Karlstad Business School
Handelshögskolan vid Karlstads universitet

Erik Framner

A Configuration User Interface for Multi-Cloud Storage Based on Secret Sharing

An Exploratory Design Study

Information systems

Master's thesis, 30 ECTS credits

Examination date: 19-02-2019
Supervisor: John Sören Pettersson

Karlstad Business School
Karlstad University SE-651 88 Karlstad Sweden
Phone:+46 54 700 10 00 Fax: +46 54 700 14 97
E-mail: handels@kau.se www.hhk.kau.se

Abstract

Storing personal information in a secure and reliable manner may be crucial for organizational as well as private users. Encryption protects the confidentiality of data against adversaries but if the cryptographic key is lost, the information will not be obtainable for authorized individuals either. Redundancy may protect information against availability issues or data loss, but also comes with greater storage overhead and cost. Cloud storage serves as an attractive alternative to traditional storage as one is released from maintenance responsibilities and does not have to invest in in-house IT-resources. However, cloud adoption is commonly hindered due to privacy concerns.

Instead of relying on the security of a single cloud, this study aims to investigate the applicability of a multi-cloud solution based on Secret Sharing, and to identify suitable options and guidelines in a configuration user interface (UI). Interviews were conducted with technically skilled people representing prospective users, followed by walkthroughs of a UI prototype.

Although the solution would (theoretically) allow for employment of less “trustworthy” clouds without compromising the data confidentiality, the research results indicate that trust factors such as compliance with EU laws may still be a crucial prerequisite in order for users to utilize cloud services. Users may worry about cloud storage providers colluding, and the solution may not be perceived as adequately secure without the use of encryption. The configuration of the Secret Sharing parameters are difficult to comprehend even for technically skilled individuals and default values could/should be recommended to the user.

Keywords: Secret Sharing, multi-cloud, data storage, user interface, design, HCI, security, privacy, usability, trust, PRISMACLOUD

Acknowledgements

I want to give a great 'thank you' to my supervisor John Sören Pettersson - a man of much patience - for helping me to figure out how to structure my thesis and how to tie everything together. You helped me to find the end of a seemingly endless tunnel.

Furthermore, I would like to give a great thanks to researchers in the PRISMACLOUD project. When it came to finding participants, defining interview questions, designing a UI proposal, and interpreting the research results, I had the luxury of being able to consult with not only my supervisor but also Simone Fischer-Hübner, Ala Sarah Alaqra, and Thomas Lorünser.

Lastly, I like to thank my parents Ulrika Framner and Tommy Johannesson, sister Lisa Framner, classmate Daniel Lindegren, as well as colleague Farzaneh Karegar for all the encouragement and for helping me to keep my spirit up in times of doubt.

Thank you. Without your help and support, I would not have been able to finish this Master's thesis.

Contents

List of Figures

List of Tables

1. Introduction

Information is a resource that can be utilized, created or processed in relation to the work performed within businesses and organizations (Alter 2006). For instance, public authorities often handle personal data of citizens in connection with the delivery of public services/functions. Such data may be sensitive and therefore require special protection to prevent unauthorized disclosure (Brodies LLP n.d.). Furthermore, there are certain demands in terms of *availability*. That is, in order for the information to become a useful asset within the organization/business, it must be accessible when it is needed by stakeholders performing a particular task. The consequence from unavailability may not simply be inconvenience. For instance, if medical records are not available at hospitals, health care professionals may not be able to ensure that the patients will receive the appropriate care and medical errors may therefore ensue (Alter 2006).

Computerized information (such as digital text documents, images, audio, and videos) is typically stored in files and folders on a device's hard drive (Moran 2015). Hardware components are far from infallible and a failure may cause information and systems to become unreachable (Alter 2006). One means for ensuring data availability and system reliability is *redundancy* (i.e., storing the data set and application in multiple areas) (Bhowmik 2017).

In a traditional computing scenario, business organizations need to set up their own in-house IT infrastructure of hardware and software, which requires extensive capital expenditure and effort. Only big corporations are typically able to afford investments on massive amounts of on-premise storage. Furthermore, procurement of hardware components for an IT infrastructure is not a one-time investment, since purchased resources may become out-dated after a few years when more powerful devices emerge. Out-dated computing resources might make it difficult for organizations to work efficiently and to compete with other businesses on the market (Bhowmik 2017).

Cloud computing serves as an attractive alternative to traditional computing models since IT resources can be provisioned at a significantly reduced cost and effort (Lorünser et al. 2016). It allows users to be equipped with storage and computing capabilities without requiring monetary investments on in-house hardware and software (Krutz & Vines 2010). Furthermore, the users are released from maintenance responsibilities as the underlying IT infrastructure is managed by the cloud provider (Chandrasekaran 2014; Happe et al. 2017). They are always provided with the latest version of computing resources without having to install software upgrades, patches or device drivers (Bhowmik 2017).

However, despite abovementioned benefits, adoption of cloud solutions may be hindered due to concerns about privacy and security issues. Fewer management responsibilities also imply less user control when data is outsourced to the cloud (Singhal et al. 2013). In similarity to traditional computing resources, cloud-based solutions also represent a target for external threat such as hackers (Krutz & Vines 2010). The cloud provider may also impose a potential threat by intruding on the confidentiality of the customers' data (Fabian et al. 2015) or disclosing the information to a third party without the users' consent (Pearson 2013). Although the term "cloud" may mislead people to believe that services are somehow floating in the air, they are still operated on land. Thus, cloud services are subject to national/international laws, and the confidentiality of data may be compromised due to law enforced disclosure (Oshri et al. 2015). Moreover, in similarity to traditional storage, the physical location of clouds may suffer from disasters such as fire, floods and earthquakes, which may cause data to be unavailable or lost (Bauer & Adams 2012).

The EU Horizon 2020 project PRISMACLOUD ("PRIvacy and Security MAintaining services in the CLOUD") aimed at developing solutions for protecting sensitive data in cloud-based environments. The feasibility of proposed solutions was illustrated by implementing and evaluating pilots for various scenarios (Alaqra et al. 2017). In a use case related to the area of e-Government, PRISMACLOUD proposed a framework called Archistar for secure distributed storage of data in the cloud. The solution applies a *Secret Sharing* protocol to a *multi-cloud* setting. This implies that data is divided into N fragments – or "chunks" – which are distributed to separate cloud storage providers (CSPs). No single chunk discloses any details about the full data, and in order to reconstruct the information into its original state, a subset of k chunks is needed. Thus, based on an assumption that cloud providers will not collude, data will be protected against unauthorized disclosure.

To perform the data splitting/fragmentation and distribution of chunks, a form of *configuration* needs to be created in an Archistar interface. Traditionally, system configurations are complex and generally performed by "system administrators" with more technical expertise than ordinary users (Xu & Zhou 2015). In regards to cloud-based solutions, there are several recent reports of incidents where governmental data has been leaked due to a misconfiguration. For instance: An Amazon S3 bucket, utilized to store classified data of the US Army and National Security Agency (NSA), was discovered in September 27th 2017 to provide public access to 47 files and folders with "Top Secret" information such as private keys to distributed intelligence systems (O'Sullivan 2017). Similarly, in April 2018, it was noticed that the British and Canadian government had accidentally exposed confidential data (e.g., security plans as well as server passwords) while using the cloud-based project management website Trello. As a result of human error or carelessness, the platform's visibility settings had manually been changed from its default value "private" to "public". Consequently, data was published on "boards" that was available on the open web and easy to find via the Google search engine (Grauer 2018).

While the aforementioned examples represent single cloud services, the PRISMACLOUD-enabled solution includes a *multi-cloud* architecture. Multi-clouds may provide a higher level of security, but also comes with greater configuration complexity (Salman 2015). While system complexity may be a necessity to match the needs of users, *complicacy* should be avoided by eliminating elements of perceived confusion (Norman 2013).

In previous research, solutions that combine Secret Sharing with a multi-cloud architecture have been evaluated by focusing on factors such as:

- *Performance* (e.g., Balasaraswathi & Manikandan 2014; Bessani et al. 2013; Chen et al. 2014; Fabian et al. 2015),
- *Availability* (e.g., Bessani et al. 2013; Gu et al. 2014), and/or
- *Cost* (e.g., Bessani et al. 2013; Chen et al. 2014; Gu et al. 2014).

To the best of the thesis author's knowledge, the emphasis seldom lies on human factors and the perspective of the user. Thus, there are little clues as to how prospective users would perceive a solution like Archistar, where they would be in charge of the configuration of the Secret Sharing mechanism and geographical distribution of data chunks. Furthermore, a user interface for decision-making support is not frequently designed and evaluated to a context such as Archistar.

1.1 Purpose

This study aims to explore the applicability of a multi-cloud storage solution based on Secret Sharing for personal or organizational use. Moreover, the purpose is to propose guidelines for configuration options that should be available in a user interface to serve as a feasible and trusted solution for secure data storage in the cloud.

Archistar, developed in the PRISMACLOUD project, will be utilized as a starting point for the investigation. It will serve as an *example* and the research results may also apply to other systems/frameworks that combine Secret Sharing with a multi-cloud setting.

1.2 Research Questions

RQ1. What are suitable configuration options and guidelines for organizational or private users with different security requirements?

RQ2. What are relevant trust factors, unique advantages, and risks of a multi-cloud storage solution based on Secret Sharing that should/could be communicated to users?

1.3 Target Group

The thesis has two intended target groups, i.e.: (1) Researchers and developers with an interest in privacy- and security-enhancing solutions for cloud-based storage. (2) Prospective users of a remote storage solution that combines Secret Sharing with a multi-cloud.

1.4 Outline of Thesis

Chapter 1 introduces the topic of the thesis as well as the study's purpose and research questions.

Chapter 2 gives a more in-depth explanation of the fundamental concept (i.e., *cloud computing*, *privacy/security concerns*, *user trust*, *Secret Sharing*, as well as the notion of *multi-clouds*). The chapter ends with a summary of the study's problem background.

Chapter 3 describes the methodological and ethical considerations as well as the approach utilized to address the research questions. Interviews were conducted and a user interface (UI) prototype was created and evaluated by performing user walkthroughs.

Chapter 4 presents the result from the conducted interviews and walkthroughs. The emphasis is on topics or themes brought up by several of the respondents/participants rather than by a single respondent/participant.

Chapter 5 interprets the research results. It discusses suitable features and elements that should be changed in the proposed configuration UI before the product/system is finalized.

Chapter 6 draws conclusions from the research findings and answers the research questions. Furthermore, the limitations of the study are briefly described.

The Appendix includes an analysis of an e-Government use case (utilized as an example in this study), material used during the interviews and user walkthroughs as well as a description of design decisions made when creating the UI proposal.

2. Literature review

2.1 Cloud computing

In network diagrams and documentation of web-based architecture, the metaphor of "cloud" has typically been used as an abstraction of the complex infrastructure that makes up the *Internet* (Erl et al. 2013; Oshri et al. 2015; Velte et al. 2010). However, Erl et al. (2013) argue that a cloud and the Internet should be regarded as two separate concepts. Typically, a cloud is owned by an individual company and offers IT-resources as a metered service to its *customers*, while the Internet provides IT-resources that are open for access by the general public (i.e., not just people subscribed to a particular company's services). Furthermore, the two concepts are usually dedicated to providing different *types* of resources. A cloud environment offers resources in the form of back-end processing capabilities, whereas the Internet mainly provides IT resources that are web content-based (i.e., information published on websites via the World Wide Web). Fundamentally, the Internet constitutes a "network of networks" (Erl et al. 2013), while cloud computing can be viewed as an extensive "network of computers" as it is typically comprised of a large number of machines (Bhowmik 2017). Another significant difference between the Internet and the cloud is that the former *enables access* to services of the latter (Erl et al. 2013; Bhowmik 2017; Oshri et al. 2015).

Armbrust et al. (2010) claims that a "cloud" constitutes hardware and software of one or multiple *datacentres* that deliver services over the Internet, while the term "cloud computing" also encompasses the service(s) being delivered. However, although resources of a cloud are housed in such a facility, not *all* datacentres should necessarily be regarded as clouds. Armbrust et al. (2010) suggest that a small or medium-sized datacentre does not qualify as a cloud. Similarly, Bhowmik (2017) describe that resources of cloud computing are typically maintained in more than a single datacentre. While IT-resources of a "traditional" datacentre can be accessed within the organization's perimeter (i.e., network boundary) (Bhowmik 2017), the cloud data centres are designed for providing *remote* access to corresponding resources (Erl et al. 2013).

According to the US National Institute of Standards and Technology (Mell & Grance 2011), cloud computing can be defined as the following:

> "[…] a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." (Mell & Grance 2011:2)

In other words, cloud computing represents network-accessed resources (such as storage and applications) that are (1) made available on an *on-demand* basis (i.e., accessible whenever the user wants or needs it), (2) *shared* among multiple users rather than dedicated to a single one, and (3) primarily maintained/controlled by an entity *other than* the users. Moreover, according to Mell and Grance (2011:2), the model that is cloud computing comprises "five essential characteristics, three service models, and four deployment models" – all of which will be described below.

2.1.1 Essential Characteristics

Resource pooling: In contrast to traditional computing models, where IT resources have minimal or no inter-connection and are managed separately as independent environments, cloud computing resources are *pooled together* (Hurwitz et al. 2010; Bhowmik 2017). By utilizing a cloud provider's high-capability infrastructure, users/customers can eliminate the need for huge investments on in-

house IT-systems. A large and flexible resource pool is (or should be) established by the cloud provider in order to meet users/customers' needs, fulfil Service Level Agreements (SLA) and offer significant cost savings (Krutz & Vines 2010). Thus, cloud computing utilizes a "multi-tenant model" which implies that numerous unrelated users/customers (i.e., tenants) can be served simultaneously by a single pool/set of resources (Mell & Grance 2011).

Broad network access: The IT-resources can be reached via a network from various types of thin and thick client devices (such as mobile phones, tablets, laptops and desktop computers) (Mell & Grance 2011). Since the users/customers are not bound to a particular device – so long as it has an Internet connection – they can typically access the service regardless of where they are located in the world (Rittinghouse & Ransome 2010). High-bandwidth network communication links are (or should be) in place between the provider and the user/customer to ensure that cloud computing will serve as an effective alternative solution to in-house hardware and software (Krutz & Vines 2010).

Rapid elasticity: It is difficult for providers to foresee the needs of customers as the demand may shift abruptly, causing spikes or drops in usage of the offered services (Mather et al. 2009:8). Furthermore, the demand and frequency of use might differ from one customer to another (e.g., some may use it daily, while others use it only a couple times per year). Due to this unpredictability, the cloud not only has to be available at all time, but also be designed to scale up and down, depending on customers' requirements (Hurwitz et al. 2010). Accordingly, cloud computing allows customers to rapidly provision computing resources when required, and release them when no longer needed (Mather et al. 2009; Mell & Grance 2011:3). From the customer's perspective, resources are seemingly unlimited and any quantity can be taken into use at any time (Mell & Grance 2011). The ability to scale up and down is accomplished due to the cloud's "elasticity". This characteristic can be compared with the properties of a rubber band, which can be stretched or folded depending on the size of the objects it is holding (Hurwitz 2010).

On-demand self-service: Rapid elasticity of cloud computing enables the fulfilment of another essential characteristic – namely, the "on-demand self-service" (Krutz & Vines 2010). Users/customers can provision needed computing capabilities (e.g., storage) without interacting with the cloud provider (Mell & Grance 2011:3). On-demand self-service implies that the user/customer can manage, deploy and schedule the use of cloud services on their own, eliminating the need of human interaction. This can result in efficiencies and cost savings for both the user/customer and the cloud provider (Krutz & Vines 2010).

