrastructure might also increase the number of notifications which may annoy users.

Table 6.2 - Principle 2: Allow users to choose

Name: Allow users to choose	
Intent	To ensure that users can specify the type of deletion they want or how they want their data to be deleted.
Motivation	The study presented in chapter 5 revealed that users have different deletion preferences, depending on the sensitivity of the data, deletion type (e.g., soft deletion) or how long an item should remain before being permanently removed from the cloud. These preferences change over time and may be affected by many factors. This desire was also highlighted in Chapter 4, where users expressed their wish to have more control over deletion. For example, they wanted to be able to determine when complete deletion should apply. This was further clarified in chapter 5, where users expressed that different motives to delete need to be handled differently. For instance, a user deleting for privacy reasons may desire the item to be deleted promptly and permanently, while the user who is deleting to organise their account may not be concerned by deletion time.
Description	The systems' interface should give users the ability to choose and apply their deletion preferences. In order that starters or novice users are not overwhelmed, these controls should be multi-layered; fine-grained and coarse-grained to allow both the novice and expert to choose what they want. Normal mode should have fewer controls to protect users from making mistakes.
Example	Course-grained deletion controls should be easily accessible through the interface, (e.g., through mouse right-click) while deletion granularity should be embedded under the settings menu.
Infrastructure Constraints	Since the user might prefer complete deletion, the process might not be immediate and might take some time. As stated earlier, data stored in the cloud might be scattered in several parts of the system and might not offer all deletion types that users may desire. Deletion preferences might be challenging to implement and may increase the number of controls, which may ultimately affect how they delete.

Table 6.3 - Principle 3: Transparent

Name: Transparent	
Intent	To ensure that users understand what is possible while using the cloud interface to delete.
Motivation	Chapter 4 has shown that users use and delete from different platform interfaces. This often leads to the construction of various and overlapping mental models which may cause unexpected results. For instance, some Android users think deleting from their cloud mobile app does not move the deleted items to the cloud bin because Android deletion does not have the concept of recycle bin. Some users also think deleting from their sync folder in their computers does not delete files in their cloud account. It is essential that the cloud system informs users of what is possible and match their expectations with minimum effort.
Description	The cloud interface and the deletion controls should be sufficiently transparent so that users to know what they can do and what they cannot do regarding deletion. It must not give users a false impression of what is possible. Cloud providers should keep in mind that users have different perceptions of the cloud and how deletion works hence different expectations when deleting. The system should be clear on what has taken place or is possible, to avoid misleading users into errors they may not recover from. Give users clear instructions on how to perform possible actions.
Example	Users should be made aware that deleting files from a sync folder or mobile app will send their files to the cloud bin.
Infrastructure Constraints	N/A

Table 6.4 - Principle 4: Consistent controls

Name: Consistent controls	
Intent	To ensure that users who need to delete can locate controls despite the platform.
Motivation	The study in chapter 4 revealed that some users found it easy to delete from the sync folder in their computer than on the web interface. They stated that they better understood and were more used to deleting from a computer than from the web. Some users indicated that they are sometimes discouraged to delete from specific platforms because of the effort that is required. As a result of inconsistent controls, some users were also found to likely delete from one particular platform than all the platforms they use to access the cloud (*see Chapter 4*).
Description	Deletion controls should be consistent across different platforms (i.e., web app, mobile app, and sync folders). While the implementation may differ, cloud users should be able to accomplish the same deletion tasks across different platforms. Deletion controls should not be hidden but be visible to users regardless of the interface.
Example	The location of cloud bin should be visible in all the platforms to allow easy recovery after deletion.
Infrastructure Constraints	Since cloud users use different devices and interfaces to access their cloud accounts, it is possible that some controls may be hidden due to limited interface size and space (e.g., cloud mobile application).

Table 6.5 - Principle 5: Sufficient and relevant information

Name: Sufficient and relevant information	
Intent	To ensure that users have sufficient information about deletion when they need it.
Motivation	The results presented in chapter 4 revealed that insufficient information on cloud deletion often leads to misunderstanding and consequent failure to delete. Lack of relevant information may lead to the construction of incomplete mental models. Users need information to have a better understanding of deletion. It is also evident (*see Coping strategies 4.2*) that users look for such information to recover from mistakes and errors. Users also stated that they wanted transparency, cloud providers should reveal their deletion practices clearly to users. In the participatory study presented in chapter 5, it is clear that relevance on its own is not enough for transparency, timing and channel are also crucial. Making relevant information available to users at the right time and place helps users make informed decisions about using the service and reducing the chances of incurring errors. Essential information should not only be restricted to privacy policies but should be made available in other places as well.
Description	The cloud system should have all the necessary information through all available channels. This information should be relevant, sufficient and found within the platform they are using. Users should not be required to leave the platform to search for more information. The vocabulary used should be consistent and straightforward. It should be possible to use the information provided to form accurate or meaningful mental models that lead to successful deletion.
Example	Information about data retention should not only be restricted to Privacy Policies but should also be made available from other channels.
Infrastructure Constraints	Availability and presentation of information about deletion may be affected by the type of device being used to access the cloud. Making all the information available through mobile phones might not be suitable compared to web interfaces.

Table 6.6 - Principle 6: Reduce errors

Name: Reduce errors	
Intent	To lower the chances of users making mistakes while deleting.
Motivation	Chapter 4 revealed that users often make assumptions about the deletion process or how the cloud works, leading to mistakes and unwanted consequences. This study also disclosed that some mistakes are due to faulty and confusing user interfaces, a user may delete the wrong item without being aware. It is also highlighted that too many steps while deleting may negatively affect the deletion mental models of users.
Description	The cloud system should help users avoid making mistakes. The system should warn users and make it difficult for them to cause errors or make mistakes when deleting. If a system error occurs, minimise its impact on the user and where possible provide ways of recovering from an error state on time. Regular deletion tasks (soft deletion) should include fewer steps than a complete deletion task (permanent) which may be irreversible. In the case of irreversible actions, deploy extra steps to caution the user of the consequences of their action. Make deletion controls distinctive to other controls, so that it is difficult for users to complete actions that cannot be reversed.
Example	Allow users to see the recently deleted files easily. To avoid eternal data loss, give users a warning before completely emptying their cloud bin.
Infrastructure Constraints	Because of the number of components involved in the storage of data in the cloud, the cloud system might still experience some technical errors which might affect the user.

Table 6.7 Summary: Mapping of the user requirements and principles

User Requirements	**Principles to be considered**
Complete deletion	*Transparency* *Allow users to choose* *Consistent controls*
Relevant, informative and immediate information	*Provide sufficient and relevant information* *Reduce errors*
Choose type of deletion	*Allow users to choose* *Consistent controls* *Reduce errors*
Help	*Provide sufficient and relevant information*
Error Handling	*Transparency* *Provide sufficient and relevant information* *Reduce errors*
Usable interface and controls	*Consistent controls* *Reduce errors*

6.4 Summary

This chapter proposed a conceptual framework for the design of usable assured deletion and presented user requirements and principles for assured deletion. It first presented and discussed the proposed conceptual framework for the design of usable assured deletion. It has discussed how the main concepts from the findings of Chapter 3, 4 and 5 relate to each other. This conceptual framework may serve the research community by providing an understanding of usable assured cloud deletion and encourage further theory development, and most importantly promote viable research to spur usable technological solutions. This chapter concludes by presenting the user requirements and principles for assured deletion in the cloud. It presented the user requirements for usable assured deletion formulated from the studies discussed in Chapters 4 and 5. Considering these user requirements and the ones presented in Chapter 3, a comparison is made to show which requirements are common from all the three setups. The similarities of these requirements have shown that complete deletion and usability are the standard requirements in all the setups, suggesting that in all situations it is vital to satisfy those to assure removal of data in the cloud. This chapter has also distilled and presented seven design principles for usable assured deletion that need to be respected during the implementation of the user requirements. Table 6.7 shows the mapping between the requirements and the principles that need to be considered when implementing them.

Chapter 7

Conclusion

The work presented in this thesis explored usable assured deletion in the cloud. It has considered this from two perspectives; the cloud service provider side and the users' side. From the cloud service provider's perspective, this work has investigated the challenges that cloud service providers face when attempting to assure deletion. From the user's perspective, this work has examined the needs of users concerning deletion – their preferences and how they want to obtain information about deletion. Using these findings, this work has examined what user requirements should be satisfied and what principles should be followed to assure deletion. Overall, this work sought out to answer the following objectives:

- What are the assured deletion challenges for cloud providers?
- What do users struggle with?
- What do users want, that is, deletion preferences and how to be informed about deletion?
- How do user and cloud provider challenges relate, what are the resulting requirements and what principles should be followed for usable assured deletion in the cloud?

7.1 Major Findings

The major findings of this thesis are presented below:

Salient features of the cloud pose challenges to assured deletion. This work has established that the notable features of the cloud pose different challenges to assured deletion. Cloud features such as virtualisation, multi-tenancy, live migration, high availability and on-demand elasticity make it difficult to assure deletion in the cloud. Data stored in the cloud is affected by these features. Chapter 3 presented an analysis of how these cloud features challenge assured deletion in the cloud and offered a systematisation of these challenges. It also presented a conceptual model depicting what areas of the cloud need to be considered to assure deletion.

Users face various challenges when deleting from the cloud. Chapter 4 of this thesis provided empirical evidence on what makes users struggle or fail to delete. This can be attributed to poorly designed cloud deletion mechanisms, their incomplete mental models, and lack of sufficient information on deletion.

Cloud users' deletion needs and preferences are diverse and complex. Chapter 4 provides empirical evidence on users' deletion practices (i.e., motivations and challenges) while Chapter 5 provides experimental evidence on users' deletion preferences. Users motivations to delete fall into four major categories: privacy-, policy-, expertise- and storage-driven. Their deletion preferences are multi-dimensional, complex, change over time and vary across individuals and groups. Chapter 4 also discusses the challenges users face when deleting from the cloud.

Information on deletion is essential to users' understanding of cloud deletion and recovering from deletion failures. Users perceive information on cloud deletion information as scarce and restricted to privacy policies. Chapter 5 demonstrates what information is considered essential concerning deletion, the time they would prefer to be provided with such information and the communication channels they prefer to have such information. It also shows that users use deletion information to inform their everyday deletion decisions and recover from failures. Consequently, information on deletion should be distributed across the platform for easy access.

There is a need for interface and deletion control mechanism improvement. Using the findings from Chapter 3, 4 and 5 , this work proposes six (6) user requirements and seven (7) principles for usable assured deletion. These requirements and principles are discussed in Chapter 6. For the cloud to offer usable assured deletion and meet the deletion needs of users, these requirements need to be satisfied. The principles should be followed when implementing the requirements. Table 6.4 present a mapping between the principles and the conditions they guide.

Assured deletion is more than inaccessibility of deleted data. Through existing literature and the results of the studies conducted as part of this thesis, it is evident that assured deletion has more than just one property. Other properties include completeness, proof of deletion, timeliness, usability, adequate feedback, fine-grained deletion, removal of data.

- ***Removal of data:*** if deleted data is not removed from the cloud when the request is made, it may end taking useful space that can be used to store other files. Also, unauthorised users (e.g., hackers) may forensically analyse their allocated storage for valuable data.

- ***Completeness:*** deletion of all data including copies and meta-data may protect individuals and enterprises alike from unintentional data disclosures.
- ***Proof of deletion:*** being able to prove data deletion will help the provider to comply with the law and gain more confidence from the tenants consequently improving their sales.
- ***Deletion granularity:*** it is vital for users to be able to specify which data they want to delete and for the cloud system to be able to delete such data without affecting other data.
- ***Timeliness:*** if data is not promptly deleted when the request is made, it may end up being visible to other users before it is truly removed from the cloud.
- ***Adequate feedback:*** Assured deletion also includes having sufficient information or feedback about deletion. For instance, if the system does not provide sufficient information about deletion, a user may delete and never be sure if the deletion has taken place.
- ***Usability:*** if deletion controls are not usable, users may make mistakes and unintentionally delete crucial data (i.e., suffer data loss).

7.2 Future Work

Deep understanding of cloud users' mental models. The work presented in this thesis uncovers different mental models constructed mainly by those participants who could not delete, and these mental models seem to influence users' deletion practices and behaviours. Hence, this opens an opportunity to understand mental models of using the cloud in general, particularly focusing on whether users use or transfer these models to the cloud from other domains (e.g., computers, and smartphones) or whether they develop new ones to cope with a new reality. It is essential to understand the extent these mental models have on deleting from the cloud. It would also be interesting to study security experts about their cloud usage concerning deletion. Understanding how they use and delete from the cloud could shed light on the differences between them and the findings of this thesis.

Assured deletion in other cloud use cases This study focused on usable assured deletion on cloud storage. Future work could consider usable assured deletion considering other use cases, e.g., using cloud storage versus virtual machines. It would be interesting to understand how deletion differs in these use cases and what providers need to do to assure the removal of data. Another cloud service delivery model that could be explored is software as a service (SaaS) delivery model. Investigating users' deletion practices from such environments would be interesting.

Evaluation of encryption tools and deletion. Some studies (e.g., Tang et al. (2010), Rahumed et al. (2011)) recommend the use of encryption tools in the cloud to protect users' privacy after deletion. However, none of the participants who took part in two studies presented in this thesis mentioned the use of encryption as a means to assure deletion. It is not clear whether users are not using such tools because of a lack of awareness or due to usability issues. Usability studies in this area would help understand how such tools could be improved, or how users could be encouraged to adopt them.

Multiparty access control. Chapter 4 (See Section 4.2.2), revealed that cloud users possess incomplete mental models about deleting from *shared* folders, which are managed by one or more users. Even if these models were complete and accurate, the issue of data management, and in particular data deletion, when multiple users are involved in the cloud is an under-explored area. Such multi-party access control has been studied in other domains such as social networks (Such and Criado 2016), and it would be interesting to study the applicability and usability of such techniques in order to support deletion in the cloud.