Measured services: Cloud computing includes metering capabilities, allowing the usage of resources to be monitored, optimized, controlled and reported in an automatic manner (Mell & Grance 2011). By measuring the usage, users/customers can be billed for the specific cloud resources that were utilized at a particular session (i.e., "pay per use") (Krutz & Vines 2010). Also, in similarity to public utilities delivered to one's house (e.g., water, electricity and natural gas), the customer can be charged for simply (the part of) the service that has been used – and not for an entire equipment (Krutz & Vines 2010; Velte et al. 2010).

2.1.2 Service Models

In previous sections, the term "resources" has been used to denote things being offered to users/customers by a cloud provider. What such resources may actually signify will be clarified below.

There are three major services offered through the cloud which are collectively referred to as "SPI" (i.e., Software-Platform-Infrastructure) (Mather et al. 2009; Krutz & Vines 2010). The most primitive out of these three services is **Infrastructure as a Service (IaaS)** (Linthicum 2009). This refers to the

provision of computer hardware – including servers, processing power, networking technology and storage (Hurwitz et al. 2010; Mell & Grance 2011) – on which arbitrary software can later be deployed and operated (Mell & Grance 2011). The customer typically constitutes an "IT-architect", and is obligated to self-maintain resources (i.e., platforms or applications) that are placed on top of the infrastructure (Chandrasekaran 2014).

Platform as a Service (PaaS) delivers more than just an infrastructure, namely, an integrated set of software that offers all essential resources for building applications (Hurwitz et al. 2010). It provides a development and application-hosting environment, comprised of e.g. programming languages, libraries and toolkits (Mell & Grance 2011). The PaaS customer may represent a "developer" who is responsible for managing his/her own application, while the provider maintains the underlying platform and infrastructure (Chandrasekaran 2014). The provider's platform includes channels for distribution and payment, meaning that customer can offer their applications to others (Mather et al. 2009).

Software as a Service (SaaS) provides applications that run on top of a cloud platform and infrastructure (Mell & Grance 2011). Typically, the service is a complete product that does not need to be supplemented with further hardware or software (Mather et al. 2009). SaaS customers are usually "end-users" that are released from all maintenance responsibilities, since the infrastructure, development platform and offered application is maintained by another entity (Chandrasekaran 2014). The application can typically be accessed on various client devices through e.g. a web browser (Mell & Grance 2011), and the consumers do not have to worry about licensing compliance, compatibility issues or patch installations (Mather et al. 2009; Krutz & Vines 2010).

Alternative service models have also been suggested in cloud computing-related literature. For instance, the term "Storage as a Service" (StaaS) has emerged due to the large number of providers that exclusively offers cloud storage on the market (Quick et al. 2013). Cloud storage falls under the umbrella of IaaS, but when it is offered independently from other infrastructure-related services it can be referred to as StaaS (Linthicum 2009; Bhowmik 2017). This service acts logically as a local set of storage space, even though it physically resides off-premise. It constitutes a resource that most of the other service models rely on (Linthicum 2009).

2.1.3 Deployment Models

A **public cloud** – also known as an "external cloud" (Velte et al. 2010) – entails that service offerings are made available for open use to the general public, and the underlying infrastructure is merely located outside of the customer's premises (Mell & Grance 2011; Bhowmik 2017). While a private cloud is meant for organization-specific use only (Bhowmik 2017), a public cloud is more appropriate for collaborative projects with external partners (Hurwitz et al. 2010).

A **private cloud** – also referred to as an "internal cloud" (Velte et al. 2010) – devotes resources exclusively to a single organization. The infrastructure may reside in-house or at an external location that is controlled or supervised by the customer (Mell & Grance 2011; Bhowmik 2017). Private clouds provide customers with higher control and overview of physical resources as well as incorporated security measures (Mather et al. 2009). A private cloud is preferable when control and security are highly important and when the customer must conform to strict data privacy requirements (Hurwitz et al. 2010). However, private clouds typically imply higher computing costs as well (Bhowmik 2017). Because the provider and the customer is usually part of the same organization, it is ultimately the organization that bears the cost of the cloud's underlying infrastructure (Halpert 2011).

A **community cloud** implies that the underlying infrastructure is shared among customers that are part of a community with e.g. the same mission, security requirements and policies. It might be owned, managed and operated by one or numerous community members, or by a third party (Mell & Grance 2011). Conceptually speaking, it resides somewhere in between private and public clouds. In contrast to a private cloud, it is employed by more than one organization (Krutz & Vines 2010). It combines the benefits of public clouds (such as multi-tenancy and pay-per-use billing) with the security and privacy level of private clouds (Bhowmik 2017). However, Krutz and Vines (2010) argue that the management of community clouds might become problematic due to undefined or shifting ownership and responsibilities. Moreover, this might make it difficult to deal with issues related to resilience, latency, privacy and security requirements.

A **hybrid cloud** is typically formed by combining the infrastructure of a private and/or community cloud with the corresponding infrastructure of a public cloud (Mell & Grance 2011; Bhowmik 2017). Thus, the customer can run non-critical applications on an external cloud infrastructure, while sensitive data and core applications are kept within the organization/community (Mather et al. 2009; Krutz & Vines 2010). Although each sub-cloud remains as a unique entity, they are tied together through standardized (or proprietary) technology that facilities portability of applications and data. An example of such technology is "cloud bursting", which acts as a load balancer between clouds (Mell & Grance 2011). Then, an application might primarily run in the customer's internal cloud, but can be relocated to an external cloud in conjunction with spikes in demand (Krutz & Vines 2010).

The distinction between public and private clouds should not be confused with the differentiation between the public and private *sector*. Typically, the public sector represents organizations that are owned, controlled and run by a government, whereas the private sector is comprised of businesses that are owned and managed by private individuals. Organizations in both sectors may offer services to the general public, but the objective of the former is to *serve citizens* while the latter is established with the motive of *making a profit* (i.e., the aim is commercial) (Surbhi 2015). When it comes to public and private clouds, on the other hand, only the former is available to the public while the latter is exclusively available to a single organization (Mell & Grance 2011).

2.1.4 Cloud Computing Stakeholders

2.1.4.1 Cloud Provider and Customer

A **cloud provider** constitutes an entity that offers a service to interested parties. The provider acquires, manages and operates the computing infrastructure as well as software that enable the cloud service (Liu et al. 2011:7). Normally, the cloud provider owns IT resources which can be leased to customers. However, some cloud providers might also resell resources from other providers (Erl et al. 2013).

Cloud customers represent an entity that utilizes the service offered by the cloud provider. The customer browses the provider's service catalogue, requests the desired service and arranges a contract with the provider, whereupon the service is provisioned (Liu et al. 2011:5). In Chandrasekaran (2014), customers of cloud services are referred to as "cloud service users". Despite its name, this actor does not always constitute *end-users*, since the term also encompasses *intermediate users* that deliver the cloud provider's services to those who will ultimately utilize it.

According to Aazam and Huh (2014), both of the aforementioned parties have various different roles. For instance, the cloud service may be handled by a service administrator and/or manager on both the provider's and the customer's side. Similarly, Erl et al. (2013) use the term "cloud resource administrator", which is described as an actor that may be hired by the cloud provider's organization

to perform management/administrative duties and to ensure that the overall cloud infrastructure will remain in operation. Such an actor may also belong to the customer, or a third-party contracted to administer cloud-based resources.

Furthermore, Erl et al. (2013) also use the term "cloud service owner" – a role that may not only be assumed by the cloud provider, but also the customer. The reason being that customers may be able to set up their own services in the provider's cloud.

2.1.4.2 Cloud Service Partners

Any individual or organization that facilitates the construction of the cloud provider's services is referred to as a "cloud service partner" by Chandrasekaran (2014). This represents a third party whose role is to assists the cloud provider or customers with tasks that might be out of the scope of their responsibilities. It may serve as a **cloud developer** that is employed to create (and integrate) components of a cloud service (Aazam & Huh 2014). Apart from developing and integrating systems, the cloud service partner may also act as a supplier of equipment (such as software and hardware) that enables the cloud service (Chandrasekaran 2014). Krutz and Vines (2010:288) mention the term "cloud enabler". This is not typically used to describe a cloud provider, but a *vendor* that offers technology that allows the provider (or other actors) to develop, distribute, operate and manage cloud solutions.

Services offered to cloud customers must conform to established regulations and policies in terms of e.g. security and performance (Bhowmik 2017:63). One can verify whether or not these agreed-upon conditions are met by employing a **cloud auditor**. This represents a third-party that can independently evaluate the cloud service and report their opinion accordingly (Liu et al. 2011:8; Bhowmik 2017). Such an unbiased assessment could help strengthen the trust relationship between customers and providers of cloud services (Erl et al. 2013).

There are a massive number of cloud providers on the market, and many might offer similar services. As a customer, one might be unaware of all available services, or be unable to recognize which service would bring the best performance. Moreover, customers might have the desire to use services from various different providers which would require additional system integration work (Bhowmik 2017:63). As cloud computing grows, the integration of cloud services can become too complex for customers to handle on their own. Consequently, the customers may request the cloud provider's services indirectly through a **cloud broker**. Such a party manages the usage, performance and delivery of cloud services, as well as negotiates relationships between the cloud provider and customer (Liu et al. 2011:8).

2.1.4.3 Cloud Carrier

While the cloud service is delivered via a cloud broker – or directly by the cloud provider – the agent in this delivery process is known as the **cloud carrier** (Bhowmik 2017:63). The role of a cloud carrier is commonly taken on by network and telecommunication providers (Erl et al. 2013; Aazam & Huh 2014). It acts as a mediator that offers connectivity and transport of cloud providers' services to the customer. These services can thereby be obtained through the customer's network-connected devices, such as computers and mobile phones (Liu et al. 2011:8-9).

2.2 Privacy and Security Concerns

Despite the previously mentioned benefits (see Section 2.1.1), there may be a reluctance to adopt cloud-based solutions due to a perceived risk of security and privacy issues (Lorünser et al. 2016; Kamara & Lauter 2010; Ren et al. 2012). In this chapter, such concerns will be described.

According to the United Nations (1948), privacy constitutes a fundamental right of every human being and should not be interfered with. It can be defined as "the right to be let alone" (Warren & Brandeis 1890:193), or "the claim of an individual to determine what information about himself or herself should be known to others" (Westin 2003:431). While security relates to mechanisms for handling all types of information, privacy merely relates to *personal data* (Pearson 2013). Art. 4 GDPR defines personal data as any information that can be used to (directly or indirectly) identify a natural person. It could be a name, an identification number, location data and an online identifier. Furthermore, it could also be information related to the natural person's physical, physiological, genetic, mental, economic, cultural or social identity.

Security of data is typically divided into three fundamental elements – i.e., Confidentiality, Integrity, and Availability (CIA) (Sloan & Warner 2013; Bhowmik 2017). Krutz and Vines (2010) suggest that this CIA triad represents a counter pole to Disclosure, Alteration, and Destruction (DAD). Similarly, Pearson (2013) argues that the security of data is ensured by implementing measures *against* impermissible access, disclosure, use, modification and destruction (Pearson 2013).

The confidentiality and integrity elements are both concerned with making restrictions against *unauthorized* individuals. The former involves preventing data disclosure/leakage to parties that are not allowed to read the information, while the latter involves protecting data from being tampered with or corrupted by aforementioned parties (Krutz & Vines 2010; Sloan & Warner 2013; Bhowmik 2017). Availability, on the other hand, relates to *authorized* individuals and should ensure access to the information in a timely and reliable manner (Krutz & Vines 2010; Bhowmik 2017).

2.2.1 Data Threats in the Cloud

2.2.1.1 Confidentiality and Integrity Issues

An incident where sensitive or confidential data is illegally released, viewed, used or stolen is referred to as a "data breach" by the Cloud Security Alliance (2017). This organization points out that data breaches are not unique to the context of cloud computing, but their surveys continuously show that such an incident is ranked as a top concern among cloud customers/users. It is suggested that data breaches may be caused by e.g. human error or by a *targeted attack*. Sloan and Warner (2013) argue that malicious external attacks on the information security may compromise more than one of the CIA elements. For instance, breaches of data confidentiality often involve violations of the data integrity as well, since the intention of the "attacker" may be to read secret data *and* to alter files (such as system logs) that might reveal the intrusion to the party owning/processing the information.

However, data stored in the cloud is not only the target of *external* threats. Cloud providers could themselves impose a potential threat towards the confidentiality of users' data, as they may be curious about its content (Fabian et al. 2015), or disclose the information to third parties without the user's consent (Pearson 2013). A current/former employee or business partner with access to the cloud provider's network or system could constitute a malicious insider threat by abusing or exceeding its access rights in a way that negatively affects the confidentiality or integrity of data (Cloud Security Alliance 2017).

Furthermore, cloud providers may have datacentres in various countries/regions, all of which has their own laws on how data should be handled (Halpert 2011; Pearson 2013). Offered cloud services are subject to national laws at the site of data origin (i.e., the client) as well as the territories of the cloud provider (Oshri et al. 2015). The provider may have to abide by law enforcement regulations in each location (Mather et al. 2009; Oshri et al. 2015), and if data is transferred between nations it is difficult

for users to prevent exposure of their data to law enforcement agencies (Mather et al. 2009; Pearson 2013).

2.2.1.2 Availability Issues and Data Loss

Halpert (2011) suggests that law enforcement may also have a disruptive impact on the *availability* of data; if law enforcement officials suspect illegal activities by any cloud customer, storage nodes within the cloud provider's data centre may be confiscated making data of multiple tenants unavailable. Mather et al. (2009) argue that the availability of data is generally affected by incidents that result in service outages/downtime. Krutz and Vines (2010) describe that even though the notion of availability includes aspects that are not purely associated with security (e.g., performance, uptime and guarantee of service), it could still be badly affected by security breaches in the form of Denial-of-Service (DoS) attacks. Bauer and Adams (2012) explain that such an attack overloads the system with fake service requests so that it cannot be accessed by legitimate users. This falls into "interruption of service" – one of the two main categories of issues that compromise the availability in the cloud. The second category is "destruction of resources" which refers to damage or loss of configuration information or other assets that prevent a service from being delivered correctly to the users.

When data is moved to the cloud, users essentially lose control over the physical security (Rittinghouse & Ransome 2010). Physical locations, on which the cloud provider's data centres reside, are subject to disasters (such as fires, floods or earthquakes) which could cause availability issues due to black-outs/outages of the datacentre's infrastructure (Bauer & Adams 2012). Furthermore, the damages from natural disasters may even result in permanent data loss (Cloud Security Alliance 2017). Examples of real-life disasters compromising service/data availability are described below:

- **Fire:** In April 20th 2014, a fire erupted at a Samsung datacentre in Gwacheon (South Korea). Consequently, the company's servers went down, causing its official website and features offered in e.g. Samsung's mobile app store to be inaccessible. Moreover, customers all over the world were unable to operate their Samsung Smart TVs, since the devices were dependent on the company's servers to function. The network outage and service disruption lasted from 06:00 to 13:30 (GMT). Although Samsung posted an official notice, apologizing for the incident, the company failed to explain why a fire at a single location could have such a significant impact on its services and devices.[1] [2]
- **Flood:** At the end of October 2012, several datacentres in New York struggled with connectivity and service issues, as an aftermath of Hurricane Sandy. InterNap and Peer 1 suffered from a flooded basement in their datacentre at 75 Broad Street which disabled crucial diesel fuel pumps, leaving them with no option to refuel generators. At 33 Whitehall Street, the flood took out servers in the datacentre of internet service provider Datagram, shutting down high-traffic sites such as Buzzfeed, Huffington Post and Gizmodo. Moreover, downtime due to generator failure was experienced by numerous tenants at 111 8th Avenue, a major communication hub owned by Google.[3] [4]
- **Earthquake:** Christchurch, New Zealand, was hit by two massive earthquakes in February 22nd and June 13th, 2011. Nearly 6000 businesses were partially or entirely destroyed. Many

[1] https://www.engadget.com/2014/04/20/samsung-com-outage-sds-fire/
[2] http://uk.pcmag.com/consumer-electronics-reviews-ratings/9618/news/fire-at-samsung-backup-data-center-takes-services-offline
[3] http://www.datacenterknowledge.com/archives/2012/10/30/major-flooding-nyc-data-centers
[4] http://www.datacenterdynamics.com/content-tracks/power-cooling/hurricane-sandy-takes-out-manhattan-data-centers/70690.fullarticle

businesses relied on electronic data which were totally or temporarily lost due to hardware damage/failure. [5]

Apart from failure of storage equipment, data loss may also be the outcome of accidental deletion (Cloud Security Alliance 2017) or cloud providers running out of business (Mather et al. 2009; Armbrust et al. 2010). Moreover, providers may also *intentionally* impact the availability of information in terms of data retention/lifecycle. They might discard seldom used information (without notifying the user) to free up storage space for cost-saving purposes (Wang et al. 2013) – or *keep* data after the user has made a request for its removal (Pearson 2013).

2.2.2 Trust

Often, people find it more difficult to trust online services in comparison to offline services (Pearson 2013). Similarly, cloud-based solutions are generally perceived as less trustworthy than in-house systems (Khan & Malluhi 2010). Sunyaev and Schneider (2013) claim that trust in the cloud's security could be a prerequisite in order for the offered service to be adopted by customers/users. However, as described by Pearson (2013), trust is not only a matter of security. Although the notion involves "hard" issues such as security measures (e.g., authentication and encryption), it also concerns "soft", subjective issues such as human psychology, experience, user-friendliness, and reputation. Khan and Malluhi (2010) describe that trust could typically be described as an act of confidence in that another entity will behave/deliver as promised. Uusitalo et al. (2010) suggest that users' trust in clouds is about giving away control and believing that actions of the trustee (i.e., the provider) will have a positive outcome in regards to something that is valuable to the trustor (i.e., user).

In PRISMACLOUD, novel security- and privacy-enabled cloud services are developed (such as Archistar) (Alaqra et al. 2017). In order for users to accept and utilize new technology, it is crucial to establish trust to overcome perceived risks and uncertainties (Li et al. 2008). New technology may earn the trust of potential customers/users by building a good reputation in terms of security and performance – a progress that is *gradual* (Khan & Malluhi 2010). Trust is highly influenced by the user's knowledge and experiences, which are continuously evolving (Firdhous et al. 2012). Although trust may be something that is established over time, once expectations from the service have been met (Gharehchopogh & Hashemi 2012): cues, clues or evidence of an entity's trustworthiness help users determine whether or not it can be trusted (Nissenbaum 1999). Poor first-hand experiences with another entity can in particular form a mistrust towards it (Khan & Malluhi 2010), but when users do not have a history of direct interaction with another entity, they might instead make a judgment based on its reputation or evaluate its trustworthiness indirectly through experiences of others (Nissenbaum 1999).

The security of cloud services can be certified by an independent auditor. The certificate would serve as a stamp of quality that (with a given degree of confidence) assures customers/users that the cloud service is secure to utilize (Khan & Malluhi 2010). Sunyaev and Schneider (2013) describe that public key certificates represent a common means for verifying the authenticity of websites and facilitating customer/user trust in the context of services for online banking or shopping. It is suggested that certification by an independent auditor can have the same positive effect on trust in cloud-based services. However, displaying a large number of trust seals (i.e., certificates) on a website may also give the impression that the provider is trying too hard to prove its trustworthiness. This could, in turn, cause scepticism among customers and lower the likelihood of completing a prospective purchase (Özpolat & Jank 2015).

[5] https://www.businessblogshub.com/2012/10/natural-disasters-and-data-loss/

When it comes to direct service interaction, users tend to trust systems that provide them with *control* over data assets (Gharehchopogh & Hashemi 2012), and less control typically implies that the system will be perceived as less trustworthy (Khan & Malluhi 2010). Data control may not only be a feature that the users desire, or a necessity in order to establish sufficient trust for cloud adoption. It may also be a legal requirement (Pearson 2013).

Apart from diminished control over storage equipment and the data's life cycle, low level of user control is also signified by a dearth of *transparency* (Pearson 2013). The notion of transparency can be defined as "the availability of information about an organisation or actor allowing external actors to monitor the internal workings or performance of that organisation" (Grimmelikhuijsen 2012:55). It can be used as a synonym for openness about decision-making in organizations or governments. The easier it is for the general public to obtain the information, the greater transparency (Ball 2009). By enhancing the transparency, users' disbelief towards a cloud service can be reduced (Khan & Malluhi 2010; Uusitalo et al. 2010). If an accident occurs in the cloud, transparency about it can prove to the users that it was not caused by the provider due to incompetence or negligence, and that the provider takes appropriate actions against the incident (Uusitalo et al. 2010).

Apart from details about how information is being handled by the cloud provider, another transparency issue related to cloud services is lack of knowledge about the *physical location* of data processing and storage (Khan & Malluhi 2010; Pearson 2013). Sitaram and Manjunath (2011:321) argue that the geographical location of data centres should ideally be known by the cloud users in advance to avoid legal issues (e.g., law enforced exposure). As pointed out by Pearson (2013), laws may place geographical restrictions on third-party processing of personal/sensitive data and thereby also limit the use of cloud services. Similarly, Halpert (2011) suggests that cloud customers should consult with providers about the countries in which they operate. If possible, they should make restrictions to countries with similar privacy and security legislations as the customers' local laws. Furthermore, Bhowmik (2017) describes that the geographical distance between the data centre and the user may implicate that data will travel a long distance via the network when it is requested by a user. Transferring large-sized data (e.g., high-definition video files) across the network may cause performance issue.

The trust in a cloud solution may vary greatly depending on its deployment model (Gharehchopogh & Hshemi 2012). The level of security/privacy as well as the cost of utilized resources may vary between the different cloud deployment models (discussed in Section 2.1.3). A public cloud is typically the least expensive (Goyal 2014) and the predominantly used model in scenarios where restrictions on cost are crucial to the customer (Pearson 2013). However, public clouds are typically less secure than private clouds (Goyal 2014), and may not be deemed as suitable for processing sensitive information since the perceived risk of data leakage or loss is too high (Pearson 2013).

2.2.3 Data Classification

Information classification can be utilized to identify which data is most crucial or sensitive to an organization (Krutz & Vines 2010). It constitutes the basis for establishing an understanding of what implications it may have to lose the security/privacy of a particular data set, as well as making decisions in regards to protection of information in the cloud (Halpert 2011). It helps to ensure that each type of data will be appropriately safeguarded (Krutz & Vines 2010; Mather et al. 2009) and by focusing security measures on the data that needs it the most, a more cost-efficient employment of data protection will be accomplished. Furthermore, classifications of data may also be performed due to legal/regulatory requirements (Krutz & Vines 2010).

In *"Security self-assessment guide for information technology systems"* by the National Institute of Standards and Technology (Swanson 2001), a High/Medium/Low data classification scheme is proposed for rating the sensitivity level of data. According to the scheme, each level implies the following:

- If data with "low" sensitivity is compromised, it could lead to minor financial loss or require administrative action (within the organization) for correction.
- If data with "medium" sensitivity is compromised, the financial loss may be more significant and *legal* actions may be required.
- If data with "high" sensitivity is compromised, the financial loss may be major and also require legal action for correction. Furthermore, the incident could cause loss of life or imprisonment.

Krutz & Vines (2010) suggest that such a classification scheme could be used to also rate the data in terms of the CIA parameters. This suggestion complies with the scheme presented in *"Standards for Security Categorization of Federal Information and Information Systems"* by FIPS Publication 199 (2004), utilized to assess the potential impact on organizational operations, assets or individuals if the confidentiality, integrity or availability of data is lost.

2.3 Multi-cloud Solution based on Secret Sharing

A multi-cloud solution based on Secret Sharing (such as Archistar) enables secure *distributed storage* of data in the cloud (Lorünser et al. 2016). The notion of distributed storage implies that information is kept as fragments (rather than entire data sets) across multiple machines (Pearson 2013). In the solution proposed in this study, information is divided into fragments/chunks which are dispersed to data centres in separate, independent clouds. Thereby, the damage in – or caused *by* – a single cloud service can be limited.

The concept of Secret Sharing and multi-cloud will be explained in detail in Section 2.3.1 and 2.3.2. Subsequently, Secret Sharing is compared to other security measures in Section 2.3.3, and privacy legislations that apply to the proposed solution are mentioned in Section 2.3.4. Lastly, some previous research on Secret Sharing solutions is described in Section 2.3.5.

2.3.1 Origin of the Secret Sharing Concept

The concept of Secret Sharing was originally invented by Blakley (1979) and Shamir (1979) with the intention to facilitate the management of cryptographic keys. It was suggested that data could ideally be protected with encryption. But in order to subsequently protect the *encryption key*, another security measure was needed since further encryption would *move* the problem, rather than solve it (Shamir 1979).

Shamir (1979) claimed that the most secure way to ensure that a key would not get into the wrong hands was to store it in a single, well-guarded location. However, this would imply great reliability issues since the key (and consequently the information protected by it) could become inaccessible due to a single misfortune at this particular storage location. Blakley (1979) argued that cryptographic systems typically involved *numerous copies* of a crucial key which are stored in several protected sites. Although this might be seen as an obvious solution to increase the reliability, Shamir (1979) pointed out that the creation and distribution of copies would also result in a higher risk of security breaches. Similarly, Blakley (1979) described that if a key is duplicated too many times, one of the copies could potentially get lost and end up in the reach of adversaries. On the other hand, if an

insufficient amount of copies is produced, one might not be able to guarantee that the entire set of keys will not be destroyed.

Instead of creating entire copies of an encryption key, Blakley (1979) and Shamir (1979) independently proposed the concept known as Secret Sharing, where a "secret" (i.e., the key or any form of data) is divided into numerous "chunks" (or fragments).[6]

When implementing a Secret Sharing solution, Blakley's (1979) idea was that the key/data owner should in advance decide on a number of misfortunes that the key/data should be safeguarded against – i.e., a abnegation incidents (data loss) and b betrayal incidents (data breaches/collusion). The former refers to events where information entrusted with a "guard" can no longer be completely reclaimed by the owner due to accidental loss or destruction. Betrayal incidents, on the other hand, constitute events where the guard discloses the information to an unauthorized individual (see Table 1).

Shamir's (1979) way of describing the Secret Sharing concept did not make a distinction between different types of data threats or incidents. Instead, a so-called (k, N)-*threshold scheme* was proposed, where the user should decide on how many chunks the key/data should be divided into (i.e., N), and the subset of chunks (i.e., k) that should be required to reconstruct the information into its original state (see Table 1).

Table 1. Parameters in Blakley's (1979) and Shamir's (1979) version of Secret Sharing respectively.

	Blakley (1979)	**Shamir (1979)**
Total number of data chunks	$a + b + 1$	N
Threshold of data reconstruction	$b + 1$	k
Protection against Data Loss (Abnegation incidents)	If a chunks are lost or destroyed, data can still be reconstructed by the data owner with $b + 1$ chunks.	So long as no more than $N - k$ chunks are destroyed or lost, the data owner can reconstruct the data.
Protection against Collusion/Data Breaches (Betrayals incidents)	If b chunks are disclosed or stolen, adversaries still do not have enough chunks to reconstruct the data.	So long as no more than $k - 1$ chunks are disclosed or stolen, adversaries cannot reconstruct the data.

In the present study, Secret Sharing algorithms based on Shamir's scheme are presupposed.

Data is divided into N chunks (i.e., fragments) that independently do not reveal any information about the original content. In order to reconstruct the data into its original state, any arbitrary set of k chunks can be used (due to some level of redundancy). Thus, if a certain chunk is inaccessible, lost or corrupted, the data owner is still able to regain the information by gathering other fragments (Lorünser et al. 2016).

In order to create a secure and reliable storage service, the chunks should be distributed to (data centres in) separate clouds. No single cloud storage provider (CSP) should have access to enough fragments to obtain plain-text and tampering with one chunk should not compromise the integrity of

[6] Chunks may also be referred to as "shares". However, in order to avoid that Secret Sharing will be confused with the notion of sharing regular information with other parties (e.g., via social media or other forms of communication), the term chunk will be used instead.

the original data. Moreover, if there is a sufficient distance between storage nodes, only one chunk will become (permanently or temporarily) unavailable to the data owner in an event of a disaster (Lorünser et al. 2016).

As indicated by Table 2, a particular level of data fragmentation is required in order to achieve data protection and data loss prevention benefits from Secret Sharing. That is, the minimum number of chunks (i.e., N) is 3 and the threshold for reconstruction (i.e., k) is 2.

Table 2. Benefits from different values on N and k.

	Protection against Data Breaches	**Protection against Data Loss**	**Description**
$N = 1, k = 1$	–	–	No data splitting. One chunk will contain all data.
$N > 1, k = 1$	–	✓	Only one chunk is needed to reconstruct the data.
$N > 1, k = N$	✓	–	All chunks are needed to reconstruct the data.
$N \geq 3,$ $N > k \geq 2$	✓	✓	More than one chunk is needed to reconstruct the data. The data will still be recoverable if some chunks are destroyed.

In comparison to traditional storage (where the full data is kept on a single device), a Secret Sharing solution may require a larger amount of storage space (due to a certain level of redundancy). The amount of storage overhead will depend on which Secret Sharing *algorithm* is employed in the solution. For instance:

- Shamir's algorithm is referred to as a *Perfect Secret Sharing* (PSS) scheme as the privacy guarantees are said to be "information theoretic" and free from errors (Bellare & Rogaway 2016), meaning that no information will be disclosed to an adversary regardless of how much computing power he/she has (Martin 2008). However, a limitation with PSS is that each chunk must have the same size as the original data, which makes it unwieldy when a large set of files is to be stored (Bellare & Rogaway 2016).
- A *Computational Secret Sharing* (CSS) scheme, on the other hand, permits data chunks to be smaller than the original information. If the size of the original data is S and the threshold for reconstruction (k) is 2, the size of each chunk would be $S/2$. However, the solution's privacy properties may no longer be information theoretic and it may still be possible for an unauthorized individual to obtain a small amount of information (Bellare & Rogaway 2016). CSS is utilized with the assumption that adversaries only have a moderate amount of computing resources (Martin 2008).

In the context of this thesis, it will be assumed that a PSS scheme will be utilized at all time.

2.3.2 Multi-cloud

When a large organization relies merely on a single cloud provider, numerous issues could ensue. For instance, the cloud service might become unavailable for a certain period of time which not only diminishes the benefit from the cloud provider's offering, but also has a negative impact on the organization that intends to utilize it. Another significant danger is the risk of permanent data loss due to e.g. a system failure (Marinescu 2017:84). To prevent data loss or availability issues, one could

argue that services should by principle not operate on a "single point of failure" and the CSP may therefore have several data centres in different regions. However, this would still imply that users' data is at risk of being permanently lost if the provider goes out of business. Thus, instead of relying on a single company, Armbrust et al. (2010) argue that high availability could only be guaranteed if *multiple* CSPs are employed.

Vukolić (2010) coined the term "inter-cloud", referring to a cloud of several independent cloud services. The idea was that security and reliability should be distributed across multiple clouds to improve the offering of each individual CSP. Furthermore, by not longer depending on a single cloud, concerns about security threats, availability issues and loss of data control could be mitigated. According to Petcu (2013), there are two types of inter-clouds, i.e. *federated cloud*[7] and *multi-cloud*. A federated cloud implies that cloud providers have formed an agreement to share resources. The users/customers interact with one of the clouds, not knowing that the utilized resources or services may reside in another. In a multi-cloud, on the other hand, there is no agreement between providers. The users/customers are not only aware of the different clouds, but also responsible for handling the provision of resources or services. The most common form of the multi-cloud concept is in turn a so-called hybrid cloud (described in Section 2.1.3), meaning that both private and public cloud storage providers are employed (Petcu 2013).