Follow-up confirmatory studies. Finally, it should be noted that the work presented in this thesis identified factors that play a role in deletion in the cloud and potential relationships between them grounded in the data obtained through semi-structured interviews and participatory action method. The next step would be to undertake confirmatory studies, to further understand these concepts and confirm the extent of their relationships.

7.3 Closing remarks

The work presented in this thesis focuses on usable assured deletion in the cloud. Despite assured deletion being an important aspect of data management, it is often overlooked. However, when considering the full cycle of the data ecosystem, it plays a significant role as data storage and access controls – the removal of data is essential. However, in the cloud data ecosystem, it is more complex and challenging, this thesis is a stepping stone in understanding the challenges. The research and the principles in this thesis formulate a solid grounding for further work in this area. As we build more and more complex and automated systems, we need to consider the full data life cycle – what data will be collected and how it will be removed from such systems. We need to empower users with regards to all the aspects of that data life cycle.

Appendix A

Deletion Practices Study

This appendix contains a copy of the questions asked during the screening stage and the interviews for the study presented in Chapter 4.

A.1 Screening questions

1. Which gender do you most identify yourself with?

   ○ Male

   ○ Female

   ○ Other

   ○ Prefer not to say

2. Is this the same gender you had at birth? ○ Yes ○ No ○ Prefer not to say

6. How old are you?

   ○ 18 - 24

   ○ 25 - 34

   ○ 35 - 44

   ○ 45 - 54

   ○ 55 - 64

   ○ 65 +

7. Please indicate the highest level of education completed.

   ○ No formal education

   ○ High school graduate

   ○ Some college

   ○ Undergraduate degree

   ○ Postgraduate degree

   ○ Professional degree

   ○ Doctorate

○ Prefer not to say

8. Employment status

   ○ Employed full time

   ○ Employed part time

   ○ Unemployed looking for work

   ○ Unemployed not looking for work

   ○ Retired

   ○ Student

   ○ Prefer not to say

9. Which of the following do you use at least once per week to access your cloud: (Check all that apply) □ Smartphone □ Desktop □ Laptop □ Tablet

10. Which of the following cloud storage services do you use at least once per month: (Check all that apply)

    □ Dropbox

    □ Box

    □ iCloud

    □ G-Drive (Google drive)

    □ OneDrive (Microsoft One drive)

    □ Amazon Cloud Drive

    □ None

    □ Other:

11. How many cloud storage accounts do you have?

    ○ No accounts

    ○ 1 Account

    ○ 2 - 3 Accounts

    ○ 4 - 5 Accounts

    ○ 6 + Accounts

12. I use this Cloud Storage account: (Check all that apply)

    □ Collaborating with co-workers, classmates, or professional contacts by jointly creating and editing files.

    □ Collaborating with friends and family by jointly creating and editing files.

    □ Sharing files that I have created with co-workers, classmates, or other professional contacts.

    □ Backing up files related to my job, school, or career.

    □ Backing up files that are not related to my job, school, or career

13. Which of the following applies to you? (Check all that apply)

    □ Uploaded data online (e.g., photos, documents)

    □ Deleted data from online (e.g., photos, documents)

    □ Shared folder/photos/documents online

    □ Deleted a Cloud account

    □ Downloaded data from online service

□ Read online privacy policies

□ Used account settings to limit data about me that could be collected or used

14. On average, how often do you run out of storage space on your Cloud Storage account?

○ I am always out of storage space.

○ At least once a month

○ At least once a year, but less than once a month

○ Less than once a year, but sometimes

○ I have never run out of storage space

○ I don't know

15. How often do you delete from your Cloud Storage account?

○ Always

○ Most of the time

○ About half the time

○ Sometimes

○ Never

16. Have you ever experienced some challenges when attempting to delete from the cloud? ○ Yes ○ No

19. Which of the following do you find it easy to use when accessing your cloud? (Check all that apply)

□ Tablet/iPad

□ Computer/Laptop

□ Smartphone

□ Web interface (Cloud website)

□ Other

A.2 Interview Guide

Thank you for participating in our study. As you read in the consent form, we will be recording the session so we can review it to make sure that we don't miss any part of our conversation. Your information will be kept confidential and will only be accessed by us. Your name will not be associated with any data I collect. Do you have any questions regarding the consent form? Do I have your permission to start the recording?

1. Do you use any of the following services or similar services? Examples: Dropbox, Box, iCloud, G-Drive, One-drive.

   - Follow-up-1: How often do you use them?

     (a) Prompt: Would you say you use them every day?

   - Follow-up-2:What do you use these services for?

     (a) Prompt: Is it for work or its personal?

   - Follow-up-3: You mentioned that you use [service/services], how do you use [it/them].

2. Do you use any of the following services or related services? Examples: Office365, Google Docs etc.

   - Follow-up-1: How often do you use them?

   - Follow-up-2: What do you use these services for?

  (a) Prompt: Is it for work or its personal?

- Follow-up-3: Can you describe to me how you use [name of the service]?

3. Do you have any particular reason why you use these services?
4. When you store your files in [service mentioned in Q1] or create a document in [service mentioned in Q2] what happens?
   - Prompt: Do you know where they are stored?
5. Have you ever deleted something you uploaded on [service mentioned in question 1]?
   - Prompt: Have you ever thought of deleting something you have uploaded online?
   - Follow-up-1: Why?
   - Follow-up-2: Can you share with me how you go about deleting a file in [service mentioned by user in Q1]?
     (a) Prompt: You mentioned that you use [name of the service], how do you delete data from [name of the service]?
6. Have you ever deleted something you uploaded on [service mentioned in question 2]?
   - Follow-up-1: Why?
   - Follow-up-2: Can you share with me how you go about deleting a file in [service mentioned by user in Q2]?
     (a) Prompt: You mentioned that you use [name of the service], how do you delete data from [name of the service]?

   [**NOTE:** If the participant claims to have never deleted anything from the cloud before, ask the following question otherwise skip it]
7. You have mentioned that you have never deleted or been asked to delete anything before, how come?
   - Follow-up-1: How do you deal with information that you no longer need?
8. Have you ever faced problems or challenges when trying to delete your data from any of your services?
   - Prompt: Can you recall a time when you wanted to delete something but could not figure out how to delete it or you could simply not just delete.

   [If participant says Yes]
   - Follow-up-1: Which service was that and how did you resolve or get around those challenges? Or how did you finally delete then?
9. Have you ever been required to recover information you have previously deleted?
   - Prompt: Have you ever needed a document or file that you had previously deleted from [service mentioned in Q1 or Q2].
   - Follow-up-1: Were you successful?
   - Follow-up-2: How did you do it?
10. Do you ever think the information [e.g., files, documents] you have previously deleted still exist somewhere online or can be shared by your service provider?
    [If participant says Yes]
    - Follow-up-1: Why?
    - Follow-up-2: What do you do to ensure that your deleted information can never be shared after you have deleted it?

    [If participant says No]
    - Follow-up-1: You mentioned that you don't think your information could be shared after it has been deleted, why?
11. After you delete your files, do you know how long it takes for [service mentioned at Q1 or Q2] to delete them from their side?
    - Prompt: How long does deletion process take?

    [Explain to the participant that you will share a scenario with them and then ask questions using the scenario. Choose one scenario per interview depending on the interviewee occupation, for example, if the interviewee is a student ask them scenario one.]

Scenario 1 *After a [late night out/party/picnic], your [friend/colleague] creates a folder in [service mentioned] and shares it with you and your other friend. He then tries to be funny and decides to upload 3 embarrassing photos of you three that you took on the night out. You are embarrassed and decide to delete all the photos from the shared folder.*

Scenario 2 *You have just joined a new team at work. Your new supervisor creates a folder in [service mentioned by participant] and shares it with you and your other colleagues. Your supervisor uploads some documents for you and your team to work on. Upon a discussion between you and your supervisor, s/he realizes you don't need one of the documents so s/he asks you to delete the document.*

[Scenario Questions]

12. What do you think will happen when you delete the [photos/document]?

    - Prompt: Will [they/it] be deleted from the shared folder or just your computer or device?

13. Will the [photos/document] be deleted from all your [friends'/colleagues'] accounts or they will only be deleted from your account?

    - Prompt: Will the deletion process affect your [friends/colleagues] too?

[End of scenario questions]

Explain to the interviewee that the questions on the scenario have ended.

14. If you were told that information you delete may never be completely deleted, what would you do differently?

15. Do you know anything about the "right to be forgotten" European ruling?

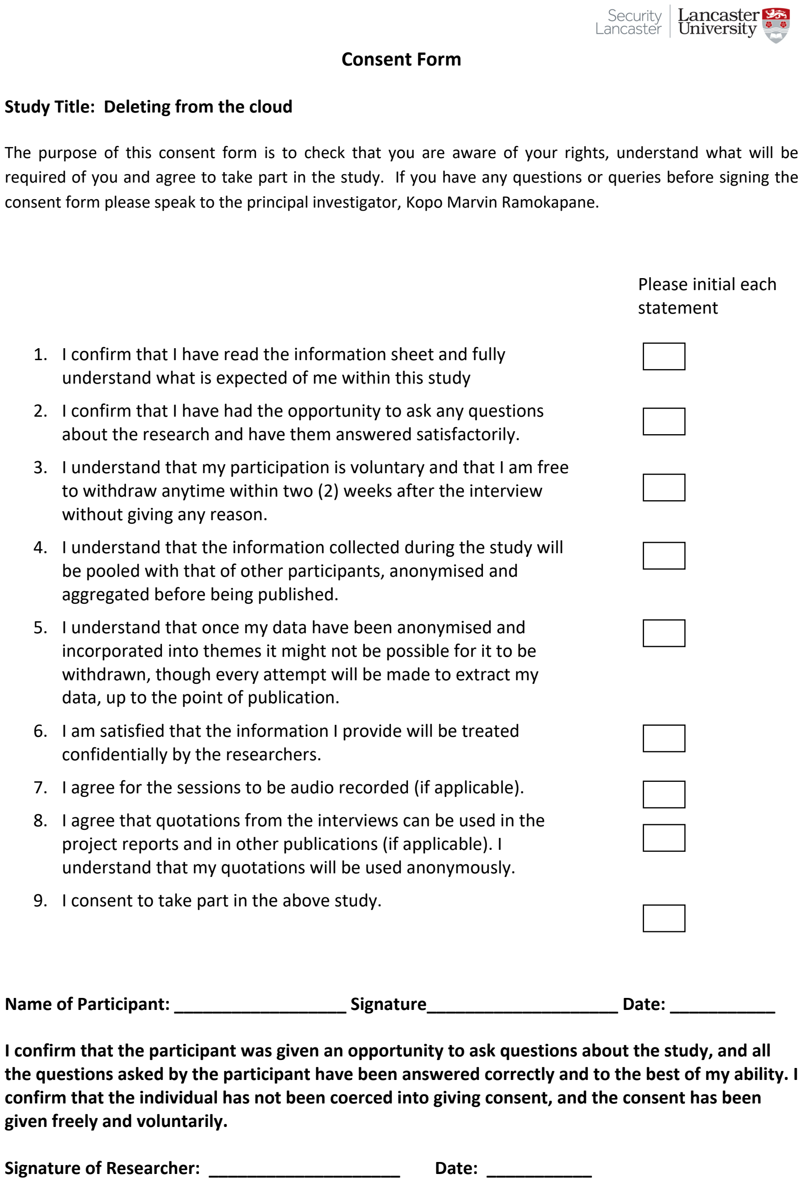
    [Explain to the user that you are at the end of the interview, ask them if they do have any questions or anything they want to share about deletion from the cloud.]

A.3 Respondents Demographics

Table A.1 Summary: Respondents Demographics. A total of 48 people responded to our advert, the table below summarizes their demographics.

		No. of participants
Gender	Male	16
	Female	32
Age	18 - 20	10
	21 - 25	18
	26 - 30	11
	31 - 35	5
	36 - 40	2
	41 - 45	0
	46 - 50	1
	51 - 55	1
Educational Background	High school/College course	14
	Bachelors	13
	Masters	11
	PhD	9
	Preferred not to say	1
Employment Status	Unemployed/Retired	2
	Full time	15
	Part-time	3
	Student	28
Cloud Services	Dropbox	22
	iCloud	20
	OneDrive	22
	Google Drive	21
	Box	13
	Microsoft Office 365	35
	Google Docs	29
	OneNote	6
Cloud Access	Smartphone	48
	Tablet	24
	Desktop	30
	Laptop	43
Cloud Activities	Uploaded files	48
	Deleted data	47
	Shared folder/files	46
	Deleted an account	30
	Recovered deleted files	15
	Downloaded files	36
	Read a service agreement	12
	None of the above	1

A.4 Consent form



Security Lancaster | Lancaster University

Consent Form

Study Title: Deleting from the cloud

The purpose of this consent form is to check that you are aware of your rights, understand what will be required of you and agree to take part in the study. If you have any questions or queries before signing the consent form please speak to the principal investigator, Kopo Marvin Ramokapane.