2.3.3 Comparison of Secret Sharing with Other Security Measures

Information security operations typically involve trade-offs between confidentiality and availability (Ioannidis et al. 2012). Although having omnipresent access to cloud data may be an attractive advantage, Menkel (2008) argues that one should draw a line on how accessible the information should be for the sake of protecting its confidentiality. Sloan and Warner (2013) argue that it is easy to maintain the confidentiality and integrity of information if one does not have to worry about its availability. That is, data can be kept safe from adversaries by eliminating the power from the storage device(s), but this would also make the information inaccessible for authorized individuals.

Arockiam and Monikandan (2014) suggest that mechanisms that protect data confidentiality can also ensure the *integrity* of data. Authentication techniques can be utilized to safeguard the integrity of data from external attacks, because if adversaries cannot *access* the cloud storage then they can also not maliciously alter or modify the stored information. However, the Cloud Security Alliance (2017) describes that weak identity and access control is one of today's major concerns in cloud computing, which enables external attackers to get a hold of the user's data.

According to Arockiam and Monikandan (2014), encryption constitutes the most common technique for ensuring data confidentiality. It represents the process of converting plain-text into an unreadable state (known as "cipher-text") using a cryptographic algorithm and a secret key.[8] However, in the context of cloud-based environments, encryption would alone not suffice in protecting the confidentiality of data. As evidenced by e.g. Hodgson (2015) and Li et al. (2013), there are techniques for breaking encryption without the secret key. Mather et al. (2009) point out that high confidentiality does not *always* imply high data integrity. It is argued that file encryption may ensure that information is not disclosed to unauthorized individuals (even if they manage to gain access to the cloud storage). However, the encrypted file may still be corrupted or tampered with and, thereby, have its integrity compromised. Furthermore, Ren et al. (2012) suggest that encryption may suffer from performance issues and be less appropriate when access is needed by a large number of individuals. As described

[7] Also called "federation of cloud" or "cloud federation".
[8] Managed by the user or a trusted guardian.

by Shamir (1979) and Blakley (1979), the more copies of an encryption key that is created and distributed, the higher the risk of adversaries obtaining the key (and the information protected by it).

As indicated in Section 2.3.1, a Secret Sharing solution would require greater efforts in order for adversaries to alter the integrity of the full data. Gaining access to and tampering with a fragment/chunk in a single cloud would not corrupt the original information so long as k chunks remain untouched and available in other clouds. Depending on the values on the Secret Sharing parameters (i.e., N and k), there will be some form of trade-off between the data *confidentiality* and *availability* in the sense that one of the two factors will be more enhanced than the other. However, when Happe et al. (2017) discusses their multi-cloud solution based on Secret Sharing (known as "Archistar"), it is pointed out that both aspects will still be improved in comparison to a single cloud system. By utilizing a solution like Archistar, the user is relying on a "non-collusion assumption" – i.e., it is presumed that employed CSPs are unaware of each other and do not collaborate behind the user's back to combine fragments/chunks and reconstruct the data. Happe et al. (2017) suggest that "untrusted" CSPs can be employed in the Archistar solution while still maintaining sufficient security of data. In Zambrano et al. (2017), it is suggested that untrusted providers are those who offer *public* (i.e., commercial) cloud storage. In the context of this thesis, it will be left to the user/customer to decide whether or not certain CSPs are trustworthy.

2.3.4 Legal Implications

Although a single chunk does not reveal any information to the individual who gets a hold of it, it is still possible to reconstruct the original data if the chunk were to be combined with $k-1$ other fragments. Therefore, dividing data into chunks and storing them in separate locations corresponds to the Art. 4(5) GDPR's definition of "pseudonymisation":

> "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."
>
> Art. 4(5) GDPR

"Additional information" that is "kept separately" would in this case represent other data chunks. As concluded in Mourby's et al. (2018) review of the General Data Protection Regulation (GDPR), personal data will remain personal data even if pseudonymisation is applied and should be handled accordingly.

Multiple regulations apply when personal data is outsourced to the cloud. For instance, Art. 28 GDPR stipulates that personal data should only be outsourced to CSPs with whom the user has a "data processing agreement". This represents a contract that stipulates that personal data should only be processed by the provider if documented instructions have been given by the customer, and that technical as well as organization security measures should be taken. Moreover, the adequacy principle of Art. 45 GDPR specifies that personal data may only be transferred to a country outside of European Union if the EU commission has determined that the destination (i.e., country, territory or sector within the country, or international organization) ensure an adequate level of protection. For instance, the *EU-US Privacy Shield* (adopted by the EU commission in 2016) allows for personal data to be transferred to the US.

2.3.5 Previous Studies of Solutions based on Secret Sharing

Prior research suggests that Secret Sharing can be employed in several different areas/fields. For instance:

- **m-Banking or m-Payment.** Wong & Kim (2016) notices an increased adoption of banking and payment applications on mobile phones. It is argued that such devices could easily be lost or stolen and, therefore, require stronger protection against unauthorized access to private information/applications. In order to improve user authentication for mobile banking and payments, Wong & Kim (2016) proposes a solution involving the use of wearable devices. With the help of Secret Sharing (or "Secret Splitting"), private data such as credit card and banking information can be divided into chunks which are subsequently stored on separate devices (i.e., a mobile phone and a wearable device). It is suggested that the solution will prevent private information from leaking in case one of the devices is lost or stolen.
- **e-Health.** Medical applications for telediagnosis or teleconsultation require physicians and patients to exchange information over insecure networks. In order to prevent unintentional disclosure of the patient's medical information to unauthorized individuals, Ulutas et al. (2011) propose a solution that combines Secret Sharing with Steganography (i.e., the practice of concealing a file/message within another). Medical images (e.g., digitized x-rays) are first split into N chunks, all of which will have a noise-like appearance. It is suggested that these may still attract the curiosity of eavesdroppers. Thus, Steganography is subsequently used to hide chunks in different "natural" cover images, along with Electronic Patient Record (EPR) information. The cover images are then distributed (by the patient) to separate physicians and subset k out of n is needed to restore the original image. If any cover image is modified during the retrieval phase, the original image will contain corrupted sections/regions once it has been reconstructed. It is argued that the proposed solution will provide three capabilities at the same time, i.e.: EPR hiding, confidentiality and authenticity.

  In Fabian et al. (2015), it is described that sharing of medical big data among healthcare organizations becomes increasingly important. Although cloud computing may provide means for the needs of collaborating healthcare professionals, privacy and security risks prevent a wide cloud adoption within the health domain. Therefore, Fabian et al. (2015) develop a "novel architecture" where encrypted Electronic Health Records (EHR) are split into chunks that are dispersed to a multi-cloud environment, consisting of several independent CSPs. Each provider is assumed to be "semi-trusted" (i.e., honest in protecting the storage against external threats, but curious about the data being stored), and will not be able to obtain any information from a single chunk.
- **e-Voting.** Neumann et al. (2014) develop a smartphone application for electronic voting, intended for real-life elections. In existing e-Voting systems, votes are typically encrypted upon submission. At the end of the election, the votes are anonymized and eventually decrypted when votes are to be counted. If a single election authority has access to the decryption key, the secrecy can be violated by disclosing a voter's identity before the anonymization. Thus, Neumann's et al. (2014) application features protocols for generating keys and decrypting votes in a distributed manner (i.e., chunks of the decryption key is stored among multiple authorities).

As indicated above, Secret Sharing could be used for secure *storage* of personal information (Wong & Kim 2016) or keys (Neumann et al. 2014), but also for the purpose of *sharing* data in a secure manner (Fabian et al. 2015; Ulutas et al. 2011).

The previously proposed solutions have been evaluated to a varying extent and with focus on different aspects. Wong and Kim (2016) evaluated their solution through a "security analysis" based on

attacks/breaches mentioned in literature. Ulutas et al. (2011) and Fabian et al. (2015) indicated the feasibility of their proposed solutions by conducting performance experiments. Out of the aforementioned studies, only Neumann et al. (2014) had human participants (as potential users) and made an assessment of their application's usability. During their evaluation, the participants succeeded with performing tasks in the application but they were unable to answer questions about the security provided by it, which indicates that there was still a lack of understandability.

As mentioned before, to the best of the thesis author's knowledge, the emphasis in Secret Sharing-related research seldom lies on human factors and the perspective of the user. Thus, there are little clues as to how prospective users would perceive a solution like Archistar, where they would be in charge of the configuration of the Secret Sharing mechanism and geographical distribution of data chunks.

2.4 Summary of Problem Background

The following provides a summary of the problem background related to research question **RQ1** (i.e., "What are suitable configuration options and guidelines for organizational or private users with different security requirements?"), and **RQ2** (i.e., "What are relevant trust factors, unique advantages, and risks of a multi-cloud storage solution based on Secret Sharing that should/could be communicated to the users?").

RQ1: When using a Secret Sharing solution (such as Archistar), decisions have to be made in regards to two fundamental issues - i.e.: (1) *The Secret Sharing parameters*. When selecting values for N and k, one is faced with a trade-off situation between data confidentiality and availability. That is, a high threshold for data reconstruction may increase the former aspect but it will also make the information less easy to obtain - not only for unauthorized individuals but also for the data owner (see Section 2.3.1). (2) *The geographical distribution of data chunks.* CSPs often have data centres in multiple areas. Thus, when a data chunk is transferred to a particular cloud, it can be stored in various physical locations which, in turn, may be subject to different forms of laws/jurisdictions as well as natural disasters (see Section 2.2.1). Depending on the sensitivity of data, the user/organization may be obligated to follow certain legal restrictions. For instance, if a certain backup/archiving project contains *personal data*, it may only be distributed to countries within the European Union - or locations that have proven to provide an equivalent level of protection (see Section 2.3.4). Configuration settings that are suitable for the user/organization depend on the information security requirements related to their backup/archiving project(s). In order to establish which requirements apply to a particular situation, some form of data classification is commonly performed (see Section 2.2.3). However, in the context of a multi-cloud secret sharing solution, it is yet to be established how the user's security requirements should be communicated and subsequently addressed/fulfilled, while finding an appropriate balance between the confidentiality and availability of data.

RQ2: There are numerous threats/concerns that may hinder the adoption and usage of (single) cloud services (see Section 2.2.1). The level of privacy/security (as well as cost) may vary depending on the cloud's deployment model (see Section 2.1.3). In the context of Archistar, a multi-cloud infrastructure is customized and utilized by the user, which can be comprised of a combination of both private/internal and public/external clouds. According to the non-collusion assumption, less "trustworthy" CSPs can be employed without compromising the confidentiality of data (see Section 2.3.3). However, a solution that provides a high level of security and privacy does not necessarily result in a high level of user trust (see Section 2.2.2). Trust issues may not only relate to the cloud as a storage medium, but also the Secret Sharing mechanism. Alternative security measures such as data encryption may be more common and familiar to the user (see Section 2.3.3). The users may have a

hard time accepting technology that they have little/no previous experience with. The confidence in individual cloud services/providers as well as new technology may be facilitated through various means or aspects - e.g., privacy seals/certificates, trust ratings, transparency (see Section 2.2.2). Whether the proposed solution will be perceived as sufficient protection, or if privacy/security concerns (typically associated with cloud services) will remain in a multi-cloud and secret sharing solution is to be determined.

3. Methodology: Interviews and User Walkthroughs of UI Prototype in a Design Study

In order to establish which "configuration options and guidelines" are suitable for users/organizations (RQ1), information about their requirements and understanding of a multi-cloud solution based on Secret Sharing would be investigated. Furthermore, their perception and acceptance of the proposed solution would be examined for the purpose of identifying "trust factors, unique advantages, and risks […] that should/could be communicated to the users" (RQ2).

In the study, Archistar will be utilized as an example of the intended solution. As earlier mentioned, Archistar is based on an e-Government use case defined in the PRISMACLOUD project. Conditions around a specific case (e.g., a particular community or organization) might typically be explored through a *case study* where a real-world setting may be observed (Bryman 2012). However, a solution like Archistar does not constitute traditional storage of data backups or archives, so conditions and requirements/needs that apply to the users' current systems may not be relevant in the case of Archistar. Since the researcher did not have access to an environment where a solution like Archistar already exists, an exploratory design study would be conducted instead. That is, interviews would be conducted with prospective users, followed-up by the creation and evaluation of a user interface (UI) proposal.

Methodological and ethical considerations are described in Section 3.1 to 3.4. Moreover, the aforementioned e-Government use case is scrutinized in Appendix A to identify prospective conditions around the intended solution. Aspects from the use case that were considered in this study are subsequently mentioned in Section 3.5. Lastly, the setup of utilized research methods and data collection procedures are described in Section 3.6 to 3.7. Design decisions in the UI proposal are described in Appendix G.

3.1 Conducting Exploratory Research

A great proportion of social research is conducted for the purpose of exploring – i.e., familiarizing oneself with a research topic. Typically, this tactic is used when a researcher examines a new area of interest or when the study subject itself is fairly new (Babbie 2012). When barely any prior research has been conducted on a specific topic, the amount of previous literature from which leads can be drawn is limited. Consequently, an "inductive" research approach should ideally be employed, which implies that the researcher attempts to develop theory from empirical data (i.e., his own observations or findings) (Kalof et al. 2008; Bryman 2012). Such an approach is sometimes referred to as "exploratory research" (Kalof et al. 2008), and is typically associated with qualitative research methods (Bryman 2012).

Another, less appropriate strategy for investigating new research topics is a so-called "deductive" approach. This entails that the researcher aims to test existing theory (Kalof et al. 2008; Bryman 2012) – hence, it is sometimes referenced as "confirmatory research" (Kalof et al. 2008). In contrast to inductive studies, this approach is typically characterised by research methods of a quantitative nature (Bryman 2012). The main distinction between qualitative- and quantitative-oriented research methods is the type of data that is being collected. Quantitative methods typically entail the use of measurements to gather "hard data" in form of numbers, which are subsequently processed and analysed statistically, whereas qualitative methods collect "soft data" in form of e.g. words, symbols or photos which are analysed in a more interpretive manner (Kalof et al. 2008; Patel & Davidson 2011; Bryman 2012; Neuman 2013).

In qualitative research, one can use a variety of data collection techniques such as document analysis, interviews and observations (Kalof et al. 2008; Neuman 2013). A disadvantage with studies that analyse documents or other forms of human communication is that they are limited to information that has already been recorded (Babbie 2012). When conducting observations, events or experiences from people's past (i.e., something that may affect their attitude, perception or behaviour; Rubin & Chisnell 2008) are not available for scrutiny. However, such information can be collected by *asking questions* (Patel & Davidson 2011).

Two techniques for gathering information by asking questions are *questionnaires* and *interviews* (Patel & Davidson 2011). The latter implies that the researcher (or a representative interviewer) verbally asks questions and records the answers given by a respondent. The former constitutes an instrument from which the respondent himself/herself both read questions and enter his/her replies (Patel & Davidson 2011; Bryman 2012).

3.1.1 Questionnaires or Interviews

Questionnaires can be distributed by post, email or as an online survey. Furthermore, they may also be filled in under the researcher's presence and supervision (Babbie 2012). Self-completion questionnaires may be practically convenient for all parties. They are typically quicker for researchers to administer as they can be distributed to a massive number of people simultaneously, and a wide geographical area can be covered at a small cost. Moreover, the respondents can in turn decide themselves which time and pace the questionnaire should be completed (Bryman 2012). However, letting respondents complete the questionnaire unsupervised is not without its issues. If a respondent finds it difficult to comprehend and therefore answer a question, there is no one present to clarify any ambiguities. Moreover, if a given answer is unclear, the researcher cannot ask the respondent follow-up questions for further explanation. There is also a bigger risk of missing data as the respondent may (intentionally or unintentionally) skip questions, meaning that the questionnaire is returned insufficiently completed (Bryman 2012; Neuman 2013). Furthermore, it cannot be confirmed whether or not the questionnaire is actually completed by the intended respondent (Neuman 2013). For instance, if the questionnaire is sent to a particular individual via post, other people in that household may help out in answering the questions (Bryman 2012).

Interviews are generally conducted in person (i.e., face-to-face) or via telephone (Patel & Davidson 2011; Babbie 2012). Compared to questionnaires, interviews typically achieve a higher response rate (Bryman 2012; Neuman 2013). It is more suitable for complicated topics (Babbie 2012), and for asking a large number of questions since a lengthy questionnaire may be perceived as off-putting and the respondent may decide at first glance to not participate (Bryman 2012). In similarity to self-completion questionnaires, interviews via telephone implicate that the researcher has less control over surrounding conditions that may distract the respondent from answering questions (Bryman 2012; Neuman 2013). Interviews via telephone are typically less expensive and quicker to administer since the interviewer does not have to spend money and time travelling to remote respondents. In regards to controversial opinions, respondents are sometimes more honest when they do not have to look the interviewer in the eyes. On the other hand, people may also be more suspicious when they cannot see the person who asks the questions. In-person interviews are more appropriate if the respondent has a hearing impairment, since the audio quality during phone calls may be inconsistent and unpredictable. During a face-to-face interaction, the researcher is also in a position where he can observe the respondents' facial expressions, meaning that he can respond to/follow up on (silent) signs of puzzlement or unease. Moreover, it is easier to use visual aid from which the respondent may e.g. be asked to select an answer (Babbie 2012).

3.1.2 Open-ended and Closed-ended Questions

According to Patel and Davidson (2011), interviews and questionnaires can both be structured or unstructured. A distinction is made between the level of "structure" and "standardization". The structure refers to the extent that respondents can freely answer the asked questions. Interviews/questionnaires with a high level of structure use closed-ended questions where the respondents' answers are limited to a list of alternatives defined by the researcher, while interviews/questionnaires with a low level of structure utilize open-ended questions, allowing the respondents to answer in their own words. Similarly, Neuman (2013) describe that *answers* received on open-ended questions are unstructured, while the corresponding responses on closed-ended questions are fixed and structured.

Standardization, on the other hand, is referred to as the extent to which the interviewer can freely choose the phrasing and sequence of questions from one interview to another. A high level of standardization entails that the questions have the same wording and order, whereas a low level of standardization implies that the way questions are presented varies between sessions (Patel & Davidson 2011). However, structure and standardization are not differentiated in all literature on social research. In Bryman (2012), "structured" and "standardized interview" are instead used as synonyms, both referring to a setup where questions are identically presented to all respondents so that they are provided with the same cues and context of questioning. It is also suggested that an interview can be "semi-structured", which means that the researcher has prepared a list of questions that are more general in nature. The researcher can divert from the list by altering its order or by asking additional questions. In interviews that are "unstructured", the researcher has simply prepared a list of *themes* that could/should be covered. Questions related to these topics are then asked in an informal manner and vary significantly between sessions.

Interviews in qualitative research are usually unstructured or semi-structured, leaving the respondents with much leeway in how to answer questions (Patel & Davidson 2011; Babbie 2012; Bryman 2012). Qualitative interviewing may not only imply that open-ended questions are asked verbally – it may also involve an open-ended questionnaire. Closed-ended questions are predominantly used in survey research since it provides greater uniformity of responses and makes it easier to process a large number of answers (Babbie 2012). Surveys involve the use of either a self-completion questionnaire or structured interviews to collect quantitative data. While quantitative-oriented interviews reflect the concerns of the researcher and focus on measuring key concepts, qualitative interviews put emphasis on investigating the viewpoint of the *respondent*. In other words, while rambling may be seen as a distraction in quantitative interviews, respondents may be encouraged to go off on tangents in interviews that are qualitative as it provides an insight into what they perceive as relevant and crucial (Bryman 2012).

Open-ended questions are useful for exploring new areas (Bryman 2012). They facilitate spontaneity and allow researchers to derive interesting answers that could not be predicted beforehand (Bryman 2012; Kalof et al. 2008). When using closed-ended questions, on the other hand, the researcher *should* foresee all possible answers to ensure that no relevant factor is omitted (Babbie 2012). The respondent's level of knowledge and understanding of a concept may be less easily determined with closed-ended questions, since the answer alternatives may be used by respondents to reply based on guesswork rather than certainty. However, the answer alternatives can also help to clarify the meaning of questions that are ambiguously phrased or difficult to communicate (Bryman 2012).

3.2 Visualizing System Design Ideas

The part of a system that the user can see, comprehend and direct represents the User Interface (UI) (Galitz 2007). From the user's perspective, it constitutes the system's itself since they do not interact with the system's underlying back-end architecture (Fadeyev 2009). The user communicates with the system by making "inputs" in the UI which, in turn, conveys the results of those inputs through "outputs" typically presented on a monitor/screen. The oldest type of UI is a so-called Command-Line Interface (CLI) where the user interacts with the system exclusively through typing. Today, the most common form is a Graphical User Interface (GUI) which is sometimes referred to as the "WIMP" since it includes Windows, Icons, Menus and Pointing mechanisms. Apart from the use of a typing device (e.g., a keyboard), the user performs tasks/actions by selecting and manipulating text or graphical objects with a pointing device (e.g., a computer mouse) (Galitz 2007).

A way to help users to understand and picture what a system concept/idea intends to deliver is to create a *prototype* or *mock-ups* of its UI. This constitutes concrete (but partial) representations of a system's design (Lowdermilk 2013). They may be used to demonstrate crucial features or the interface of a proposed system, but it is primarily a way for designers/developers to involve clients in the evaluation of design ideas (Benyon 2014). They are created to have something tangible to test with prospective users (Lowdermilk 2013). Letting users try to work with a preliminary prototype allows for a collection of user inputs in an on-going design process (Garrett 2010). Visual representations of options and possibilities make it easier for the potential user to effectively communicate what they are (or would be) looking for (Lowdermilk 2013). A great advantage of using a prototype is that a design can be evaluated without significant (or any) programming efforts (Lowdermilk 2013; Rubin & Chisnell 2008) which, in turn, ensures that time and resources are not spent on implementing a solution that ultimately does not work (Lowdermilk 2013). In contrast to a final product, prototypes are easily disposable, meaning that early design proposals can be extensively revised or even completely replaced with another design that works better (Rubin & Chisnell 2008).

A prototype can be classified as *low-fidelity* or *high-fidelity*. The former typically implies that the prototype is static with no real interactivity. It may e.g. represent sketches on paper or a digital representation of a basic interface. It has limited/no functionality so people are forced to imagine or "fake" the behaviour of the intended system. A high-fidelity prototype is usually created in a computer software so that it is interactive and mostly "work" as a real product/system would. Its appearance is also more detailed and reminiscent of a real product/system (Saffer 2010). A low-fidelity prototype gives the impression that the commitment made to the overall design is still small. Test participants may, therefore, feel more encouraged to question core concepts as well as features. Prototypes that are too polished and rich in details and functionality, on the other hand, might cause the test participant to assume that the product/system is near completion. Consequently, they may be less likely to comment on questionable designs when a high-fidelity prototype is employed (Lowdermilk 2013). Furthermore, if a UI (prototype) is too visually appealing, it may be presumed that a system (will) have great usability and prospective issues in the design may be overlooked during evaluations (Moran 2017). However, in cases where complex functionality is to be presented, a richer and more complete prototype (that is interactive) may be preferable as it is hard for participants to imagine how it works if they are unable to play around with it. In such a scenario, more accurate feedback may be received through the use of a high-fidelity prototype. Nevertheless, participants should be made aware that it is simply a prototype so that it is clear that the design is not final (Saffer 2010).

3.3 Evaluating System Design Ideas

Throughout the PRISMACLOUD project, a *human-centred design* approach was utilized (Alaqra et al. 2017). This is defined by ISO 9241-210:2010 as an "approach to systems design and development that aims to make interactive systems more usable by focusing on the use of the system and applying human factors/ergonomics, and usability knowledge and techniques". According to Rubin and Chisnell (2008), there are numerous techniques/methods that can be utilized in a human-centred design approach to evaluate prototypes or concepts of systems – e.g., *focus groups, walkthroughs, heuristic evaluations*, and *usability testing*.

In **heuristic evaluations**, a system or prototype is reviewed by a usability or human factors specialist based on usability principles (i.e., heuristics) from previous research/literature or the expert's own professional experience (Rubin & Chisnell 2008). Typically, this technique may easily identify "surface issues" in the user interface (e.g., misalignments and unclear labels). However, the more complex the interface is, the more likely it is that serious issues will be missed. The evaluator is typically a specialist in usability or human factors but not necessarily an expert of the area/domain that the system falls into. Problems related to user tasks and workflows – which are more critical to the users – may therefore *not* be discovered. Moreover, heuristics are typically general in nature and might be interpreted differently by different evaluators. That is, aspects classified as a problem by one evaluator may not be seen as an issue by others. The severity of an issue may also be estimated differently. Thus, heuristic evaluations should preferably be supplemented with other methods to solidify the result from such an assessment (Wilson 2013).

In **usability testing**, representative end-users are observed while they perform given tasks in a system/prototype. A formal approach may be employed where true experiments are conducted to confirm a set of hypotheses. Alternatively, an informal approach may be utilized where an iterative cycle of tests are performed to identify usability deficiencies and to shape the system design progressively. Usability testing may be great for observing behaviour and measuring performance (Rubin & Chisnell 2008). However, compared to focus groups or interviews, it is less appropriate for exploring people's preferences and opinions. Feedback on the concept (and appearance) of a system are difficult to collect (Wilson 2009). During a usability test, the moderator (and prospective observers) should keep the interaction with the test participant at a minimum since their comments may provide the participants with cues and influence their behaviour and ability to perform the tasks in the system/prototype. The moderator and test participant could instead interact through pre- and post-test questions (Rubin & Chisnell 2008). However, users typically do not memorize the content at hand when utilizing a website/system (Galitz 2007) and information that is no longer in sight tends to also be out of the user's mind (Brinck et al. 2002). Furthermore, the participant is often focused on solving a problem/task rather than mentioning perceived oddities in the user interface (Wilson 2009).

While heuristic evaluations and usability testing may be conducted in a later stage of the system's development lifecycle, **focus groups** can be used to evaluate preliminary concepts with a group of representative users. The objective of focus groups is to discover how acceptable a particular concept is, but the method may also be used to confirm or identify the characteristics of end-users. It is useful for gathering information about systems that people are (un)willing to purchase, and to explore people's feelings and way of thinking. The method is less appropriate for learning about the real behaviour of prospective users. People are only reporting what they feel like telling the researcher, which may not reflect the reality (Rubin & Chisnell 2008). Given that focus groups constitute a group activity, they tend to end up at *consensus*. The majority or loudest opinion will often speak for the entire group, since participants with another view may be reluctant to express their disagreement

(Cooper et al. 2014; Wilson 2009). Thus, focus groups may be insufficient for eliciting *all* opinions and behaviour patterns that should be accommodated by a system (Cooper et al. 2014).

In **walkthroughs**, people go through numerous screens of a system/prototype while being asked for their reaction (Wilson 2009). Such a method may be utilized to evaluate the sequence of steps in a particular task (Galitz 2007) and to investigate how a prospective – or hypothetical – user may fare with the intended system (Rubin & Chisnell 2008). The method helps identify labelling and object placement issues in the interface at an early stage in the design process. In comparison to usability testing, people are more likely to make comments on e.g. inefficient tasks, privacy concerns and designs that they dislike during a walkthrough (Wilson 2009). Often, walkthroughs do not involve actual end-users (Preece et al. 2015). The system/prototype may rather be evaluated by designer or researcher colleagues (assuming the role of a user), which are guided through a task in order to envision how actual users will utilize the system (Rubin & Chisnell 2008). However, the walkthrough method comes in various forms which involve different sets and types of participants:

- **Cognitive walkthrough** – Usability specialists go through the system/prototype from the user's perspective, and perform tasks that represent typical work assignments for the user (Vu & Proctor 2011). This approach is based on the notion that real users are not required for the evaluation. Instead, a single or group of usability experts are employed (Wilson 2013). The focal point is on assessing the ease of learning (Preece et al. 2015; Vu & Proctor 2011; Wilson 2013).
- **Pluralistic walkthrough** – involves a group with representative users, developers as well as usability specialists, each of which assumes the role of an end-user. A few prototype screens are given to each evaluator who writes down the sequence of actions that they would take to move from one screen to another (without consulting with the other group members). Subsequently, a group discussion is held where each evaluator describes the actions they have suggested on each screen (Preece et al. 2015).
- **User walkthrough** – individual users are asked to walk through a system or prototype in order to gain insight into how well the system suits the user's expectations (Vu & Proctor 2011). In walkthroughs with actual or prospective users, they are often asked to "think aloud" while interacting with the system or prototype. Perceived difficulties of using the prototype can thereby be elicited (Noyes & Baber 1999). The researcher/walkthrough moderator should pay attention to whether the user comprehends the system and how to complete a given task; whether the user thinks that the interface provides adequate feedback in response to his/her actions; and whether the user knows where to go after a particular action is performed. This research method essentially constitutes a variation of semi-structured interviews. The researcher's areas of concern can be used as talking points to learn about the user's opinion on the matter. Questions that the *user* asks about the system (e.g. in relation to their current working tools and environment) should also be taken into consideration (Vu & Proctor 2011).

3.4 Ethical Considerations

According to the Swedish Research Council (2002), there are four main principles of ethics in social science, two of which concerns the phase where data is to be collected, i.e.:

(1) *The principle of Information* ("Informationskravet") – stipulates *what* the researcher is obligated to inform the respondent about prior to their participation in the study. Namely, the respondent's task(s), as well as the purpose of and conditions around their participation should be clarified. Furthermore, any elements that may affect the respondent's willingness to partake in the study should be explained (Swedish Research Council 2002).

(2) *The principle of Consent* ("Samtyckeskravet") – stipulates that the researcher should obtain the consent from the respondent (or from a parent/legal guardian if the respondent is under 15 years of age). The respondent has the right to decide to what extent they will take part in the study. If the respondent wishes to discontinue their participation, they should be able to do so without suffering from negative consequences or being pressured into changing their mind (Swedish Research Council 2002).

In accordance with the aforementioned principles, Bryman (2012) describe the importance of *informed consent*. That is, the respondents should be given as many details as necessary to make an informed decision about their participation in the study. Similarly, the Ethical Guidelines defined by the UK *Social Research Association* (2003:28) describes that obtaining informed consent ensures the respondents understand the confines of their participation, what they will be exposed to, and the risks that may incur. The researcher should not deliberately withhold information that may cause reluctance to partake in the study. Furthermore, the respondents should be informed that they are entitled to refuse involvement at any stage, for whatever reason, and that data given to the researcher can be withdrawn.

The remaining two principles, defined by the Swedish Research Council (2002), apply to the storage of collected data and reporting of findings:

(3) *The principle of Confidentiality* ("Konfidentialitetskravet") – stipulates that personal identifiable information, collected during the study, should be reported and stored in a confidential manner so that the respondent's identity is not disclosed to third parties (Swedish Research Council 2002).
(4) *The principle of Data Usage* ("Nyttjandekravet") – stipulates that collected information should only be utilized for the sake of the study, and not for commercial/non-scientific purposes. If data is used for actions that directly affect the respondent, a special consent has to be obtained from the individual in question (Swedish Research Council 2002).