Please initial each statement

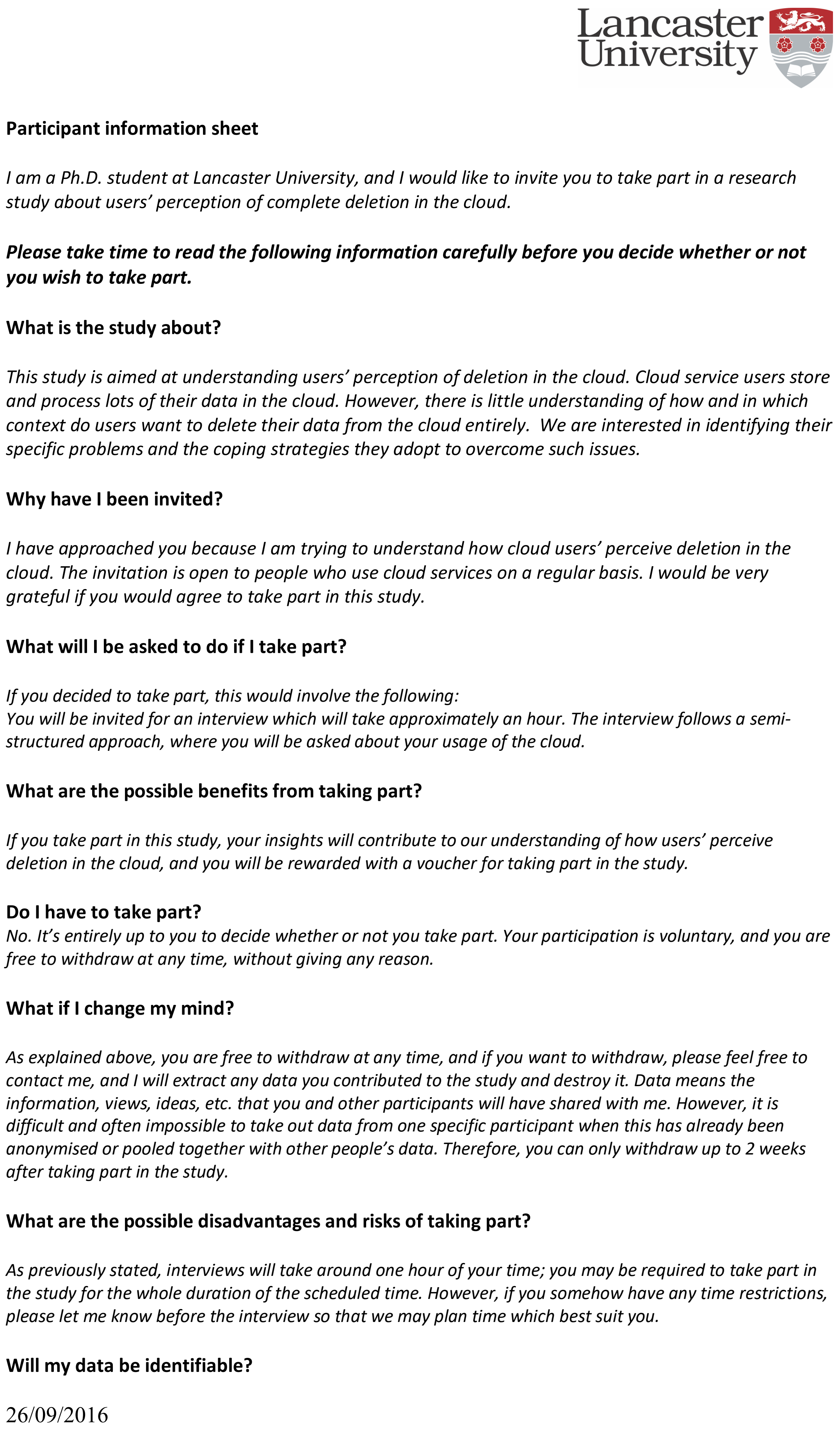
1. I confirm that I have read the information sheet and fully understand what is expected of me within this study □
2. I confirm that I have had the opportunity to ask any questions about the research and have them answered satisfactorily. □
3. I understand that my participation is voluntary and that I am free to withdraw anytime within two (2) weeks after the interview without giving any reason. □
4. I understand that the information collected during the study will be pooled with that of other participants, anonymised and aggregated before being published. □
5. I understand that once my data have been anonymised and incorporated into themes it might not be possible for it to be withdrawn, though every attempt will be made to extract my data, up to the point of publication. □
6. I am satisfied that the information I provide will be treated confidentially by the researchers. □
7. I agree for the sessions to be audio recorded (if applicable). □
8. I agree that quotations from the interviews can be used in the project reports and in other publications (if applicable). I understand that my quotations will be used anonymously. □
9. I consent to take part in the above study. □

Name of Participant: ___________________ **Signature**_____________________ **Date:** ___________

I confirm that the participant was given an opportunity to ask questions about the study, and all the questions asked by the participant have been answered correctly and to the best of my ability. I confirm that the individual has not been coerced into giving consent, and the consent has been given freely and voluntarily.

Signature of Researcher: _____________________ **Date:** ___________

A.5 Information Sheet



Participant information sheet

I am a Ph.D. student at Lancaster University, and I would like to invite you to take part in a research study about users' perception of complete deletion in the cloud.

Please take time to read the following information carefully before you decide whether or not you wish to take part.

What is the study about?

This study is aimed at understanding users' perception of deletion in the cloud. Cloud service users store and process lots of their data in the cloud. However, there is little understanding of how and in which context do users want to delete their data from the cloud entirely. We are interested in identifying their specific problems and the coping strategies they adopt to overcome such issues.

Why have I been invited?

I have approached you because I am trying to understand how cloud users' perceive deletion in the cloud. The invitation is open to people who use cloud services on a regular basis. I would be very grateful if you would agree to take part in this study.

What will I be asked to do if I take part?

If you decided to take part, this would involve the following:
You will be invited for an interview which will take approximately an hour. The interview follows a semi-structured approach, where you will be asked about your usage of the cloud.

What are the possible benefits from taking part?

If you take part in this study, your insights will contribute to our understanding of how users' perceive deletion in the cloud, and you will be rewarded with a voucher for taking part in the study.

Do I have to take part?
No. It's entirely up to you to decide whether or not you take part. Your participation is voluntary, and you are free to withdraw at any time, without giving any reason.

What if I change my mind?

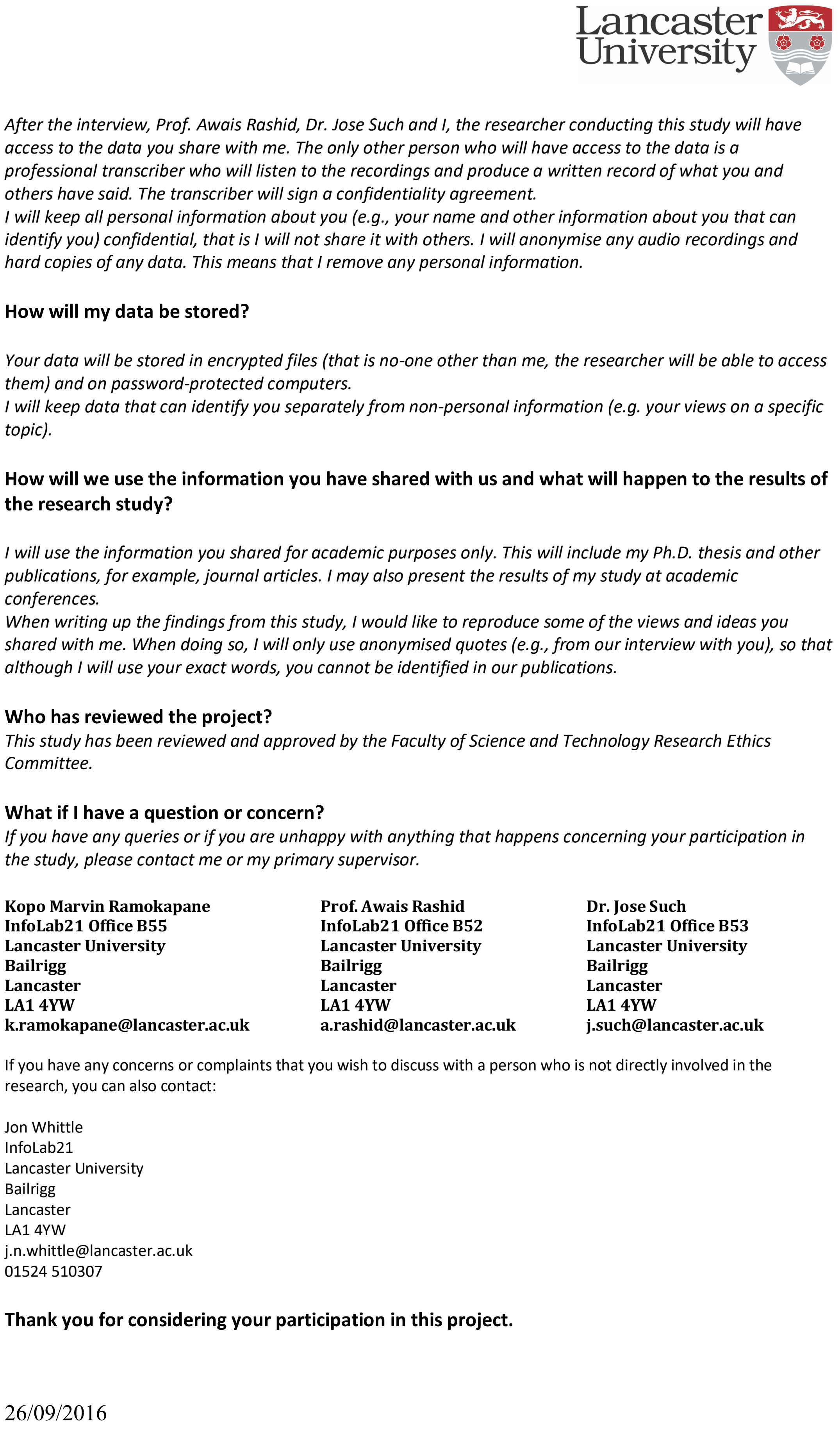
As explained above, you are free to withdraw at any time, and if you want to withdraw, please feel free to contact me, and I will extract any data you contributed to the study and destroy it. Data means the information, views, ideas, etc. that you and other participants will have shared with me. However, it is difficult and often impossible to take out data from one specific participant when this has already been anonymised or pooled together with other people's data. Therefore, you can only withdraw up to 2 weeks after taking part in the study.

What are the possible disadvantages and risks of taking part?

As previously stated, interviews will take around one hour of your time; you may be required to take part in the study for the whole duration of the scheduled time. However, if you somehow have any time restrictions, please let me know before the interview so that we may plan time which best suit you.

Will my data be identifiable?

26/09/2016



After the interview, Prof. Awais Rashid, Dr. Jose Such and I, the researcher conducting this study will have access to the data you share with me. The only other person who will have access to the data is a professional transcriber who will listen to the recordings and produce a written record of what you and others have said. The transcriber will sign a confidentiality agreement.
I will keep all personal information about you (e.g., your name and other information about you that can identify you) confidential, that is I will not share it with others. I will anonymise any audio recordings and hard copies of any data. This means that I remove any personal information.

How will my data be stored?

Your data will be stored in encrypted files (that is no-one other than me, the researcher will be able to access them) and on password-protected computers.
I will keep data that can identify you separately from non-personal information (e.g. your views on a specific topic).

How will we use the information you have shared with us and what will happen to the results of the research study?

I will use the information you shared for academic purposes only. This will include my Ph.D. thesis and other publications, for example, journal articles. I may also present the results of my study at academic conferences.
When writing up the findings from this study, I would like to reproduce some of the views and ideas you shared with me. When doing so, I will only use anonymised quotes (e.g., from our interview with you), so that although I will use your exact words, you cannot be identified in our publications.

Who has reviewed the project?
This study has been reviewed and approved by the Faculty of Science and Technology Research Ethics Committee.

What if I have a question or concern?
If you have any queries or if you are unhappy with anything that happens concerning your participation in the study, please contact me or my primary supervisor.

Kopo Marvin Ramokapane
InfoLab21 Office B55
Lancaster University
Bailrigg
Lancaster
LA1 4YW
k.ramokapane@lancaster.ac.uk

Prof. Awais Rashid
InfoLab21 Office B52
Lancaster University
Bailrigg
Lancaster
LA1 4YW
a.rashid@lancaster.ac.uk