According to the Swedish Ethics Review Act (SFS 2003:460), an ethical approval[9] is required for research performed in Sweden if it aims to *influence* or run the risk of *damaging* a person – mentally or physically. The same applies if the research that involves processing of sensitive personal data.

3.5 Considerations Related to an e-Government Use Case

As earlier mentioned, this study will utilize Archistar and the corresponding e-Government use case as an example of a multi-cloud solution based on Secret Sharing. The use case is described and analysed in Appendix A whereon the following interpretations are made:

(1) The solution can be used for both backups and archiving of data.
(2) Both private/internal and public/external CSPs can be employed in the multi-cloud solution.
(3) The customers may be public authorities – or citizens/private businesses.
(4) The end-users are individuals with great IT knowledge.
(5) Instead of relying on an external auditor, the end-users should be provided with means to evaluate the multi-cloud solutions themselves.

The intended solution would/could include multiple different features, of which this study focuses specifically on the creation of configurations for data that are to be protected in the cloud. System configurations are typically rather complex and in the context of organizations, such tasks may be

[9] From the Swedish Ethical Review Authority.

primarily performed by system administrators. On the other hand, means for *auditing* would not be the focal point of this study but in order for such features to be successfully utilized in (the final product of) the intended solution, it would be assumed that the end-user *may* need a high technical knowledge to validate whether security-related requirements are fulfilled. Although less knowledgeable individuals may be capable to use *some* of the system's features, it would be presumed that only users with a higher understanding of IT may be able to utilize *all* functionality to its full potential.

Respondents/participants in the interviews and the evaluation of the UI proposal (described in Section 3.6 and 3.7) should represent prospective end-users. In other words, they should have an IT knowledge that exceeds the understanding of "lay-users".[10] The study would not be restricted to only people working at a public authority, meaning that private individuals and/or workers at private companies would also be considered as suitable candidates. Moreover, in order to answer interview questions, the respondents/participants should have previous experience of cloud storage and be *qualified* to perform system administrative work. Since the intended solution will first and foremost be utilized in organizations, interview questions should preferably be answered from an organizational point of view. However, if the respondents would only be able to answer from a private use-perspective, their replies would still be considered as valuable for the study. There were no requirements in terms of the size of their organization so long as it utilized – or was planning for using – cloud storage.

During the interviews, the solution's perceived applicability when utilizing private/internal or public/external CSPs would be investigated. During the evaluation of the UI prototype, the participants would be tasked with creating a configuration (further described in Section 3.7.1). They would be asked to select both private/internal and public/external CSPs to utilize in the multi-cloud solution. When specifying the characteristics of their data, the intended solution could be thought of as a security measure for both data backups and archiving projects.

3.6 Setup of Interviews

A total of 16 individuals were interviewed; each interview session was either conducted in person or through a peer-to-peer application (i.e., Skype or GoToMeeting). 12 respondents were based in Sweden, 2 in Germany, 1 in Austria, and 1 in Italy.

Among the respondents were one *Software Developer* based in Germany, and one Swedish *Security Engineer* at a multinational IT-company. 2 respondents worked in a profession involving customer support in Sweden: One as an *IT Consultant*, and one as a *Manager in Consulting* that provided security services/products to customers. 2 respondents had already heard of Archistar since their organizations had intended to use the solution in the future: One of them was a *Capacity Manager* at LISPA (i.e. a regional ICT provider in Lombardy, Italy), and the other one was a co-founder of a start-up cloud infrastructure company in Austria. Moreover, 5 *PhD students*, 2 *Lecturers* and 1 *Professor* were interviewed at a university in Sweden – all within the area of Computer Science. Other participants working on a campus were 1 *IT-Security Coordinator* and 1 *Project Leader/IT architect*, both of which were knowledgeable in how data backups are stored and handled at their respective universities in Sweden.

A considerable proportion of the respondents based in Sweden were working at a Computer Science department at a university that constituted a public authority. None of the respondents was directly involved in the PRISMACLOUD project, and did not have in-depth knowledge of its on-going

[10] That is, individuals that lacks professional IT knowledge or experience.

research prior to the interview. Other respondents were recruited through colleagues within the project or respondents that already had participated in the study (i.e., via so-called snowball sampling).

Before starting the interview, each respondent was shown a short introduction video[11] to be familiarized with the concept of Secret Sharing. They were informed about the purposes and circumstances of the study. Moreover, they provided an informed consent for any data collection and voice recording during the interviews (see Appendix B). Since no sensitive personal data were to be collected or processed during the interviews, no ethical approval was needed according to the Swedish Ethics Review Act (SFS 2003:460).

A standardized questionnaire was utilized during the interviews (see Appendix C). Thereby, questions were asked verbally while simultaneously presented in a fillable PDF form, which would serve as visual aid. The respondents would be able to answer interview questions both in spoken and written form, without having to keep prospective answer alternatives in memory. During interview sessions conducted via a peer-to-peer application, the questionnaire would also ensure that the respondent would "receive" and be able to answer questions even if the audio would be of poor quality.

The questionnaire had been created by consulting with other researchers within PRISMACLOUD. Furthermore, it was presented during a meeting with a project advisory board, where meeting participants tried to fill it out and pointed out flaws/ambiguities. Based on the received feedback, the questionnaire was revised and finalized, before interviews were conducted.

Some interview questions were closed-ended since the respondents should be requested to choose between particular options (e.g., for classifying data and specifying the preferred type of security measures). Closed-ended questions were followed up by open-ended enquires, asking the respondents to justify selected answer alternatives in their own words. Thus, even though closed-ended questions were used, qualitative data was still collected while these were answered.

Although a standardized questionnaire was used, questions were sometimes verbalised by the researcher with a slightly different phrasing. Thereby, the wording in the PDF form would be complemented with clarifications when needed. Furthermore, during some interviews, questions were covered in a different order to fit the discussion more properly. Additional questions (outside of the PDF form) were sometimes asked when the respondent's answers needed further explanation.

Respondents that were not interviewed in-person, were sent the questionnaire and introductory material via e-mail. The respondent opened the material on their local device and shared their screen with the researcher via Skype or GoToMeeting. Once the interview was over, the respondent was asked to return the completed questionnaire to the researcher via e-mail.

The interviews had been estimated to take 40 minutes each. In order to keep within this time frame, each topic/section was given a specific number of minutes in which corresponding questions should be answered. All topics were discussed during each interview session, but every question in the PDF form was not asked in some cases due to the time limit.

The structure of the interview questions can be summarized in Table 3.

[11] https://www.youtube.com/watch?v=4_jx2V1z-2U

Table 3. Structure of interview questions.

Section	Topics	Subtopics
#1	**Background questions:**	• Demographic information. • Classification of cloud data in terms of confidentiality, integrity and availability requirements. • Data threat perceived as most severe in the cloud (i.e., data loss, availability issues or breaches of data confidentiality).
#2-3	**Geographical distribution:**	• Requirements in terms of the Secret Sharing parameters (i.e., N and k). • Requirements in terms of minimum distance between servers. • Trusted countries/regions for prevention of data loss/availability issues, breaches of data confidentiality, and collusion between CSPs.
#4	**Factors for trust in CSPs:**	• Perceived importance of compliance with privacy legislations, possession of a privacy/trust seal as well as high trust ratings.
#5	**Applicability of Secret Sharing:**	• Perceived adequacy of Secret Sharing in the context of private clouds, public clouds, community clouds with private individuals and community clouds with public authorities.
#6	**Security measures:**	• Perceived advantage and disadvantage of Secret Sharing's keyless nature. • Preferred security measures out of Secret Sharing and Standard Encryption for protection of data in the cloud.
#7	**Security trade-offs:**	• Prioritization of key factors (i.e., Security, Cost, Usability, Performance, Reliability). • Perceived trade-offs/correlations between them.

3.7 Setup of User Walkthroughs

In parallel with the interviews, a pilot study was conducted within PRISMACLOUD where a prototype for a preliminary Archistar user interface (UI) was developed and evaluated. The prototype was created by the thesis author and Pettersson[12] in collaboration with LISPA representatives.[13] Furthermore, during a PRISMACLOUD plenary meeting in July 2017, LISPA representatives presented a mock-up for an alternative UI solution (which had not been evaluated). Each of the previously proposed UI solutions is further discussed in Appendix F.

In the confines of this thesis, a *new* UI prototype would be developed based on the requirements collected in the interviews. The prototype would constitute a UI proposal for a decision-making support system, corresponding to the solution's Dealer component (described in Appendix A, Section 1.3). That is, the focal point would be means for creating "configurations" for upcoming data backup/archiving projects. A configuration serves as a description/plan of how the Secret Sharing mechanism and multi-cloud infrastructure should be arranged to safeguard data. The UI prototype was created using the prototyping software *Axure RP8*[14]. Thereby, some level of logic could be simulated on each screen/step of the configuration process. Researchers within the PRISMACLOUD project was

[12] PhD., Pettersson, J.S.
[13] This pilot study was reported in the PRISMACLOUD deliverable *D3.2 HCI Guidelines* (Alaqra et al. 2017).
[14] https://www.axure.com/

consulted during the creation of the UI prototype (see Appendix G for a description of design decisions).

The user walkthroughs were performed while the PRISMACLOUD project was active and the prototype being evaluated should be seen as a *preliminary* UI proposal – and not a final product. The overall structure of the configuration task was yet to be confirmed. The purpose of the assessment would not be to simply identify usability issues (as in Heuristic evaluations), or to verify whether the user would be able to perform a well-defined task (as in Usability testing). The intention was to also evaluate whether the proposed workflow (i.e., the division into – and sequence of – steps) would be realistic and suit the situation of prospective users. In order to do so, the user's perception and opinion on the matter would be investigated. Moreover, providing the respondents/participants with a tangible example of how the solution may look like would allow for further elicitation and refinement of user requirements. Even though focus groups (or group walkthroughs) would allow the researcher to gather information about the user's acceptance of the Archistar concept, varied opinions may be difficult to identify in a group activity. Thus, the prototype would be evaluated through *individual* walkthroughs with 5 prospective users.

The walkthrough participants constituted 1 *Administrative Director* at a municipality's IT department, 1 *IT Security Coordinator* at a Swedish University (who also had the position of a Data Protection Officer), 2 *System Administrators* at a Swedish University, as well as 1 *IT Security Expert* (who worked part-time as a *Security Consultant* in industry for many years). The participants were recruited because of their complementing expertise. The participants were recruited because of their complementing expertise. In similarity to the previously conducted interviews, the participants represented people qualified to perform system administrative work and that would be able to determine whether the proposed solution is suitable for the intended user.

Before the walkthroughs were conducted with the aforementioned participants, a test round of the setup was performed with the author's supervisor during which some (logical) inaccuracies were discovered in the prototype (i.e., missing hotspots and links navigating the user to the wrong scene). These issues were addressed before the "real" walkthrough sessions.

The participant was briefed about the purpose of and conditions around the study. They were informed that it was entirely voluntary to partake in the study and that their participation would constitute a valuable contribution to science. No compensation was paid. They were asked to give their consent, allowing the researcher to gather data through screen and voice recording, and to use the collected information in the study. No special categories of data were to be collected or processed in connection with the walkthroughs. Therefore, no ethical approval was needed according to the Swedish Ethics Review Act (2003:460). Before the UI prototype was shown to the participants, they were then given a short verbal introduction to the proposed Archistar solution (see Appendix D). Furthermore, they were shown the video[15] used in the previously conducted interviews.

During each walkthrough session, the prototype was presented on a laptop computer. The participant was assigned to operate the laptop to create a configuration for a backup/archiving project in the prototype. They were instructed to think that the project involved data that they normally handle/back up in their daily work. The configuration process was divided into 5 steps, of which the first 3 involved a selection of configuration settings (see Section 3.7.1 for a description of each step in the configuration process). Thus, these 3 steps would constitute the focal point of the evaluation.

[15] https://www.youtube.com/watch?v=4_jx2V1z-2U

The researcher had prepared a list of general questions about the perceived meaning and relevance of UI elements as well as the feasibility of the proposed solution (see Appendix E). Each walkthrough session would involve the same configuration task (with the same sequence of steps) but due to the study's exploratory nature, the author would utilize an approach that is only semi-structured. That is, the author would allow for additional (more specific) questions to be discussed if/when unforeseen topics would emerge during each walkthrough session.

3.7.1 Steps in the Configuration Task

On the **first configuration step** (see Figure 1), the user should prioritize three protection goals/aspects - i.e.:

- "Cost Minimization – Low Cost"
- "Data Protection – High Confidentiality"
- "Data Loss Prevention – High Availability"

Each goal/aspect is represented by an item that should be dragged and dropped in one out of three drop areas to indicate how important it is to the user. (The two protection goals/aspects that are *not* the user's *lowest* priority will be combined and determine how some of the other configuration steps will look like.)

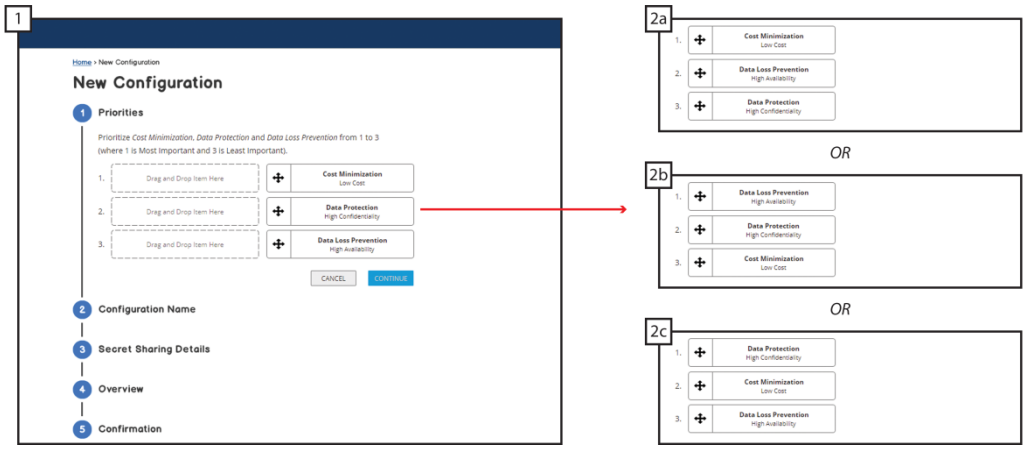


Figure 1. The first configuration step of the walkthrough of the UI prototype.

On the **second step** (see Figure 2), the user should specify a *reference name* to the configuration. (If all configurations created by the user were to be summarized in a *list* in the final system, the reference name could help the user to tell them apart.)

Depending on the previously made prioritization, the user may also have (the option) to add a layer of encryption on the second configuration step. That is:

- If "Cost Minimization – Low Cost" and "Data Loss Prevention – High Availability" are most important, encryption should be *avoided*. The feature is hidden [2a & 3a].
- If "Data Protection – High Confidentiality" and "Data Loss Prevention – High Availability" are most important, encryption should be *optional*. The feature is disabled but can be enabled via a radio button [2b & 3b].
- If "Cost Minimization – Low Cost" and "Data Protection – High Confidentiality" are most important, encryption should be *required* [2c & 3c].