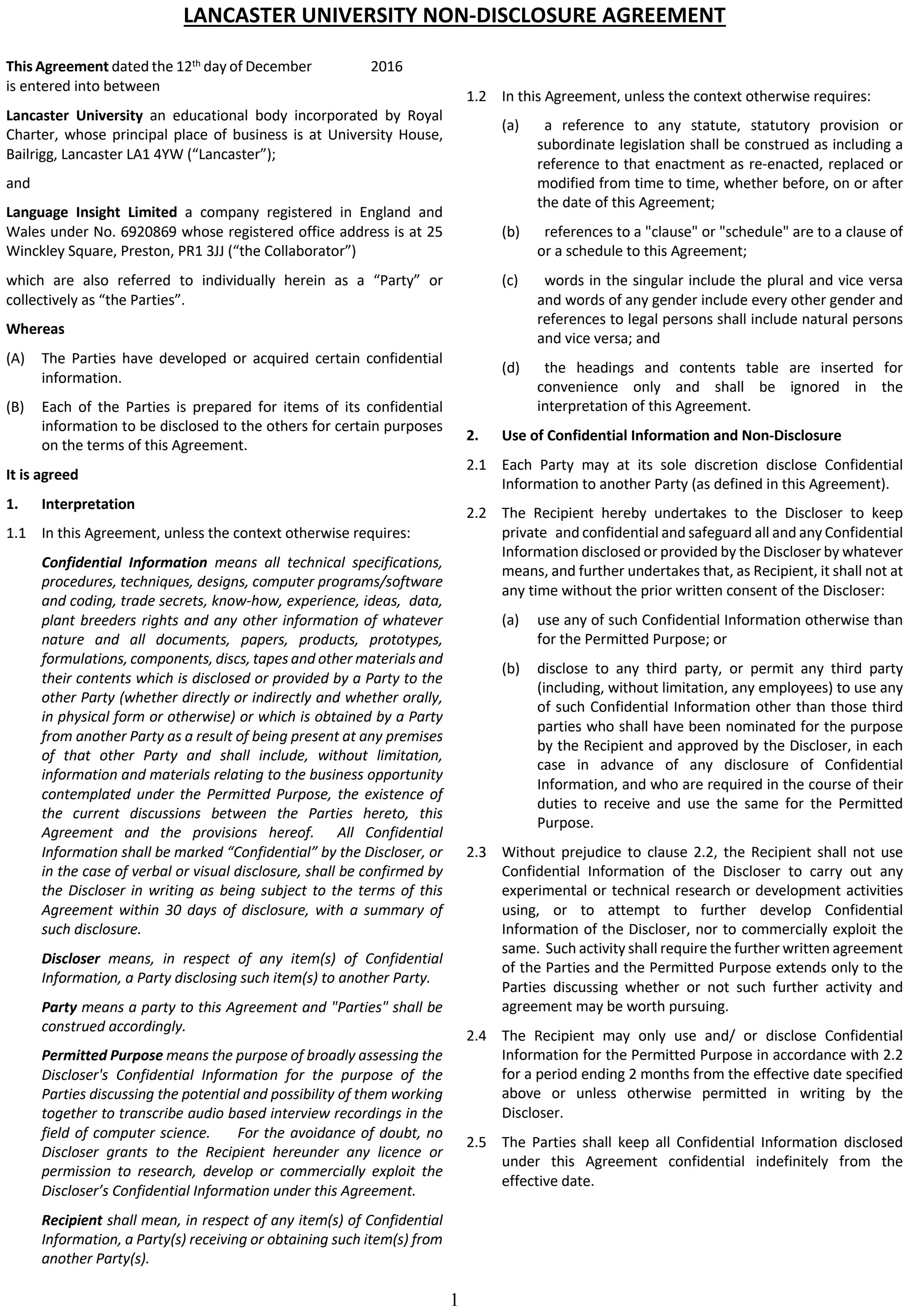
Dr. Jose Such
InfoLab21 Office B53
Lancaster University
Bailrigg
Lancaster
LA1 4YW
j.such@lancaster.ac.uk

If you have any concerns or complaints that you wish to discuss with a person who is not directly involved in the research, you can also contact:

Jon Whittle
InfoLab21
Lancaster University
Bailrigg
Lancaster
LA1 4YW
j.n.whittle@lancaster.ac.uk
01524 510307

Thank you for considering your participation in this project.

A.6 Non-Disclosure Agreement



LANCASTER UNIVERSITY NON-DISCLOSURE AGREEMENT

This Agreement dated the 12th day of December 2016 is entered into between

Lancaster University an educational body incorporated by Royal Charter, whose principal place of business is at University House, Bailrigg, Lancaster LA1 4YW ("Lancaster");

and

Language Insight Limited a company registered in England and Wales under No. 6920869 whose registered office address is at 25 Winckley Square, Preston, PR1 3JJ ("the Collaborator")

which are also referred to individually herein as a "Party" or collectively as "the Parties".

Whereas

(A) The Parties have developed or acquired certain confidential information.

(B) Each of the Parties is prepared for items of its confidential information to be disclosed to the others for certain purposes on the terms of this Agreement.

It is agreed

1. Interpretation

1.1 In this Agreement, unless the context otherwise requires:

Confidential Information *means all technical specifications, procedures, techniques, designs, computer programs/software and coding, trade secrets, know-how, experience, ideas, data, plant breeders rights and any other information of whatever nature and all documents, papers, products, prototypes, formulations, components, discs, tapes and other materials and their contents which is disclosed or provided by a Party to the other Party (whether directly or indirectly and whether orally, in physical form or otherwise) or which is obtained by a Party from another Party as a result of being present at any premises of that other Party and shall include, without limitation, information and materials relating to the business opportunity contemplated under the Permitted Purpose, the existence of the current discussions between the Parties hereto, this Agreement and the provisions hereof. All Confidential Information shall be marked "Confidential" by the Discloser, or in the case of verbal or visual disclosure, shall be confirmed by the Discloser in writing as being subject to the terms of this Agreement within 30 days of disclosure, with a summary of such disclosure.*

Discloser *means, in respect of any item(s) of Confidential Information, a Party disclosing such item(s) to another Party.*

Party *means a party to this Agreement and "Parties" shall be construed accordingly.*

Permitted Purpose *means the purpose of broadly assessing the Discloser's Confidential Information for the purpose of the Parties discussing the potential and possibility of them working together to transcribe audio based interview recordings in the field of computer science. For the avoidance of doubt, no Discloser grants to the Recipient hereunder any licence or permission to research, develop or commercially exploit the Discloser's Confidential Information under this Agreement.*

Recipient *shall mean, in respect of any item(s) of Confidential Information, a Party(s) receiving or obtaining such item(s) from another Party(s).*

1.2 In this Agreement, unless the context otherwise requires:

(a) a reference to any statute, statutory provision or subordinate legislation shall be construed as including a reference to that enactment as re-enacted, replaced or modified from time to time, whether before, on or after the date of this Agreement;

(b) references to a "clause" or "schedule" are to a clause of or a schedule to this Agreement;

(c) words in the singular include the plural and vice versa and words of any gender include every other gender and references to legal persons shall include natural persons and vice versa; and

(d) the headings and contents table are inserted for convenience only and shall be ignored in the interpretation of this Agreement.

2. Use of Confidential Information and Non-Disclosure

2.1 Each Party may at its sole discretion disclose Confidential Information to another Party (as defined in this Agreement).

2.2 The Recipient hereby undertakes to the Discloser to keep private and confidential and safeguard all and any Confidential Information disclosed or provided by the Discloser by whatever means, and further undertakes that, as Recipient, it shall not at any time without the prior written consent of the Discloser:

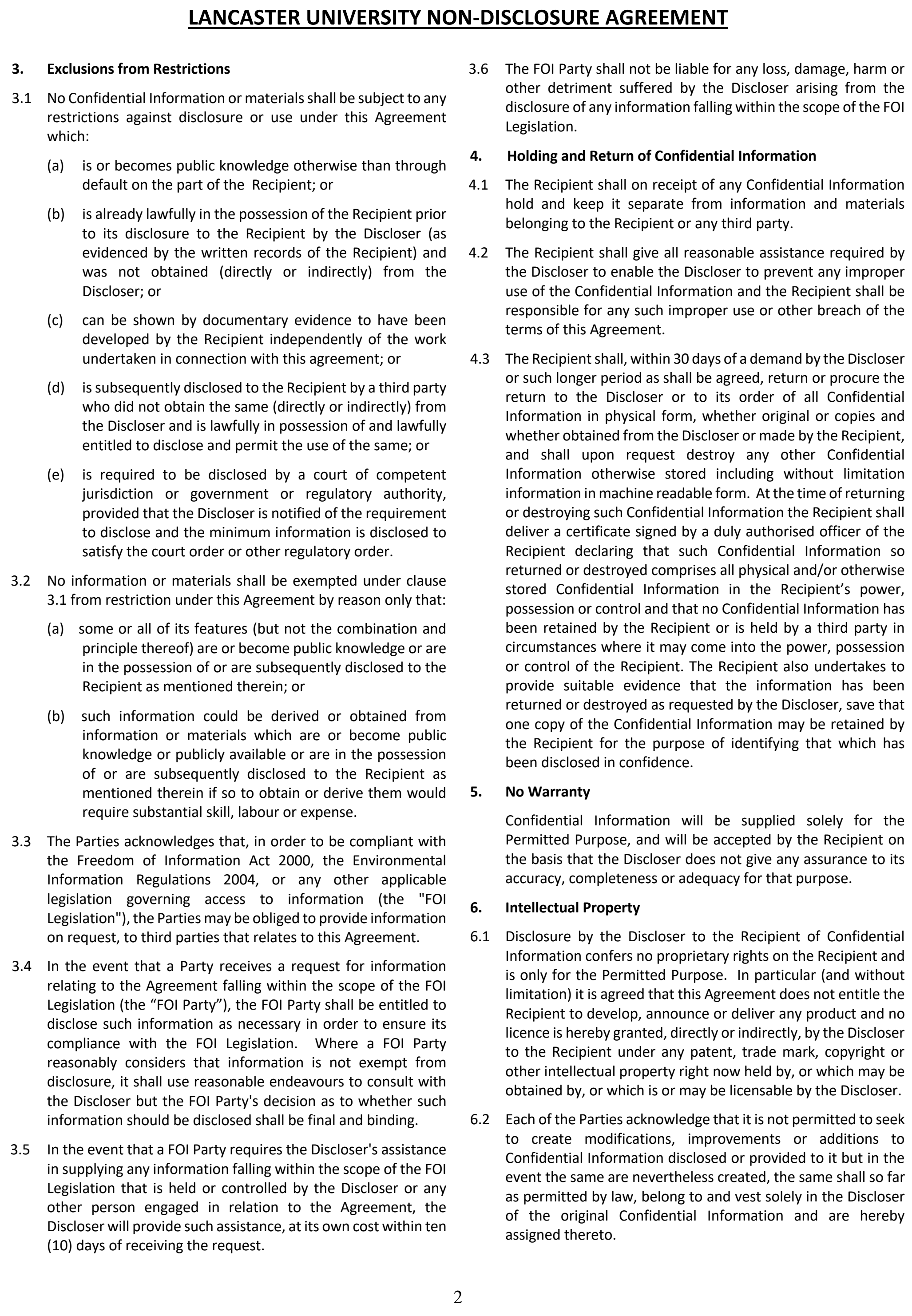
(a) use any of such Confidential Information otherwise than for the Permitted Purpose; or

(b) disclose to any third party, or permit any third party (including, without limitation, any employees) to use any of such Confidential Information other than those third parties who shall have been nominated for the purpose by the Recipient and approved by the Discloser, in each case in advance of any disclosure of Confidential Information, and who are required in the course of their duties to receive and use the same for the Permitted Purpose.

2.3 Without prejudice to clause 2.2, the Recipient shall not use Confidential Information of the Discloser to carry out any experimental or technical research or development activities using, or to attempt to further develop Confidential Information of the Discloser, nor to commercially exploit the same. Such activity shall require the further written agreement of the Parties and the Permitted Purpose extends only to the Parties discussing whether or not such further activity and agreement may be worth pursuing.

2.4 The Recipient may only use and/ or disclose Confidential Information for the Permitted Purpose in accordance with 2.2 for a period ending 2 months from the effective date specified above or unless otherwise permitted in writing by the Discloser.

2.5 The Parties shall keep all Confidential Information disclosed under this Agreement confidential indefinitely from the effective date.



LANCASTER UNIVERSITY NON-DISCLOSURE AGREEMENT

3. Exclusions from Restrictions

3.1 No Confidential Information or materials shall be subject to any restrictions against disclosure or use under this Agreement which:

(a) is or becomes public knowledge otherwise than through default on the part of the Recipient; or

(b) is already lawfully in the possession of the Recipient prior to its disclosure to the Recipient by the Discloser (as evidenced by the written records of the Recipient) and was not obtained (directly or indirectly) from the Discloser; or

(c) can be shown by documentary evidence to have been developed by the Recipient independently of the work undertaken in connection with this agreement; or

(d) is subsequently disclosed to the Recipient by a third party who did not obtain the same (directly or indirectly) from the Discloser and is lawfully in possession of and lawfully entitled to disclose and permit the use of the same; or

(e) is required to be disclosed by a court of competent jurisdiction or government or regulatory authority, provided that the Discloser is notified of the requirement to disclose and the minimum information is disclosed to satisfy the court order or other regulatory order.

3.2 No information or materials shall be exempted under clause 3.1 from restriction under this Agreement by reason only that:

(a) some or all of its features (but not the combination and principle thereof) are or become public knowledge or are in the possession of or are subsequently disclosed to the Recipient as mentioned therein; or

(b) such information could be derived or obtained from information or materials which are or become public knowledge or publicly available or are in the possession of or are subsequently disclosed to the Recipient as mentioned therein if so to obtain or derive them would require substantial skill, labour or expense.

3.3 The Parties acknowledges that, in order to be compliant with the Freedom of Information Act 2000, the Environmental Information Regulations 2004, or any other applicable legislation governing access to information (the "FOI Legislation"), the Parties may be obliged to provide information on request, to third parties that relates to this Agreement.

3.4 In the event that a Party receives a request for information relating to the Agreement falling within the scope of the FOI Legislation (the “FOI Party”), the FOI Party shall be entitled to disclose such information as necessary in order to ensure its compliance with the FOI Legislation. Where a FOI Party reasonably considers that information is not exempt from disclosure, it shall use reasonable endeavours to consult with the Discloser but the FOI Party's decision as to whether such information should be disclosed shall be final and binding.

3.5 In the event that a FOI Party requires the Discloser's assistance in supplying any information falling within the scope of the FOI Legislation that is held or controlled by the Discloser or any other person engaged in relation to the Agreement, the Discloser will provide such assistance, at its own cost within ten (10) days of receiving the request.

3.6 The FOI Party shall not be liable for any loss, damage, harm or other detriment suffered by the Discloser arising from the disclosure of any information falling within the scope of the FOI Legislation.

4. Holding and Return of Confidential Information

4.1 The Recipient shall on receipt of any Confidential Information hold and keep it separate from information and materials belonging to the Recipient or any third party.

4.2 The Recipient shall give all reasonable assistance required by the Discloser to enable the Discloser to prevent any improper use of the Confidential Information and the Recipient shall be responsible for any such improper use or other breach of the terms of this Agreement.