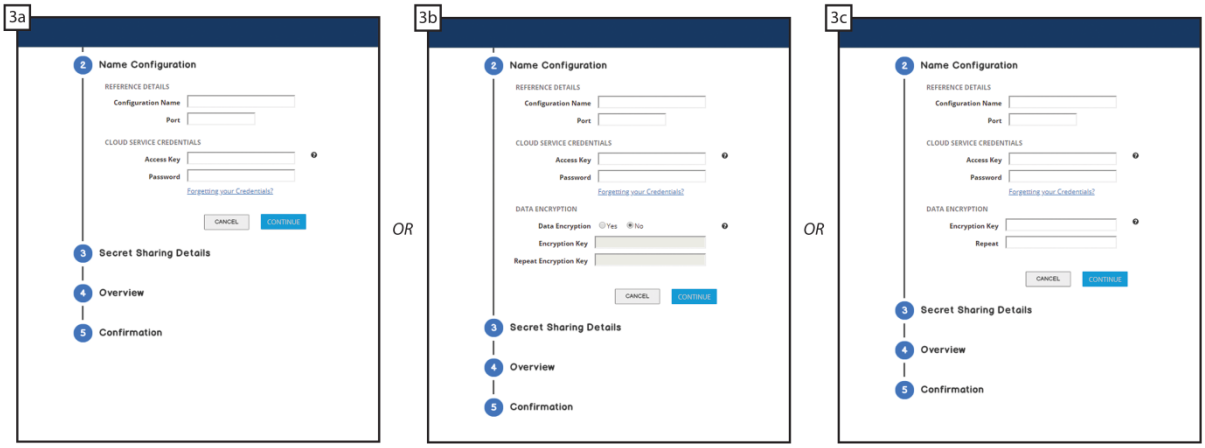


Figure 2. The second configuration step of the walkthrough of the UI prototype.

On the **third configuration step** (see Figure 3), the user should indicate their storage space needs by describing some attributes of the data that is to be backed up/archived (i.e., "Data Retention Period", "Estimated Size of Data" and "Rate of [Size] Increase").

Furthermore, the user should select values for the Secret Sharing parameters (i.e., "Number of Chunks (N)" and "[Data] Restore Threshold (k)"). The input fields would initially be populated with default values that would serve as a recommendation by the system based on the prioritization made in the first configuration step. The "Availability Rate" and "Downtime per Year" provided by the selected combination of N and k would be indicated beside the input fields. During the walkthrough, the participants would be asked to manually change the values to $N = 3$ and $k = 2$. Subsequently, the "Availability Rate" and "Downtime per Year" would be automatically updated in the user interface [5].

Subsequently, the user should select locations to which data chunks should be geographically distributed. The selection would be made from a map where all available data centres are marked with cloud icons. The map features:

- *Filters* – allowing the user to hide/show locations of e.g. public/external CSPs on the map [6].
- *Layers* – allowing the user to utilize different map views so that e.g. "Earthquake Risks" in various areas can be examined [7].

When a cloud icon is selected on the map, information about the corresponding provider's "[Service] Offering", "Location" and "Service Credibility" is listed beside the map [8-9].

When a location/CSP is selected, the corresponding service offering is added to the shopping cart [10]. The "Total Cost" or "Left of Budget" is automatically updated.

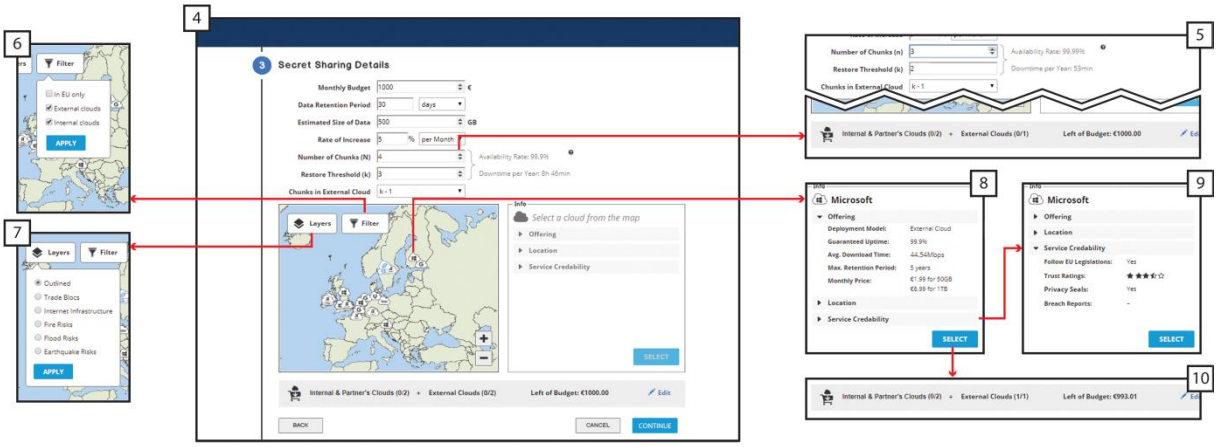


Figure 3. The third configuration step of the walkthrough of the UI prototype

On the **fourth configuration step** (see Figure 4), the user would be provided with an overview of the shopping cart in its entirety before the configuration is completed [11].

On the **fifth configuration step**, the user receives a confirmation that the configuration has been successfully created [12].

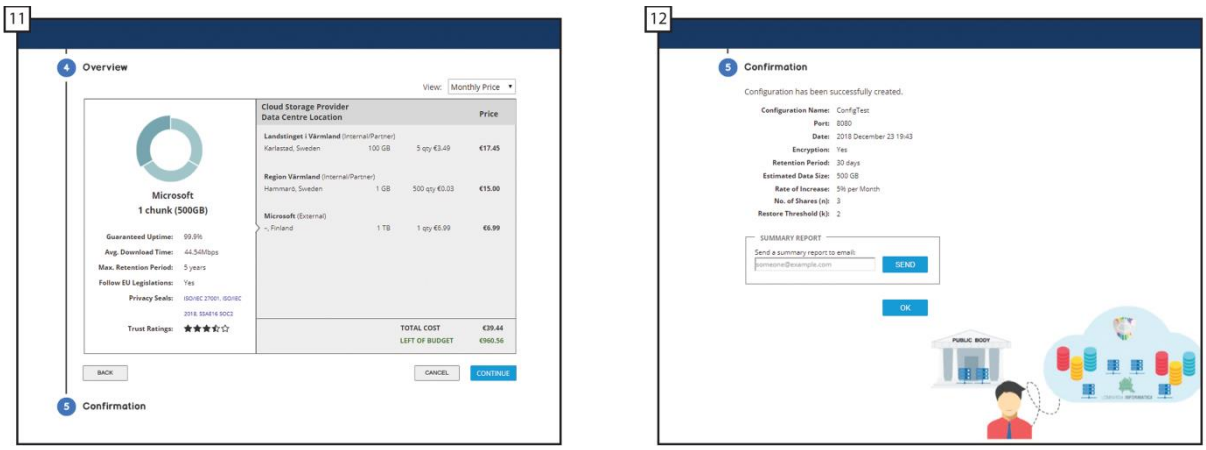


Figure 4. The fourth and fifth configuration step of the walkthrough of the UI prototype

4. Results

Once notes and recordings from the 16 interviews and the 5 user walkthroughs had been summarized/transcribed, a *thematic analysis* was conducted. That is, by comparing data from different interviews and walkthroughs, recurring and contradictory answers were identified and categorized.

The following section is not aligned with the structure of the interviews or walkthroughs but instead presents the outcome of the analysis. The structure below is by topics discussed, especially topics brought up by more than one person.

4.1 Themes topics during the Interviews

Answers regarding Storage Habits and Trust in the Cloud:

- **Most respondents stored more than one type of data in the cloud.** When asked about the type of data that they store – or intend to store – in the cloud, 7 respondents answered from an organizational viewpoint. Some of them did not specify any particular data types because stored data of customers/students/employees could involve any sort of information/files (so long as it follows the organization's regulations). The remaining 9 respondents seemed unable to answer from the perspective of their organization as a whole, but still mentioned data related to their work. 15 out of 16 respondents indicated that cloud storage was used for various types of data or purposes. Common answers where: Documents, code/projects, photos, publications, and general data backups.
- **Data with high requirements for Availability *and* Confidentiality/Integrity may be stored in the respondent's cloud storage.** 15 out of 16 respondents stored a type of data in the cloud whose *Availability* could be valued to the same extent as its *Confidentiality* and/or *Integrity*. 9 of them stated that the highest level of (i.e., strictest possible) requirements were needed for Availability and at least one of the other parameters.
- **Most respondents prioritized protection against one data threat over another.** When asked which type of data threat is regarded as most severe (i.e., most important to be protected against) in the cloud, 11 out of 16 respondents selected "Data Loss" over "Breaches of Data Confidentiality" – or *vice versa*. 4 respondents had previous experience of losing data, either in traditional or cloud-based storage. However, when it came to data breaches, 4 respondents stated that they were unaware of whether or not they had been subject to such a threat.
- **Cost may be an important factor for some users/organizations.** 5 respondents indicated that the proposed solution has to be affordable and cost-beneficial in order for them to use it. In contrast, 5 other respondents seemed to think that the cost was a less crucial factor, either because they were not in a position where they personally would have to pay for the solution within their organization or because they saw a trade-off with security. However, one of them still acknowledged that the cost could be important for others.
- **Personal mistrust towards clouds.** 6 respondents expressed some level of personal mistrust towards the cloud as a storage medium. 2 of them argued that they would generally not store their own sensitive data in clouds, because they are not secure by nature. One respondent stated that there were certain data that would not be placed in a *public* cloud. Another 2 of them indicated that – although they did use cloud storage – they would avoid clouds as much as possible.
- **Organizational/legal restrictions on cloud storage.** 6 respondents described that there were organizational or legal restrictions to what is stored in the cloud. Guidelines and privacy regulations were mentioned by 5 out of 6, of which one of them described that they currently were not allowed to store *any* information in the cloud within her institution (due to data loss and privacy issues) – a rule that she admitted to breaking. Another one explained that his university would rather store sensitive data in a secure local system without Internet access, since the

university would not like to be held accountable in case of a data breach in the cloud.

Answers regarding the Configuration of Secret Sharing Parameters and the Geographical Distribution:

- **Implications of different values on k may not be completely clear.** 8 out of 16 respondents described aspects that they would need to know in order to make a decision in terms of total number of chunks (N) and the threshold for data reconstruction (k). That is, 4 of them mentioned *trustworthiness of CSPs*, whereas *availability (of individual storage nodes/clouds)* and *sensitivity of data* were each brought up by 2 respondents.

  Moreover, 5 respondents considered what different values on N could implicate specifically: 4 of them indicated that a higher number of chunks could result in *higher costs*. *Lower risk of data loss* and *higher processing time* was also suggested by one respondent each. However, the consequences of different values on k appeared to be reflected upon by only 2 respondents. In other words, the Secret Sharing parameters were seldom considered in relation to each other. Therefore, values/numbers selected by the majority of respondents seemed to be arbitrary rather than definite.

  During the interviews, respondents were provided with the hint that the minimum values were $N = 3$ and $k = 2$. 5 respondents kept these numbers for all data. Another 6 respondents suggested several combinations of N and k, all of which also included the minimum values as a potential option. The highest selected numbers were $N = 12$ and $k = 5$ and were suggested by a respondent who had some previous familiarity with Secret Sharing.
- **Divided answers on the question about minimum distance.** When asked about the minimum distance between data centres, the respondents had the options to specify a distance in *kilometres*, an *administrative level* that data centres should be divided into and *climate zones* in which they should be located. The majority of the respondents (i.e., 14 out of 16) answered in terms of administrative level, whereas answers in kilometres were the least common (i.e., only given by 8 out of 16). Out of the answers received on the first two options, "Different Countries" and "100km" were the most frequent and were suggested by 9 and 4 respondents respectively. The respondents that selected the same minimum distance in kilometres did not necessarily have the same distance requirements in terms of administrative level.

  Climate zones were selected by 9 respondents, all of which chose "Cold" and/or "Temperate". Later, when the "most trusted countries/regions for data loss prevention" were to be specified, 8 out of 9 respondents contradicted themselves by selecting countries/regions that were either partly or entirely outside of the selected climate zones.
- **Europe, Canada and Australia & New Zealand, most trusted areas for preventing both data loss and breaches.** In order to prevent data loss, the top trusted countries/regions were: (1) Rest of Europe[16], (2) Canada, (3) Australia & New Zealand, (4) US, and (5) Japan. When asked about the reasoning behind their rankings, 5 respondents spoke about *laws/regulations*, while *safety from natural disasters* and *the infrastructure's reliability* were mentioned by 4 respondents each. Furthermore, *connectivity/access time* and *political stability* and were both brought up by 3 respondents.

  In order to prevent breaches of data confidentiality, the top trusted countries/regions were: (1) EU (incl. EEA), (2) Canada, (3) Australia & New Zealand, (4) Japan, and (5) US. *Laws/regulations* represented a determining factor for 9 respondents, while *stability* and *(geo)politics* were each mentioned by 4 respondents. Moreover, 2 respondents spoke in terms of *democracy*. Some respondents indicated that certain countries were excluded from both "most trusted" lists because they did not have enough knowledge about the conditions in these

[16] I.e., areas in Europe outside of the continent's southern/southeastern portion.

countries/regions.

Answers regarding Trust in CSPs:

- **Contradictions regarding the trustworthiness of CSPs.** 9 out of 16 respondents stated that they, in the position of a Secret Sharing user, would consider utilizing CSPs that they normally would not trust in a single cloud solution. However, when later asked how important it would be that CSPs follow *privacy legislation(s)*, all 9 of them argued that it was either important or very important. Moreover, 5 out of 9 thought it was important or very important that the CSPs have a *trust/privacy seal*, and 7 out of 9 thought it was important or very important that the providers have *high trust ratings*.
- **Mixed level of concerns regarding collusion.** When given the statement "I am concerned that cloud storage providers will collaborate and reconstruct my data behind my back", 7 out of 16 respondents disagreed or strongly disagreed, while 6 respondents answered the opposite.
- **Distribution to EU/EEA countries may be required to prevent collusion and to increase trust.** When asked to select three countries to prevent collusion between CSPs, 5 respondents chose nations that were all located in EU/EEA, while 3 respondents selected two EU/EEA-countries and 4 respondents selected one. The most commonly selected area was Germany and the Nordic countries (indicating that the respondents preferred to have their data nearby), followed by Canada and US. 7 respondents stated that a reason for their selection was *laws/regulations*. 5 respondents argued that a determining factor would be *political relationships* between countries (e.g., whether they are in conflict with each other or have non-mutual political interests), while 2 respondents mentioned *publically known incidents in the past*.

  When the respondents were asked about which privacy laws the CSPs should follow, 10 out of 16 respondents answered "European" or "EU legislation". 8 respondents also suggested that CSPs should be compliant with local/national laws of a specific country (such as Germany).

Answers regarding the Perceived Applicability of Secret Sharing:

- **Secret Sharing may be adequate and beneficial for Private clouds.** 6 respondents stated that Secret Sharing, as a security measure, would be adequate for private clouds and 5 respondents indicated that they would benefit from or be interested in using it. 1 respondent only saw benefits for private use, while 2 respondents thought it was unnecessary altogether. 3 respondents seemed sceptical that e.g. only the data owner would be able to reconstruct the data.
- **Secret Sharing may be inadequate for Public clouds.** When asked whether Secret Sharing is secure enough for public clouds, 3 respondents stated that it was, and 6 respondents indicated that it would not be (perceived as) sufficient. One of the latter argued that he personally thought it was adequate, but many people would think differently due to the difficulty of determining the level of security and a general wisdom that nothing is 100% secure.
- **Secret Sharing potentially not trusted by public bodies in Community clouds.** 3 respondents thought secret sharing would be adequate for community clouds with public bodies, but one of them pointed out that these institutions have too high security standards to actually trust it. 3 respondents argued that it would not be secure enough. One argued that province governments and city councils do not have a high IT maturity, while another one stated that they are too conservative and would not trust such a solution.

Answers regarding the Perceived Advantages of Secret Sharing's Keyless Nature:

- **No key loss issues.** 6 respondents indicated that Secret Sharing would be less vulnerable since one does not have to worry about losing a key.
- **Convenient for the user.** 6 other respondents seemed to think that it would be beneficial for

practical reasons. 3 of them suggested that it might be less complicated for the user and one of them argued that the usability would be improved. Moreover, one of them stated that the user would get rid of issues related to the key management lifecycle.