4.3 The Recipient shall, within 30 days of a demand by the Discloser or such longer period as shall be agreed, return or procure the return to the Discloser or to its order of all Confidential Information in physical form, whether original or copies and whether obtained from the Discloser or made by the Recipient, and shall upon request destroy any other Confidential Information otherwise stored including without limitation information in machine readable form. At the time of returning or destroying such Confidential Information the Recipient shall deliver a certificate signed by a duly authorised officer of the Recipient declaring that such Confidential Information so returned or destroyed comprises all physical and/or otherwise stored Confidential Information in the Recipient’s power, possession or control and that no Confidential Information has been retained by the Recipient or is held by a third party in circumstances where it may come into the power, possession or control of the Recipient. The Recipient also undertakes to provide suitable evidence that the information has been returned or destroyed as requested by the Discloser, save that one copy of the Confidential Information may be retained by the Recipient for the purpose of identifying that which has been disclosed in confidence.

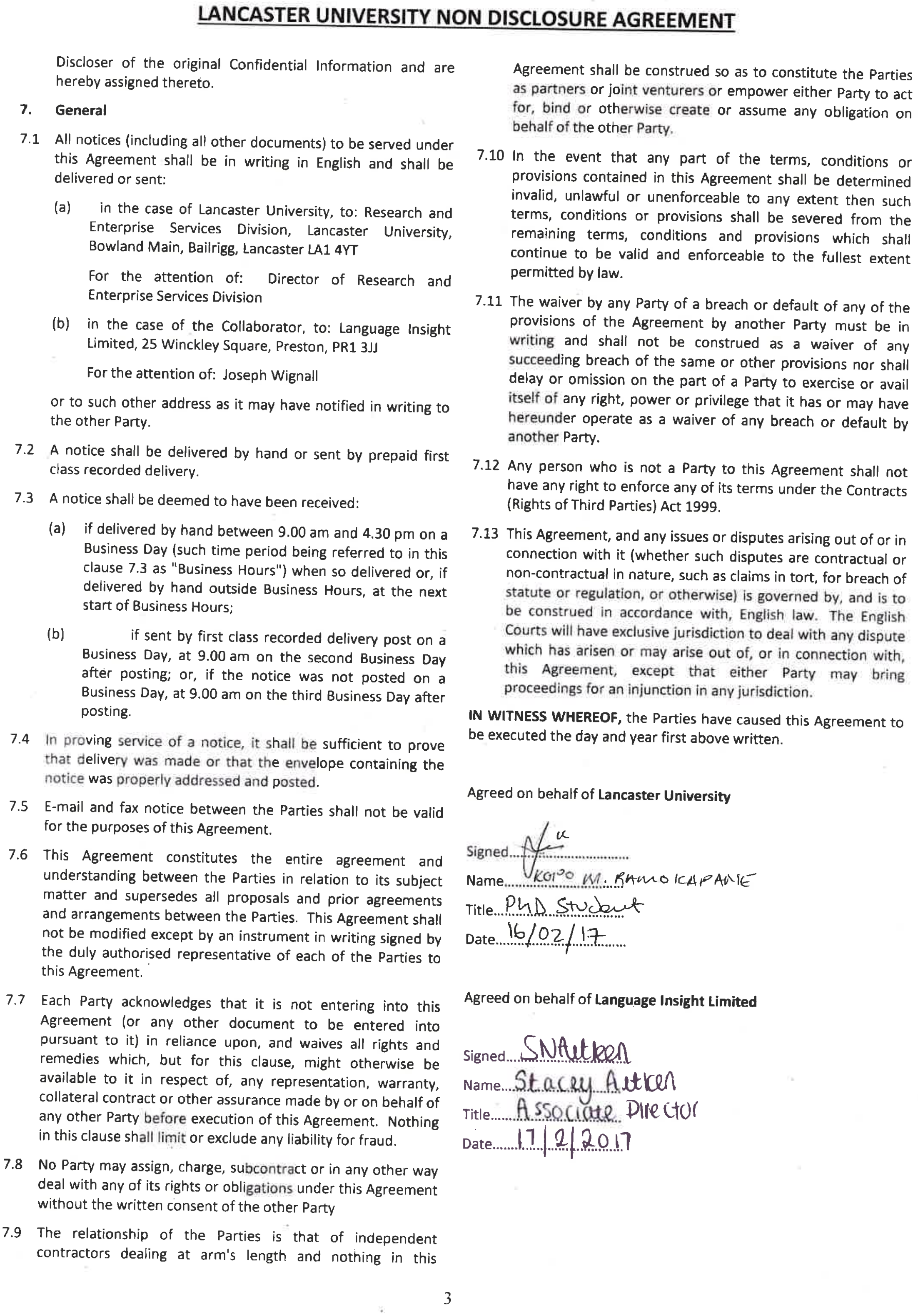
5. No Warranty

Confidential Information will be supplied solely for the Permitted Purpose, and will be accepted by the Recipient on the basis that the Discloser does not give any assurance to its accuracy, completeness or adequacy for that purpose.

6. Intellectual Property

6.1 Disclosure by the Discloser to the Recipient of Confidential Information confers no proprietary rights on the Recipient and is only for the Permitted Purpose. In particular (and without limitation) it is agreed that this Agreement does not entitle the Recipient to develop, announce or deliver any product and no licence is hereby granted, directly or indirectly, by the Discloser to the Recipient under any patent, trade mark, copyright or other intellectual property right now held by, or which may be obtained by, or which is or may be licensable by the Discloser.

6.2 Each of the Parties acknowledge that it is not permitted to seek to create modifications, improvements or additions to Confidential Information disclosed or provided to it but in the event the same are nevertheless created, the same shall so far as permitted by law, belong to and vest solely in the Discloser of the original Confidential Information and are hereby assigned thereto.



LANCASTER UNIVERSITY NON DISCLOSURE AGREEMENT

Discloser of the original Confidential Information and are hereby assigned thereto.

7. **General**

7.1 All notices (including all other documents) to be served under this Agreement shall be in writing in English and shall be delivered or sent:

(a) in the case of Lancaster University, to: Research and Enterprise Services Division, Lancaster University, Bowland Main, Bailrigg, Lancaster LA1 4YT

For the attention of: Director of Research and Enterprise Services Division

(b) in the case of the Collaborator, to: Language Insight Limited, 25 Winckley Square, Preston, PR1 3JJ

For the attention of: Joseph Wignall

or to such other address as it may have notified in writing to the other Party.

7.2 A notice shall be delivered by hand or sent by prepaid first class recorded delivery.

7.3 A notice shall be deemed to have been received:

(a) if delivered by hand between 9.00 am and 4.30 pm on a Business Day (such time period being referred to in this clause 7.3 as "Business Hours") when so delivered or, if delivered by hand outside Business Hours, at the next start of Business Hours;

(b) if sent by first class recorded delivery post on a Business Day, at 9.00 am on the second Business Day after posting; or, if the notice was not posted on a Business Day, at 9.00 am on the third Business Day after posting.

7.4 In proving service of a notice, it shall be sufficient to prove that delivery was made or that the envelope containing the notice was properly addressed and posted.

7.5 E-mail and fax notice between the Parties shall not be valid for the purposes of this Agreement.

7.6 This Agreement constitutes the entire agreement and understanding between the Parties in relation to its subject matter and supersedes all proposals and prior agreements and arrangements between the Parties. This Agreement shall not be modified except by an instrument in writing signed by the duly authorised representative of each of the Parties to this Agreement.

7.7 Each Party acknowledges that it is not entering into this Agreement (or any other document to be entered into pursuant to it) in reliance upon, and waives all rights and remedies which, but for this clause, might otherwise be available to it in respect of, any representation, warranty, collateral contract or other assurance made by or on behalf of any other Party before execution of this Agreement. Nothing in this clause shall limit or exclude any liability for fraud.

7.8 No Party may assign, charge, subcontract or in any other way deal with any of its rights or obligations under this Agreement without the written consent of the other Party

7.9 The relationship of the Parties is that of independent contractors dealing at arm's length and nothing in this Agreement shall be construed so as to constitute the Parties as partners or joint venturers or empower either Party to act for, bind or otherwise create or assume any obligation on behalf of the other Party.

7.10 In the event that any part of the terms, conditions or provisions contained in this Agreement shall be determined invalid, unlawful or unenforceable to any extent then such terms, conditions or provisions shall be severed from the remaining terms, conditions and provisions which shall continue to be valid and enforceable to the fullest extent permitted by law.

7.11 The waiver by any Party of a breach or default of any of the provisions of the Agreement by another Party must be in writing and shall not be construed as a waiver of any succeeding breach of the same or other provisions nor shall delay or omission on the part of a Party to exercise or avail itself of any right, power or privilege that it has or may have hereunder operate as a waiver of any breach or default by another Party.

7.12 Any person who is not a Party to this Agreement shall not have any right to enforce any of its terms under the Contracts (Rights of Third Parties) Act 1999.

7.13 This Agreement, and any issues or disputes arising out of or in connection with it (whether such disputes are contractual or non-contractual in nature, such as claims in tort, for breach of statute or regulation, or otherwise) is governed by, and is to be construed in accordance with, English law. The English Courts will have exclusive jurisdiction to deal with any dispute which has arisen or may arise out of, or in connection with, this Agreement, except that either Party may bring proceedings for an injunction in any jurisdiction.

IN WITNESS WHEREOF, the Parties have caused this Agreement to be executed the day and year first above written.

Agreed on behalf of **Lancaster University**

Signed............................

Name....KOPO M. RAMOKAPANE

Title....PhD Student

Date....16/02/17......

Agreed on behalf of **Language Insight Limited**

Signed....SNAitken

Name....Stacey Aitken

Title....Associate Director

Date....17/2/2017

Appendix B

Deletion Preferences Study

B.1 Activity Script

PAR 1 - Data sorting task

1. Introduce the task to participant

    - Show participants the list of types of data and ask them which of that information do they have stored in the cloud.
    - Ask them to sort this data into different groups, the groups can be of anything.

2. Get everyone around the table and ask them to sort the previous data types as a group. Let them talk amongst themselves.

PAR 2 - Deletion Metaphor

Metaphor

1. Show diagram depicting how waste is handled in a household.

    - Show participants the list of types of data and ask them which of that information do they have stored in the cloud.
    - Ask them to sort this data into different groups, the groups can be of anything.

2. Give participants examples of waste and asked participants to sort it according to how they would deal with such waste.
3. Ask them to give reasons.

Deletion preferences

1. Setup the table for the new activity. Talk about deletion and the metaphor. Similarities and differences
2. Introduce the new activity to users.

    - Ask them about their previous data sorting tasks. Ask them how they would delete data in such groups or what is important and unique about such data.
    - Introduce the props to them showing the different types of deletion.
    - Ask them to group data into the different types of deletion.
    - Encourage them to share their reasons and discuss amongst themselves.

    [After completing the task]

3. Ask participants to think about different cases.
    - Ask them to show how those preferences would change if it was a shared folder.
        - Shared between colleagues, family and public

PAR 3 - Information requirements

1. Setup and introduce the new activity.
    - Talk to participants about different information available on the internet. Discuss and make a list of different channels used with participants, what they are, their advantages and disadvantages.
    - Talk about cloud information, deletion information.
    - Ask participants to give examples of where cloud deletion information can be found.
2. Task 1: Ask participants to sort out given information according to how important it is to know with regards to deletion.
    - During the task, ask them about their decisions when they are not clear.
3. Task 2: Ask participants to state when they would want to know or see this information.
    - Use the groups they make and combine where necessary to have 3 groups: before, during, and after deleting.

    [Prompt Questions]

    - Which information should be displayed before using the cloud?
        - before deleting from the cloud?
        - during deletion?
        - after deletion?
4. Task 3: Ask participants to consider the groupings they have made, and ask them how or where this information should be made available. [Prompt Questions]
    - Which information should be made available in the adverts, policies, etc?

B.2 Survey Instrument

1. Which gender do you most identify yourself with?

    ○ Male

    ○ Female

    ○ Other

    ○ Prefer not to say

2. Is this the same gender you had at birth? ○ Yes ○ No ○ Prefer not to say

6. How old are you?

    ○ 18 - 24

    ○ 25 - 34

    ○ 35 - 44

    ○ 45 - 54

    ○ 55 - 64

    ○ 65 +

7. Please indicate the highest level of education completed.

    ○ No formal education

    ○ High school graduate

    ○ Some college

    ○ Undergraduate degree

    ○ Postgraduate degree

    ○ Professional degree

    ○ Doctorate

    ○ Prefer not to say

8. Employment status

    ○ Employed full time

    ○ Employed part time

    ○ Unemployed looking for work

    ○ Unemployed not looking for work

    ○ Retired

    ○ Student

    ○ Prefer not to say

9. Which of the following do you use at least once per week to access your cloud: (Check all that apply)
   □ Smartphone □ Desktop □ Laptop □ Tablet

10. Which of the following cloud storage services do you use at least once per month: (Check all that apply)

    □ Dropbox

    □ Box

    □ iCloud

    □ G-Drive (Google drive)

    □ OneDrive (Microsoft One drive)

    □ Amazon Cloud Drive

    □ None

    □ Other:

11. How many cloud storage accounts do you have?

    ○ No accounts

    ○ 1 Account

    ○ 2 - 3 Accounts

    ○ 4 - 5 Accounts

    ○ 6 + Accounts

12. I use this Cloud Storage account: (Check all that apply)

    □ Collaborating with co-workers, classmates, or professional contacts by jointly creating and editing files.

    □ Collaborating with friends and family by jointly creating and editing files.

    □ Sharing files that I have created with co-workers, classmates, or other professional contacts.

- □ Backing up files related to my job, school, or career.
- □ Backing up files that are not related to my job, school, or career

13. Which of the following applies to you? (Check all that apply)

- □ Uploaded data online (e.g., photos, documents)
- □ Deleted data from online (e.g., photos, documents)
- □ Shared folder/photos/documents online
- □ Deleted a Cloud account
- □ Downloaded data from online service
- □ Read online privacy policies
- □ Used account settings to limit data about me that could be collected or used

14. On average, how often do you run out of storage space on your Cloud Storage account?

- ○ I am always out of storage space.
- ○ At least once a month
- ○ At least once a year, but less than once a month
- ○ Less than once a year, but sometimes
- ○ I have never run out of storage space
- ○ I don't know

15. How often do you delete from your Cloud Storage account?

- ○ Always
- ○ Most of the time
- ○ About half the time
- ○ Sometimes
- ○ Never

16. Which of the following best describes why you delete from your Cloud storage services? (Check all that apply)

- □ To get more space.
- □ To clean my account
- □ To get rid of old files
- □ To preserve my privacy
- □ My work or organization policies require that I delete from data stored in these services
- □ Delete because I don't trust my provider
- □ I don't delete

17. Which of the following do you find it easy to delete your cloud data from? (Check all that apply)

- □ Tablet/iPad
- □ Computer/Laptop
- □ Smartphone
- □ Web interface (Cloud website)
- □ Other

18. Have you ever experienced some challenges when attempting to delete from the cloud? ○ Yes ○ No

21. How often would you be interested in seeing the summary of all your deleted data from your cloud provider?

    ○ Everyday

    ○ Every week

    ○ Every month

    ○ Every quarter

    ○ Every year

    ○ Never

22. Where should the information about cloud deletion be made available for users? (Check all that apply)

    □ Privacy Policies

    □ Advertisements

    □ Frequently Asked Questions (FAQs)

    □ Cloud Service Blog pages

    □ Pop-up dialogues

    □ Account home page

    □ One of the above

23. If anyone other than me deletes a file I uploaded into a shared folder I should be notified.

    ○ Strongly agree

    ○ Somewhat agree

    ○ Neither agree nor disagree

    ○ Somewhat disagree

    ○ Strongly disagree

24. If anyone other than me deletes any file in a shared folder I should be notified

    ○ Strongly agree

    ○ Somewhat agree

    ○ Neither agree nor disagree

    ○ Somewhat disagree

    ○ Strongly disagree

25. If I delete any file from a shared folder, other members of the shared folder should be notified

    ○ Strongly agree

    ○ Somewhat agree

    ○ Neither agree nor disagree

    ○ Somewhat disagree

    ○ Strongly disagree

26. How important is the following information about cloud deletion. On scale (Extremely important & Very important & Moderately important & Slightly important & Not at all important)

    - Info on deleting from shared folders

- Info on deletion of copies
- Info on how to delete using different devices
- Info on how to delete from trash can
- Info on how how long data gets deleted
- Info on recovery after deletion
- Info on who has access to deleted information

27. Cloud interfaces provide enough guidance (as needed) to help me delete from the cloud. On scale (Strongly agree, Somewhat agree, Neither agree nor disagree, Somewhat disagree, Strongly disagree)

    - Cloud interfaces provide enough guidance (as needed) to help me delete.
    - Cloud services have all the features I need to delete from the cloud.
    - Cloud apps in my smartphone have all the features I need to delete from the cloud.
    - Information on deletion is enough and available for me all the time.
    - I think cloud computing allows me to do what I need to.
    - Cloud storage services completely delete my data when I delete.

References

I. M. Abbadi, J. Lyle, et al. Challenges for provenance in cloud computing. In *TaPP*, 2011.

R. Abu-Salma, M. A. Sasse, J. Bonneau, A. Danilova, A. Naiakshina, and M. Smith. Obstacles to the adoption of secure communication tools. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 137–153. IEEE, 2017.

A. Adams and M. A. Sasse. Users are not the enemy. *Communications of the ACM*, 1999. ISSN 00010782. doi: 10.1145/322796.322806.

N. Aharony. An exploratory study on factors affecting the adoption of cloud computing by information professionals. *The Electronic Library*, 33(2):308–323, 2015.

S. Ahmed and M. Haag. Entering the field: Decisions of an early career researcher adopting classic grounded theory. *Grounded Theory Review*, 15(2), 2016.

A. Albeshri, C. Boyd, and J. G. Nieto. Geoproof: proofs of geographic location for cloud computing environment. In *Distributed Computing Systems Workshops (ICDCSW), 2012 32nd International Conference on*, pages 506–514. IEEE, 2012.

H. Almuhimedi, S. Wilson, B. Liu, N. Sadeh, and A. Acquisti. Tweets are forever: a large-scale quantitative analysis of deleted tweets. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 897–908. ACM, 2013.

H. Almuhimedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, and Y. Agarwal. Your location has been shared 5,398 times!: A field study on mobile app privacy nudging. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 787–796. ACM, 2015.

P. Anthonysamy, P. Greenwood, and A. Rashid. A method for analysing traceability between privacy policies and privacy controls of online social networks. In *Privacy Technologies and Policy - First Annual Privacy Forum, APF 2012, Limassol, Cyprus, October 10-11, 2012, Revised Selected Papers*, pages 187–202, 2012. doi: 10.1007/978-3-642-54069-1_12. URL https://doi.org/10.1007/978-3-642-54069-1_12.

P. Anthonysamy, P. Greenwood, and A. Rashid. Social networking privacy: Understanding the disconnect from policy to controls. *Computer*, 46(6):60–67, June 2013. ISSN 0018-9162. doi: 10.1109/MC.2012.326.

J. Arnold. *OpenStack Swift: Using, Administering, and Developing for Swift Object Storage*. O'Reilly Media, 2014. ISBN 9781491903872. URL https://books.google.co.uk/books?id=ALHXBAAAQBAJ.

G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song. Provable data possession at untrusted stores. In *Proceedings of the 14th ACM conference on Computer and communications security*, pages 598–609. Acm, 2007.

O. Ayalon and E. Toch. Retrospective privacy: Managing longitudinal privacy in online social networks. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, page 4. ACM, 2013.

M. Backes and F. Bendun. Poster: Forcing the cloud to forget by attesting data deletion. 2015.

M. Barhamgi, A. K. Bandara, Y. Yu, K. Belhajjame, and B. Nuseibeh. On Protecting Privacy in the Cloud. *IEEE Computer*, 2016a.

M. Barhamgi, A. K. Bandara, Y. Yu, K. Belhajjame, and B. Nuseibeh. Protecting privacy in the cloud: Current practices, future directions. *Computer*, 49(2):68–72, 2016b.

A. F. Barsoum and M. A. Hasan. Provable possession and replication of data over cloud servers. *Centre For Applied Cryptographic Research (CACR), University of Waterloo, Report*, 32:2010, 2010.

L. Bauer, L. F. Cranor, S. Komanduri, M. L. Mazurek, M. K. Reiter, M. Sleeper, and B. Ur. The Post Anachronism: The Temporal Dimension of Facebook Privacy. *Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society - WPES '13*, 2013. ISSN 15437221. doi: 10.1145/2517840.2517859.

M. Bellare, A. Boldyreva, and A. O'Neill. Deterministic and efficiently searchable encryption. In *Annual International Cryptology Conference*, pages 535–552. Springer, 2007.

M. Bennett. A review of the literature on the benefits and drawbacks of participatory action research. *First Peoples Child & Family Review*, 1(1):19–32, 2004.

K. Benson, R. Dowsley, and H. Shacham. Do you know where your cloud files are? In *Proceedings of the 3rd ACM workshop on Cloud computing security workshop*, pages 73–82. ACM, 2011.

J. Bergold and S. Thomas. Participatory research methods: A methodological approach in motion. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 13(1), 2012. ISSN 1438-5627. doi: 10.17169/fqs-13.1.1801. URL http://www.qualitative-research.net/index.php/fqs/article/view/1801.

A. F. Blackwell. The reification of metaphor as a design tool. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(4):490–530, 2006.

S. J. Blanchard and I. Banerji. Evidence-based recommendations for designing free-sorting experiments. *Behavior research methods*, 48(4):1318–1336, 2016.

J. Blythe and L. J. Camp. Implementing mental models. In *2012 IEEE symposium on Security and privacy workshops*, pages 86–90. IEEE, 2012.

J. Blythe, J. Camp, and V. Garg. Targeted Risk Communication for Computer Security. *Proceedings of the 16th International Conference on Intelligent User Interfaces. ACM, 2011.*, 2011. doi: 10.1145/1943403.1943449.

G. A. Bowen. Naturalistic inquiry and the saturation concept: a research note. *Qualitative Research*, 8(1):137–152, 2008. doi: 10.1177/1468794107085301. URL https://doi.org/10.1177/1468794107085301.

K. D. Bowers, A. Juels, and A. Oprea. Hail: A high-availability and integrity layer for cloud storage. In *Proceedings of the 16th ACM conference on Computer and communications security*, pages 187–198. ACM, 2009a.

K. D. Bowers, A. Juels, and A. Oprea. Proofs of retrievability: Theory and implementation. In *Proceedings of the 2009 ACM workshop on Cloud computing security*, pages 43–54. ACM, 2009b.

K. D. Bowers, M. Van Dijk, A. Juels, A. Oprea, and R. L. Rivest. How to tell if your cloud files are vulnerable to drive crashes. In *Proceedings of the 18th ACM conference on Computer and communications security*, pages 501–514. ACM, 2011.

V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

C. Bravo-Lillo, S. Komanduri, L. F. Cranor, R. W. Reeder, M. Sleeper, J. Downs, and S. Schechter. Your Attention Please: Designing Security-decision UIs to Make Genuine Risks Harder to Ignore. *SOUPS '13: Proceedings of the Ninth Symposium on Usable Privacy and Security*, 2013. doi: 10.1145/2501604.2501610.

A. Bryman. *Social research methods*. Oxford university press, 2015.

L. Burkon. Quality of service attributes for software as a service. *Journal of Systems Integration*, 4(3):38–47, 2013.

C. Cachin, K. Haralambiev, H.-C. Hsiao, and A. Sorniotti. Policy-based secure deletion. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 259–270. ACM, 2013.

L. J. Camp. Mental models of privacy and security. *IEEE Technology and Society Magazine*, 2009a. ISSN 02780097. doi: 10.1109/MTS.2009.934142.

L. J. Camp. Mental models of privacy and security. *IEEE Technology and society magazine*, 28(3), 2009b.

R. H. Campbell, M. Montanari, and R. Farivar. A middleware for assured clouds. *Journal of Internet Services and Applications*, 3(1):87–94, 2012.

R. H. Campbell, C. A. Kamhoua, and K. A. Kwiat. *Assured Cloud Computing*. John Wiley & Sons, 2018.

N. Cao, C. Wang, M. Li, K. Ren, and W. Lou. Privacy-preserving multi-keyword ranked search over encrypted cloud data. *IEEE Transactions on parallel and distributed systems*, 25(1):222–233, 2014.

R. Capra, E. Vardell, and K. Brennan. File synchronization and sharing: User practices and challenges. *Proceedings of the Association for Information Science and Technology*, 51 (1):1–10, 2014.

L. Chaoling, C. Yue, and Z. Yanzhou. A data assured deletion scheme in cloud storage. *China Communications*, 11(4):98–110, 2014.

K. Charmaz. *Constructing grounded theory: A practical guide through qualitative analysis*. Sage, 2006.

K. Charmaz. *Constructing grounded theory*. Sage, 2014.

Cisco. *Cisco Global Cloud Index: Forecast and Methodology, 2016–2021 White Paper*, 2018 (accessed August 15, 2018). URL https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.html.

J. W. Clark, P. Snyder, D. McCoy, and C. Kanich. I saw images i didn't even know i had: Understanding user perceptions of cloud storage privacy. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1641–1644. ACM, 2015.

J. Curn. *How a bug in Dropbox permanently deleted my 8000 photos*, 2014 (accessed August 15, 2018). URL https://medium.com/@jancurn/how-bug-in-dropbox-permanently-deleted-my-8000-photos-cb7dcf13647b.

R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky. Searchable symmetric encryption: improved definitions and efficient constructions. *Journal of Computer Security*, 19(5): 895–934, 2011.

S. Diesburg, C. Meyers, M. Stanovich, M. Mitchell, J. Marshall, J. Gould, A.-I. A. Wang, and G. Kuenning. Trueerase: Per-file secure deletion for the storage data path. In *Proceedings of the 28th Annual Computer Security Applications Conference*, pages 439–448. ACM, 2012.

S. Diesburg, C. Meyers, M. Stanovich, A.-I. A. Wang, and G. Kuenning. Trueerase: Leveraging an auxiliary data path for per-file secure deletion. *ACM Transactions on Storage (TOS)*, 12(4):18, 2016.

S. M. Diesburg and A.-I. A. Wang. A survey of confidential data storage and deletion methods. *ACM Computing Surveys*, 2010. ISSN 03600300. doi: 10.1145/1824795.1824797.

A. Dix, J. E. Finlay, G. D. Abowd, and R. Beale. *Human-Computer Interaction (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2003. ISBN 0130461091.

C. Dong, G. Russello, and N. Dulay. Shared and searchable encrypted data for untrusted servers. *Journal of Computer Security*, 19(3):367–397, 2011.

X. Dong, W. Zhang, X. Hu, and K. Liu. A cloud-user watermarking protocol protecting the right to be forgotten for the outsourced plain images. *Int. J. Digit. Crime For.*, 10 (4):118–139, Oct. 2018. ISSN 1941-6210. doi: 10.4018/IJDCF.2018100109. URL https://doi.org/10.4018/IJDCF.2018100109.

P. Dourish, E. Grinter, J. Delgado De La Flor, and M. Joseph. Security in the wild: user strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing*, 8(6):391–401, 2004.

J. B. Earp, A. I. Anton, L. Aiman-Smith, and W. H. Stufflebeam. Examining internet privacy policies within the context of user privacy values. *IEEE Transactions on Engineering Management*, 52(2):227–237, May 2005. ISSN 0018-9391. doi: 10.1109/TEM.2005.844927.

C. C. Erway, A. Küpçü, C. Papamanthou, and R. Tamassia. Dynamic provable data possession. *ACM Transactions on Information and System Security (TISSEC)*, 17(4):15, 2015.

J. Fan and F. Vercauteren. Somewhat practical fully homomorphic encryption. *IACR Cryptology ePrint Archive*, 2012:144, 2012.

T. Fifield, D. Fleming, A. Gentle, L. Hochstein, J. Proulx, E. Toews, and J. Topjian. *OpenStack Operations Guide*. O'Reilly Media, 2014. ISBN 9781491906309. URL https://books.google.co.uk/books?id=jQ5pAwAAQBAJ.

J. L. Fleiss, B. Levin, and M. C. Paik. *Statistical methods for rates and proportions*. John Wiley & Sons, 2013.

D. Freed, J. Palmer, D. Minchala, K. Levy, T. Ristenpart, and N. Dell. Digital Technologies and Intimate Partner Violence: A Qualitative Analysis with Multiple Stakeholders. *Proc. ACM Hum.-Comput. Interact*, 2017. ISSN 2573-0142. doi: 10.1145/3134681.

S. Furnell. Why users cannot use security. *Computers &amp; Security*, 24(4):274–279, 2005.

GDPR. *Article 17*, 2018 (accessed October 15, 2018). URL http://www.privacy-regulation.eu/en/article-17-right-to-erasure-{%}27right-to-be-forgotten{%}27-GDPR.htm.

R. Geambasu, T. Kohno, A. A. Levy, and H. M. Levy. Vanish: Increasing data privacy with self-destructing data. In *USENIX Security Symposium*, pages 299–316, 2009.

B. G. Glaser and A. L. Strauss. *The discovery of grounded theory: Strategies for qualitative research*. Transaction publishers, 2009.

B. G. Glaser et al. The discovery of grounded theory strategies for qualitative research. Technical report, 1967.

L. Gou, M. X. Zhou, and H. Yang. Knowme and shareme: understanding automatically discovered personality traits from social media and user sharing preferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 955–964. ACM, 2014.

A. B. Habib, T. Khanam, and R. Palit. Simplified file assured deletion (sfade)-a user friendly overlay approach for data security in cloud storage system. In *Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on*, pages 1640–1644. IEEE, 2013.

S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In *Proceedings of the 18th ACM Conference on Computer and Communications Security*, CCS '11, pages 491–500, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0948-6. doi: 10.1145/2046707.2046765. URL http://doi.acm.org/10.1145/2046707.2046765.

J. Hallett and D. Aspinall. Apppal for android. In *Proceedings of the 8th International Symposium on Engineering Secure Software and Systems - Volume 9639*, ESSoS 2016, pages 216–232, Berlin, Heidelberg, 2016. Springer-Verlag. ISBN 978-3-319-30805-0. doi: 10.1007/978-3-319-30806-7_14. URL https://doi.org/10.1007/978-3-319-30806-7_14.

A. Henry. *Scan and Save Images of Your Passport and Prescriptions When Traveling*, (Last accessed Sep 14, 2018). URL https://lifehacker.com/scan-and-save-images-of-your-passport-and-prescriptions-927527185.

J. Huang and D. M. Nicol. Trust mechanisms for cloud computing. *Journal of Cloud Computing: Advances, Systems and Applications*, 2(1):9, 2013.

M. E. Hussein, S. Hirst, V. Salyers, and J. Osuji. Using grounded theory as a method of inquiry: Advantages and disadvantages. *The Qualitative Report*, 19(27):1–15, 2014.

P. Ilia, I. Polakis, E. Athanasopoulos, F. Maggi, and S. Ioannidis. Face/off: Preventing privacy leakage from photos in social networks. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 781–792. ACM, 2015.

I. Ion, N. Sachdeva, P. Kumaraguru, and S. Čapkun. Home is safer than the cloud! Privacy Concerns for Consumer Cloud Storage. *Soups '11*, 2011. ISSN 1450309119. doi: 10.1145/2078827.2078845.

I. Ion, R. Reeder, and S. Consolvo. "... no one can hack my mind": Comparing expert and non-expert security practices. In *SOUPS*, volume 15, pages 1–20, 2015.

G. Irazoqui, T. Eisenbarth, and B. Sunar. A shared cache attack that works across cores and defies vm sandboxing–and its application to aes. In *Security and Privacy (SP), 2015 IEEE Symposium on*, pages 591–604. IEEE, 2015.

Q. Ismail, T. Ahmed, A. Kapadia, and M. K. Reiter. Crowdsourced exploration of security configurations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 467–476. ACM, 2015.

Q. Ismail, T. Ahmed, K. Caine, A. Kapadia, and M. Reiter. To permit or not to permit, that is the usability question: Crowdsourcing mobile apps' privacy permission settings. *Proceedings on Privacy Enhancing Technologies*, 2017(4):119–137, 2017.

Y. Jabareen. Building a conceptual framework: philosophy, definitions, and procedure. *International journal of qualitative methods*, 8(4):49–62, 2009.

S. Jasanoff, G. E. Markle, J. C. Peterson, and T. Pinch. *Handbook of science and technology studies*. Sage publications, 2001.

M. Johnson, S. Egelman, and S. M. Bellovin. Facebook and privacy: it's complicated. In *Proceedings of the eighth symposium on usable privacy and security*, page 9. ACM, 2012.

P. N. Johnson-Laird. Mental models and human reasoning. *Proceedings of the National Academy of Sciences of the United States of America*, 2010. ISSN 1091-6490. doi: 10.1073/pnas.1012933107.

K. Jonsen and K. A. Jehn. Using triangulation to validate themes in qualitative studies. *Qualitative Research in Organizations and Management: An International Journal*, 4(2): 123–150, 2009.

A. Juels and B. S. Kaliski Jr. Pors: Proofs of retrievability for large files. In *Proceedings of the 14th ACM conference on Computer and communications security*, pages 584–597. Acm, 2007.

S. Kamara and K. Lauter. Cryptographic cloud storage. In *International Conference on Financial Cryptography and Data Security*, pages 136–149. Springer, 2010.

P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder. A nutrition label for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, page 4. ACM, 2009.

M. T. Khan, M. Hyun, C. Kanich, and B. Ur. Forgotten but not gone: Identifying the need for longitudinal data management in cloud storage. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 543. ACM, 2018.

N. J. King and V. Raja. Protecting the privacy and security of sensitive customer data in the cloud. *Computer Law & Security Review*, 28(3):308 – 319, 2012. ISSN 0267-3649. doi: https://doi.org/10.1016/j.clsr.2012.03.003. URL http://www.sciencedirect.com/science/article/pii/S0267364912000556.

F. J. Krautheim. Private virtual infrastructure for cloud computing. In *HotCloud*, 2009.

R. L. Krutz and R. D. Vines. *Cloud security: A comprehensive guide to secure cloud computing*. Wiley Publishing, 2010.

B. L Berg. Qualitative research methods for the social sciences. 2001.

J. Lai, J. Xiong, C. Wang, G. Wu, and Y. Li. A secure cloud backup system with deduplication and assured deletion. In *International Conference on Provable Security*, pages 74–83. Springer, 2017.

L. Lee, J. Lee, S. Egelman, and D. Wagner. Information disclosure concerns in the age of wearable computing. In *Proceedings of the 2016 Workshop on Usable Security*, 2016.

P. G. Leon, B. Ur, Y. Wang, M. Sleeper, R. Balebako, R. Shay, L. Bauer, M. Christodorescu, and L. F. Cranor. What matters to users?: factors that affect users' willingness to share information with online advertisers. In *Proceedings of the ninth symposium on usable privacy and security*, page 7. ACM, 2013.

B. Liu, M. S. Andersen, F. Schaub, H. Almuhimedi, S. Zhang, N. Sadeh, A. Acquisti, and Y. Agarwal. Follow my recommendations: A personalized privacy assistant for mobile app permissions. In *Symposium on Usable Privacy and Security*, 2016a.

F. Liu, Y. Yarom, Q. Ge, G. Heiser, and R. B. Lee. Last-level cache side-channel attacks are practical. In *Security and Privacy (SP), 2015 IEEE Symposium on*, pages 605–622. IEEE, 2015.

R. Liu, J. Cao, K. Zhang, W. Gao, J. Liang, and L. Yang. When privacy meets usability: unobtrusive privacy permission recommendation system for mobile apps based on crowdsourcing. *IEEE Transactions on Services Computing*, 2016b.

Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 61–70. ACM, 2011.

R. Lu, X. Lin, X. Liang, and X. S. Shen. Secure provenance: the essential of bread and butter of data forensics in cloud computing. In *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, pages 282–292. ACM, 2010.

Y. Luo, M. Xu, S. Fu, and D. Wang. Enabling assured deletion in the cloud storage by overwriting. In *Proceedings of the 4th ACM International Workshop on Security in Cloud Computing*, pages 17–23. ACM, 2016.

K. M. MacQueen, E. McLellan, K. Kay, and B. Milstein. Codebook development for team-based qualitative analysis. *CAM Journal*, 10(2):31–36, 1998.

D. C. Marinescu. *Cloud Computing: Theory and Practice*. Elsevier Science, 2017. ISBN 9780128128114. URL https://books.google.co.uk/books?id=O9smDwAAQBAJ.

C. Marshall and J. C. Tang. That syncing feeling: early user experiences with the cloud. In *Proceedings of the Designing Interactive Systems Conference*, pages 544–553. ACM, 2012.

P. Massa, C. Leonardi, B. Lepri, F. Pianesi, and M. Zancanaro. If you are happy and you know it, say "i'm here": investigating parents' location-sharing preferences. In *Human-Computer Interaction*, pages 315–332. Springer, 2015.

C. Massey, T. Lennig, and S. Whittaker. Cloudy forecast: an exploration of the factors underlying shared repository use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2461–2470. ACM, 2014.

T. Matthews, K. O'Leary, A. Turner, M. Sleeper, J. P. Woelfer, M. Shelton, C. Manthorne, E. F. Churchill, and S. Consolvo. Stories from survivors: Privacy & security practices when coping with intimate partner abuse. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2189–2201. ACM, 2017.

M. L. Mazurek, J. Arsenault, J. Bresee, N. Gupta, I. Ion, C. Johns, D. Lee, Y. Liang, J. Olsen, B. Salmon, et al. Access control for home data sharing: Attitudes, needs and practices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 645–654. ACM, 2010.

A. M. McDonald and L. F. Cranor. The cost of reading privacy policies. *ISJLP*, 4:543, 2008.

W. Melicher, M. Sharif, J. Tan, L. Bauer, M. Christodorescu, and P. G. Leon. (do not) track me sometimes: users' contextual preferences for web tracking. *Proceedings on Privacy Enhancing Technologies*, 2016(2):135–154, 2016.

R. C. Merkle. Protocols for public key cryptosystems. In *1980 IEEE Symposium on Security and Privacy*, pages 122–122. IEEE, 1980.

K. Micinski, D. Votipka, R. Stevens, N. Kofinas, M. L. Mazurek, and J. S. Foster. User interactions and permission use on android. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 362–373, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4655-9. doi: 10.1145/3025453.3025706. URL http://doi.acm.org/10.1145/3025453.3025706.

G. Misra and J. M. Such. How socially aware are social media privacy controls? *Computer*, 49(3):96–99, 2016.

G. Misra and J. M. Such. React: Recommending access control decisions to social media users. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pages 421–426. ACM, 2017.

Z. Mo, Y. Qiao, and S. Chen. Two-party fine-grained assured deletion of outsourced data in cloud systems. In *2014 IEEE 34th International Conference on Distributed Computing Systems (ICDCS)*, pages 308–317. IEEE, 2014a.

Z. Mo, Q. Xiao, Y. Zhou, and S. Chen. On deletion of outsourced data in cloud computing. In *2014 IEEE 7th International Conference on Cloud Computing*, pages 344–351. IEEE, 2014b.

M. Mondal, J. Messias, S. Ghosh, K. Gummadi, and A. Kate. Longitudinal privacy management in social media: The need for better controls. *IEEE Internet Computing*, 2017.

A. Murillo, A. Kramm, S. Schnorf, and A. D. Luca. "if i press delete, it's gone" - user understanding of online data deletion and expiration. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, pages 329–339, Baltimore, MD, 2018. USENIX Association. ISBN 978-1-931971-45-4. URL https://www.usenix.org/conference/soups2018/presentation/murillo.

M. Naehrig, K. Lauter, and V. Vaikuntanathan. Can homomorphic encryption be practical? In *Proceedings of the 3rd ACM workshop on Cloud computing security workshop*, pages 113–124. ACM, 2011.

M. Nebeling, M. Geel, O. Syrotkin, and M. C. Norrie. Mubox: Multi-user aware personal cloud storage. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1855–1864. ACM, 2015.

K. A. Neuendorf. *The content analysis guidebook*. Sage, 2016.

L. R. Newton. Data-logging in practical science: research and reality. *International Journal of Science Education*, 22(12):1247–1259, 2000.

J. Nielsen. Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 152–158. ACM, 1994.

J. Nielsen. *Mental Models*, 2010 (accessed June 1, 2019). URL https://www.nngroup.com/articles/mental-models/.

J. Nielsen. *Mental Models*, 2019 (accessed June 1, 2019). URL https://www.interaction-design.org/literature/article/a-very-useful-work-of-fiction-mental-models-in-design.

J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256. ACM, 1990.

D. Norman. *The design of everyday things: Revised and expanded edition*. Constellation, 2013.

L. S. Nowell, J. M. Norris, D. E. White, and N. J. Moules. Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1): 1609406917733847, 2017. doi: 10.1177/1609406917733847. URL https://doi.org/10.1177/1609406917733847.

N. Nthala and I. Flechais. "if it's urgent or it is stopping me from doing something, then i might just go straight at it": a study into home data security decisions. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*, pages 123–142. Springer, 2017.

D. Ocean. *Data leakage*, 2014 (accessed August 14, 2018). https://www.digitalocean.com/company/blog/resolved-lvm-data-issue/.

J. S. Olson, J. Grudin, and E. Horvitz. A study of preferences for sharing and privacy. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1985–1988. ACM, 2005.

K. E. Olson, M. A. O'Brien, W. A. Rogers, and N. Charness. Diffusion of technology: Frequency of use for younger and older adults. *Ageing International*, 36(1):123–145, Mar 2011. ISSN 1936-606X. doi: 10.1007/s12126-010-9077-9. URL https://doi.org/10.1007/s12126-010-9077-9.

OpenStack. *OpenStack High Availability*, 2018 (accessed May 25, 2019)a. https://docs.openstack.org/ha-guide/.

OpenStack. *OpenStack Image and Instances*, 2018 (accessed May 25, 2019)b. https://docs.openstack.org/glance/pike/admin/troubleshooting.html.

OpenStack. *OpenStack: Data privacy concerns*, 2019 (accessed May 25, 2019)a. https://docs.openstack.org/security-guide/tenant-data/data-privacy-concerns.html.

OpenStack. *OpenStack Reaper*, 2019 (accessed May 25, 2019)b. https://docs.openstack.org/swift/pike/overview_reaper.html.

OpenStack. *OpenStack Official*, 2019 (accessed May 25, 2019)c. http://www.openstack.org/.

I. Papagiannis and P. Pietzuch. Cloudfilter: practical control of sensitive data propagation to the cloud. In *Proceedings of the 2012 ACM Workshop on Cloud computing security workshop*, pages 97–102. ACM, 2012.

V. Pappas, V. P. Kemerlis, A. Zavou, M. Polychronakis, and A. D. Keromytis. Cloudfence: Data flow tracking as a cloud service. In *Research in Attacks, Intrusions, and Defenses*, pages 411–431. Springer, 2013.

T. Pasquier, J. Bacon, J. Singh, and D. Eyers. Data-centric access control for cloud computing. In *Symposium on Access Control Models and Technologies*. ACM, ACM, 2016.

D. F. Polit and C. T. Beck. Generalization in quantitative and qualitative research: Myths and strategies. *International journal of nursing studies*, 47(11):1451–1458, 2010.

I. Pollach. What's wrong with online privacy policies? *Communications of the ACM*, 50(9): 103–108, 2007.

S. Portigal. *Interviewing users*. Rosenfeld Media, 2013.

C. Priebe, D. Muthukumaran, D. O'Keeffe, D. Eyers, B. Shand, R. Kapitza, and P. Pietzuch. Cloudsafetynet: Detecting data leakage between cloud tenants. In *Proceedings of the 6th edition of the ACM Workshop on Cloud Computing Security*, pages 117–128. ACM, 2014.

E. Rader. Yours, mine and (not) ours: social influences on group information repositories. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2095–2098. ACM, 2009.

E. Rader and J. Slaker. The importance of visibility for folk theories of sensor data. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 257–270, Santa Clara, CA, 2017. USENIX Association. ISBN 978-1-931971-39-3. URL https://www.usenix.org/conference/soups2017/technical-sessions/presentation/rader.

A. Rahumed, H. C. Chen, Y. Tang, P. P. Lee, and J. C. Lui. A secure cloud backup system with assured deletion and version control. In *Proceedings of the International Conference on Parallel Processing Workshops*, 2011. ISBN 9780769545110. doi: 10.1109/ICPPW.2011.17.

J. Reardon, S. Capkun, and D. Basin. Data node encrypted file system: Efficient secure deletion for flash memory. In *Proceedings of the 21st USENIX conference on Security symposium*, pages 17–17. USENIX Association, 2012.

J. Reardon, D. Basin, and S. Capkun. Sok: Secure data deletion. In *Security and Privacy (SP), 2013 IEEE Symposium on*, pages 301–315. IEEE, 2013.

J. Reardon, D. Basin, and S. Capkun. On secure data deletion. *Security & Privacy, IEEE*, 12 (3):37–44, 2014.

S. Reeves, A. Kuper, and B. D. Hodges. Qualitative research methodologies: ethnography. *Bmj*, 337:a1020, 2008.

M. Reilly. *Is Facebook Targeting Ads at Sad Teens?*, 2017 (accessed October 22, 2018). https://www.technologyreview.com/s/604307/is-facebook-targeting-ads-at-sad-teens/.

T. Ristenpart, E. Tromer, H. Shacham, and S. Savage. Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. In *Proceedings of the 16th ACM conference on Computer and communications security*, pages 199–212. ACM, 2009.

J. Rosen. The right to be forgotten. *Stan. L. Rev. Online*, 64:88, 2011.

K. Roulston. *Reflective interviewing: A guide to theory and practice*. Sage, 2010.

W. B. Rouse and N. M. Morris. On looking into the black box: Prospects and limits in the search for mental models. *Psychological bulletin*, 100(3):349, 1986.

S. Ruoti, J. Andersen, T. Hendershot, D. Zappala, and K. Seamons. Helping johnny understand and avoid mistakes: A comparison of automatic and manual encryption in email. *CoRR*, 2015.

G. W. Ryan and H. R. Bernard. Techniques to identify themes. *Field methods*, 15(1):85–109, 2003.

N. Santos, K. P. Gummadi, and R. Rodrigues. Towards trusted cloud computing. *HotCloud*, 9(9):3, 2009.

F. Schaub, R. Balebako, and L. F. Cranor. Designing effective privacy notices and controls. *IEEE Internet Computing*, 21(3):70–77, May 2017. ISSN 1089-7801. doi: 10.1109/MIC.2017.75.

Z. Shen, L. Li, F. Yan, and X. Wu. Cloud computing system based on trusted computing platform. In *Intelligent Computation Technology and Automation (ICICTA), 2010 International Conference on*, volume 1, pages 942–945. IEEE, 2010.

S. Sheng, L. Broderick, C. A. Koranda, and J. J. Hyland. Why johnny still can't encrypt: evaluating the usability of email encryption software. In *Symposium On Usable Privacy and Security*, pages 3–4, 2006.

R. Singh, S. Kumar, and S. K. Agrahari. Ensuring data storage security in cloud computing. *IOSR Journal of Engineering*, 2(12):17–21, 2012.

M. Sleeper, J. Cranshaw, P. G. Kelley, B. Ur, A. Acquisti, L. F. Cranor, and N. Sadeh. 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*, pages 3277–3286. ACM, 2013.

M. Sleeper, W. Melicher, H. Habib, L. Bauer, L. F. Cranor, and M. L. Mazurek. Sharing Personal Content Online: Exploring Channel Choice and Multi-Channel Behaviors. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016. doi: 10.1145/2858036.2858170.

R. H. Sloan and R. Warner. Beyond notice and choice: Privacy, norms, and consent. *J. High Tech. L.*, 14:370, 2014.

S. W. Smith. Humans in the loop: Human-computer interaction and security. *IEEE Security &amp; privacy*, 1(3):75–79, 2003.

H. Srinivas. *Is it safe to store personal IDs like scanned copies of passport, on Google drive?*, (last accessed Sep 11, 2018). URL https://www.quora.com/Is-it-safe-to-store-personal-IDs-like-scanned-copies-of-passport-on-Google-drive.

K.-J. Stol, P. Ralph, and B. Fitzgerald. Grounded theory in software engineering research: a critical review and guidelines. In *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*, pages 120–131. IEEE, 2016.

J. M. Such and N. Criado. Resolving Multi-Party Privacy Conflicts in Social Media. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1851–1863, 2016. ISSN 10414347. doi: 10.1109/TKDE.2016.2539165.

G. Sweeten, E. Sillence, and N. Neave. Digital hoarding behaviours: Underlying motivations and potential negative consequences. *Computers in Human Behavior*, 85:54–60, 2018.

B. Tang, Z. Chen, G. Hefferman, S. Pei, T. Wei, H. He, and Q. Yang. Incorporating intelligence in fog computing for big data analysis in smart cities. *IEEE Transactions on Industrial informatics*, 13(5):2140–2150, 2017.

J. C. Tang, J. R. Brubaker, and C. C. Marshall. What do you see in the cloud? understanding the cloud-based user experience through practices. In *IFIP Conference on Human-Computer Interaction*, pages 678–695. Springer, 2013.

Y. Tang, P. P. Lee, J. C. Lui, and R. Perlman. Fade: Secure overlay cloud storage with file assured deletion. In *International Conference on Security and Privacy in Communication Systems*, pages 380–397. Springer, 2010.

Y. Tang, P. P. Lee, J. C. Lui, R. Perlman, et al. Secure overlay cloud storage with access control and assured deletion. *IEEE transactions on dependable and secure computing*, 9 (6):903–916, 2012.

UMASS. *Screening Participants*, 2015 (accessed June 1, 2019). URL https://www.umass.edu/research/guidance/screening-activities-used-determine-eligibility-participation-research.

B. Ur, J. Bees, S. M. Segreti, L. Bauer, N. Christin, and L. F. Cranor. Do users' perceptions of password security match reality? In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 3748–3760, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3362-7. doi: 10.1145/2858036.2858546. URL http://doi.acm.org/10.1145/2858036.2858546.

A. Valdez. *Everything You Need to Know About Facebook and Cambridge Analytica*, 2018 (accessed October 22, 2018). https://www.wired.com/story/wired-facebook-cambridge-analytica-coverage/.

M. Van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan. Fully homomorphic encryption over the integers. In *Advances in cryptology–EUROCRYPT 2010*, pages 24–43. Springer, 2010.

P. Van Liesdonk, S. Sedghi, J. Doumen, P. Hartel, and W. Jonker. Computationally efficient searchable symmetric encryption. In *Workshop on Secure Data Management*, pages 87–100. Springer, 2010.

G. Venkatadri, E. Lucherini, P. Sapiezynski, and A. Mislove. Investigating sources of pii used in facebook's targeted advertising. *Proceedings on Privacy Enhancing Technologies*, 1:18, 2018.

A. Voida, J. S. Olson, and G. M. Olson. Turbulence in the clouds: challenges of cloud-based information work. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2273–2282. ACM, 2013.

S. Voida, W. K. Edwards, M. W. Newman, R. E. Grinter, and N. Ducheneaut. Share and share alike: exploring the user interface affordances of file sharing. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 221–230. ACM, 2006.

L.-N. Wang, Z.-W. Ren, R.-W. Yu, F. Han, and Y.-F. Dong. A data assured deletion approach adapted for cloud storage. *Dianzi Xuebao(Acta Electronica Sinica)*, 40(2):266–272, 2012.

Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li. Enabling public auditability and data dynamics for storage security in cloud computing. *IEEE transactions on parallel and distributed systems*, 22(5):847–859, 2011a.

Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. G. Leon, and L. F. Cranor. I regretted the minute I pressed share: A qualitative study of regrets on Facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, page 10. ACM, 2011b.

R. Wash. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security - SOUPS '10*, 2010. ISBN 9781450302647. doi: 10.1145/1837110.1837125.

R. Wash and E. Rader. Influencing mental models of security. In *Proceedings of the 2011 workshop on New security paradigms workshop - NSPW '11*, 2011. ISBN 9781450310789. doi: 10.1145/2073276.2073283.

G. J. Watson, R. Safavi-Naini, M. Alimomeni, M. E. Locasto, and S. Narayan. Lost: location based storage. In *Proceedings of the 2012 ACM Workshop on Cloud computing security workshop*, pages 59–70. ACM, 2012.

S. Weber, M. Harbach, and M. Smith. Participatory design for security-related user interfaces. In *Usable Security (NDSS USEC)*, 2015. URL http://www.internetsociety.org/sites/default/files/04_1_3.pdf.

WhatsAppInc. *Backing up to Google Drive*, (accessed June 11, 2019). URL https://faq.whatsapp.com/en/android/28000019/.

A. Whitten and J. D. Tygar. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In *Usenix Security*, volume 1999, 1999.

Y. Xu and M. J. Lee. Shopping as a social activity: Understanding people's categorical item sharing preferences on social networks. In *Companion Proceedings of the 23rd International on Intelligent User Interfaces: 2nd Workshop on Theory-Informed User Modeling for Tailoring and Personalizing Interfaces (HUMANIZE)*, 2018.

L. Xue, J. Ni, Y. Li, and J. Shen. Provable data transfer from provable data possession and deletion in cloud storage. *Computer Standards & Interfaces*, 54:46–54, 2017.

L. Xue, Y. Yu, Y. Li, M. H. Au, X. Du, and B. Yang. Efficient attribute-based encryption with attribute revocation for assured data deletion. *Information Sciences*, 2018.

Y. Yu, L. Xue, Y. Li, X. Du, M. Guizani, and B. Yang. Assured data deletion with fine-grained access control for fog-based industrial applications. *IEEE Transactions on Industrial Informatics*, 2018.

L. Zeng, Z. Shi, S. Xu, and D. Feng. Safevanish: An improved data self-destruction for protecting data privacy. In *2nd IEEE International Conference on Cloud Computing Technology and Science*, pages 521–528. IEEE, 2010.

O. Q. Zhang, M. Kirchberg, R. K. Ko, and B. S. Lee. How to track your data: The case for cloud computing provenance. In *2011 Third IEEE International Conference on Coud Computing Technology and Science*, pages 446–453. IEEE, 2011a.

Y. Zhang, A. Juels, A. Oprea, and M. K. Reiter. Homealone: Co-residency detection in the cloud via side-channel analysis. In *2011 IEEE symposium on security and privacy*, pages 313–328. IEEE, 2011b.

Y. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart. Cross-vm side channels and their use to extract private keys. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 305–316. ACM, 2012.