- **Better performance.** 2 respondents believed that Secret Sharing could bring performance advantages. One of them stated that the solution would consume fewer resources on the client's device, since more interactions and operations are performed on the 'server side'. This was argued to be beneficial for mobile devices.

Answers regarding the Perceived Disadvantages of Secret Sharing's Keyless Nature:

- **Not as secure.** 4 respondents thought that the data would be less securely protected without a key. All of them appeared to (incorrectly) assume that not only the data owner would be able to gather all the chunks and reconstruct the data, so a key would therefore be necessary. Moreover, another respondent had become used to encrypting 'everything' and seemed to desire a key out of old habit.
- **Level of security not easily proven.** 4 respondents described that it would be hard to know whether the solution is secure or not. One of them mentioned the difficulty to determine whether the data are handled as promised by the providers.
- **Risk of Collusion.** On a similar note, 3 respondents described the plausibility of data being disclosed or stolen due to collusion between CSPs. It was suggested that such an incident could be enforced by law enforcement, or that the providers could collaborate illegally behind the users back.
- **More to keep track of.** 2 respondents argued that one has to handle more storage spaces and user accounts since the data chunks are distributed to separate CSPs.
- **Complicated.** 2 respondents suggested that certain tasks would become more difficult to perform when the data are split up into chunks and distributed to different providers. The tasks mentioned where restoring from data backups, and sharing data (e.g. pictures) with family members or friends.

Answers regarding Security Measure Preferences:

- **Secret Sharing inadequate for sensitive data.** Only 2 out of 16 respondents stated that they would like to use Secret Sharing on its own to protect *all* their data in the cloud. However, both of them pointed out that if they were to store sensitive data in the cloud, they would like to combine Secret Sharing with Encryption. Similarly, 4 other respondents described that an extra layer of encryption would be needed for (highly) sensitive or classified information. Mentioned example were data in health care applications and documents that only an employer should be able to unlock.
- **Encryption perceived as stronger protection than Secret Sharing.** 5 respondents argued that data would be better protected with Encryption than Secret Sharing. One of them worked for a company that provided data storage to its clients and a customer-managed encryption key could, therefore, be required by legislation. Another one of them argued that authorities and institutions should always use Encryption since this is the only adequate security measure against data breaches. 3 respondents stated that they would like to use only Standard Encryption for data in the cloud.
- **Secret Sharing combined with Encryption seen as the best protection.** Although some appeared to be satisfied with just utilizing Encryption, 12 respondents recognized that a combination of Secret Sharing and Encryption would be the best solution.

  6 respondents preferred to use both security measures for all sets of data in the cloud. 4 of them had earlier specified that all their data could have high requirements in terms of Integrity

and/or Confidentiality, indicating that the information may be rather sensitive. 2 of them admitted that both security measures would be needed simply because they did not have sufficient knowledge about how Secret Sharing would work in reality.

Numerous respondents indicated that they would need to know more about the proposed Archistar solution in order to give definite answers on the interview questions. Their trust in the Secret Sharing concept may depend on…

- …the organization behind the solution, and the underlying infrastructure.
- …if one can assure that the data are protected according to its classification.
- …if one can prove that one cannot extract any information by gathering one chunk.
- …if one can share information with others without giving out the login to one's user account.
- …if the reconstruction of chunks is performed on the user's local device. If data chunks are combined in the Secret Sharing application before being transferred to the user's device, the benefits of the solution would be lost since it is still a single target point for e.g. attackers.
- …if data availability is high enough. One should be able to assure that the data are not lost or inaccessible when it is needed. Need to plan for permanent loss of some shares.
- …if the Secret Sharing mechanism is open-source.
- …if the data centres, in which data chunks are stored, are certified. Certifications have to be made by a well-known and trustworthy organization.

4.2 Themes identified during the User Walkthroughs

Comments regarding the Proposed Prioritization of Protection Goals/Aspects:

- **Proposed set of prioritization options mainly perceived as sufficient for the configuration.** On the first configuration step, where users' priorities should be specified, 3 out of 5 participants stated that the available options covered all relevant protection goals/aspects for backups/archives in the cloud. For instance, one participant described scenarios where each of the three protection goals/aspects would be the highest priority. "Data Protection – High Data Confidentiality" would be regarded as the most important factor when the data is (highly) sensitive. "Data Loss Prevention – High Data Availability" would be the main priority when backing up the most crucial information within the organization (i.e., data that the organization's survival may depend on). For the remaining data (i.e., information that is neither sensitive nor critical for the organization's existence), "Cost Minimization" would be prioritized first since the expenses need to be kept low.
- **The difference and meaning of prioritization options may not be clear to all users.** Although the protection goals/aspects prioritized on the first configuration step were accepted by most participants, they seemed to lack some clarity. One participant questioned the difference between "Data Loss Prevention" and "Data Protection", not noticing the subtitles of aforementioned options (i.e., "High Data Availability" and "High Data Confidentiality"). Furthermore, several participants indicated that "Data Availability" could have several different meanings. One of them pointed out that "Data Loss Prevention" and "High Data Availability" does not necessarily refer to the same thing. Another participant described that "High Data Availability" typically implies either *24/7 uptime* or *quick access time*, but both factors were still included in the "Availability" parameter of the data classification model utilized by his organization. The latter participant argued that if the protection goals/aspects become too similar in the first configuration step, it will be difficult for the user to determine which should be prioritized first. A third participant also spoke about availability in terms of quick access time: some data in the participant's organization needed upon request to be recovered within a limited time period. To avoid latency issues, there were restrictions on the maximum distance between the corresponding organization and employed

data centres.

- **A description of specific needs/restrictions may be preferred over the proposed prioritization.** One participant questioned what "Cost Minimization" and "High Data Confidentiality" actually implied. Instead of prioritizing protection goals/aspects in the UI, this participant preferred a more concrete input into the system, where the user would specify how confidential the data actually is as well as a budget for the archiving/backup project. Moreover, the prioritization procedure did not communicate that Archistar will increase *both* data availability and confidentiality (in comparison to a single cloud solution) and that the purpose of the prioritization is to simply find a suitable trade-off between the two factors (i.e., determining which to improve *the most*).

Comments regarding Security Measure Preferences:

- **The option to add Encryption may increase trust.** On the second configuration step, 5 out of 5 participants wanted to have the option to add encryption for *all* types of data – not only for sensitive information (i.e., when "Data Protection" is highly prioritized). One participant stated that the need for encryption was a matter of *trust*: instead of relying on a single CSP (e.g., Google or Amazon) he would have to trust the Archistar solution, not knowing what actually happens with the data behind the UI.

Comments regarding the Specification of Data Attributes/Characteristics:

- **Input fields for certain data attributes were unclear or difficult to specify.** On the third configuration step, comments were made indicating that the UI did not properly consider the fact that backup systems typically store multiple *versions* of a particular data set. That is, one participant questioned whether "Estimated Size of Data" referred to the size of the original information or all retention points combined, whereas 2 other participants assumed that it may denote the maximum amount of data that the backup/archiving project should be able to hold. Furthermore, one of the latter participants asked whether the field for "Data Retention Period" concerns a certain copy or *all versions* of the data set.

  2 participants also pointed out that the "Rate of [Size] Increase" would be difficult to estimate/predict.
- **Administrators may not have the responsibility to make (organizational) budgeting decisions.** 3 out of 5 participants argued that a budget should be determined prior to the configuration process (carried out by an administrator). The remaining 2 did not make such an objection but still had difficulties estimating a typical "Monthly Budget" on the third configuration step, since they were not responsible for the finances/economy at their respective organization. In other words, it was indicated that certain configuration settings cannot be determined by a single stakeholder/actor in a large-scale organization.

Comments on the Configuration of Secret Sharing parameters:

- **Configuration of Secret Sharing parameters confusing and difficult.** On the third configuration step, pre-selected values for the parameters N and k were also provided based on the prioritization made on step one. Despite the feedback given by the UI, the consequences of different combinations of N and k were (still) not entirely clear to all participants. One participant appeared to be confused by the term "threshold" and was unsure whether a high value on k would imply greater or lower protection against data breaches. Another participant questioned how different values on N and k would influence costs, since this would be a main determining factor when deciding upon a suitable configuration. A third participant expressed that he wanted are a more explicit explanation of what different values would imply (e.g., how will a low or high N

affect the availability as well as the confidentiality of data?).

2 participants stated that they would stick with the default values recommended by the system. One of them pointed out that frequent/expert users may desire the option to change the values for the Secret Sharing parameters but as a first-time user, he would like to be provided with defaults.

- **Alternative solutions for the configuration of Secret Sharing parameters proposed.** Rather than selecting values for the Secret Sharing parameters and receiving feedback about the "Availability Rate" accomplished from the value combination, 2 participants proposed alternative solutions. One of them suggested that the procedure should be *reversed*. That is, the user should select a desired "Availability Rate" from a dropdown list whereupon the system should provide him/her with appropriate values for N and k. The other participant proposed a similar solution but argued that the dropdown list could include options that would relate to both availability and confidentiality (e.g., “slow but secure”, “fast but less secure”). Alternatively, the latter participant thought that CSPs could be selected *first* in the configuration process and the chosen number of providers could potentially serve as the appropriate value for N.

Comments on the Geographical Distribution:

- **Allowing administrators to select geographical destinations for data chunks perceived as sensible.** 4 out of 5 participants indicated that they thought it would be realistic to allow the user to choose locations to which data chunks should be distributed. When it came to the level of abstraction in which locations should be specified, the participants had mixed opinions. One participant suggested that the user should be able to select specific data centres. 2 participants argued that the location details should be country-specific, while another thought that it was only relevant to know whether the data will reside within or outside of the European Union.
- **Selecting locations from a map view may be inappropriate or misleading.** The idea of selecting data centre locations from a map was also received with mixed opinions. 2 participants argued that the administrator should rather choose locations from a simple *list*. 2 other participants appreciated the various views/layers that the map was featuring: One of them suggested that layers for natural disaster risks could come to great use when making risk and vulnerability analyses.

  However, another participant questioned how accurate the layers for flood and fire risks actually could be. The participant worked within a city that historically had suffered from flooding in certain areas, but the site of the organization’s data centres was far above sea level and therefore outside of these risk areas. In other words, even if a city is typically marked as “unsafe”, data centres could be located on an altitude that keeps them out of reach for floods. Furthermore, the participant argued that even if data centres of a CSP is not subject to fires, data could still become unavailable due to a fire at a nearby power station or on the premises of an Internet provider. Instead of a visual representation of natural disaster risks, it was suggested that data centres could be given a “risk value” (possibly as part of a certification).
- **Administrators may *not* be able to choose freely whichever CSPs they want.** Although most participants seemed to think that administrators could/should partake in picking destinations for data chunks, there were indications that they would not have boundless *options* to choose from. That is, 2 participants described that they would merely be able to utilize CSPs that their organizations already have a contract agreement with. Therefore, one of them argued that only contracted providers should be visible to the user/administrator in the UI.
- **Crucial factors for entering a contract agreement with a CSP.** According to one of the participants who was obligated to utilize contracted providers only, their current contract agreements states that data should be stored within the European Union at all time. Another participant described that contract agreements would only be entered with CSPs that can be *fully*

trusted – something that his organization typically determines by evaluating information about service guarantees, measures against natural disasters, as well as ways of handling and reporting incidents. While the Archistar solution would theoretically allow for less trustworthy providers to be employed without putting the data confidentiality at risk (as discussed in Section 2.3.4), the latter participant appeared to think that trust would still be critical.

- **Multiple types of users may have to be involved in the configuration process.** 2 participants stated that some of the details provided in the information container/box (next to the map) would *not* be relevant for administrators (e.g., "Corruption Perception Index"). One of them previously suggested that the map could be useful when making risk and vulnerability analyses. However, the other participant explained that it would not be the administrator's responsibility to determine whether or not a CSP and its data centre location(s) is trustworthy/reliable. Instead, such decisions would rather be made by individuals in charge of service procurements.

  Similarly, a third participant argued that the map would serve as a great tool before contract agreements have been established. The second participant (mentioned above) appeared to have the same opinion, but pointed out that drafting and agreeing upon contracts with CSPs is also not the duty of administrators. It was suggested that the proposed UI may be more suitable for private use or small companies (where one stakeholder may serve multiple different roles) but in the context of a large-scale organization, a single individual would not be able to make all decisions requested in the UI proposal at hand.
- **Misconceptions about legal aspects related to Archistar.** 2 participants made comments that indicated that users may have concerns or the wrong idea about Archistar in relation to EU laws. One of them questioned whether it would be legal to set locations restrictions when distributing data chunks to public/commercial clouds in the EU, since he (wrongfully) assumed that "Free Movement in EU" allows services/data to be moved anywhere within the union. The other participant described a scenario where it was believed to be less crucial to use GDPR compliant CSPs. That is, if each cloud only stores a single data chunk and no information can be obtained from individual chunks, the participant claimed that the data stored in the cloud would no longer constitute personal information (regardless of its original content) and the GDPR would therefore not apply. However, as discussed in Section 2.3.4, Secret Shares classify as "pseudonymous data" and will, therefore, still represent personal information pursuant to the GDPR.

5. Discussion

The following discussion is divided into four main sections. The first two focuses on fundamental decisions that have to be made when creating a configuration of the proposed solution – i.e., the selection of values on the Secret Sharing parameters (see Section 5.1) and the geographical distribution of data chunks (see Section 5.2). In the remaining two sections, the emphasis is on the solution's applicability – i.e., whether it is perceived as a sufficient security measure for cloud data (see Section 5.3) and the extent to which the presented UI is suitable for various user groups with different needs/requirements (see Section 5.4).

5.1 Configuration of Secret Sharing parameters

The interviews indicated that there are numerous aspects that might be essential for the user to know in order to select suitable values for the Secret Sharing parameters. That is, how will the total number of chunks (N) and the threshold for data reconstruction (k) influence the *Cost* as well as the protection of *Data Confidentiality* and *Availability*? Furthermore, even though the Archistar solution would (in theory) allow for less trustworthy CSPs to be utilized without putting the data privacy at risk, some interview respondents indicated that the extent to which CSPs can be *trusted* may still be crucial for the user when selecting values for N and k.

5.1.1 Determining Factors: Data Confidentiality and Availability

In security systems, there is typically a trade-off situation between data confidentiality and availability (Ioannidis et al. 2012) – and the same applies to Archistar. In the context of a Secret Sharing solution, high confidentiality (and integrity) requirements entail that k data chunks should be kept out of the reach of adversaries. High availability requirements, on the other hand, denote that k chunks should be available to *authorized* individuals at all time, meaning that no more than $N - k$ chunks could be destroyed or inaccessible simultaneously. While the latter scenario may demand a relatively *low* data reconstruction threshold (k), the opposite may be true for the former.

Some interview respondents focused on what different values on N would imply, whereas the consequence of different values on k was seldom reflected upon. If the meaning of different values on N and k are not fully understood, or if the parameters are not considered in relation to each other, users may not be able to address the aforementioned security trade-off. Thus, they may fail to find an appropriate balance between the two conflicting aspects (i.e., the confidentiality or availability of data might end up being insufficiently protected/maintained).

One could argue that a configuration UI should help the user by providing a *recommendation* of suitable values on N and k. However, in order for the system to do so, the user has to specify their protection goals/needs first. Due to the security trade-off, one could argue that a *prioritization* should be made between data availability and confidentiality (i.e., even though both factors may be crucial, the user would have to decide which of the two is the *most* important to protect/maintain). During the interviews, many respondents described that they stored data in the cloud that could have high requirements in terms of both availability and confidentiality/integrity. Specifying equally high requirements for both factor would not indicate any priority, so the data classification scheme utilized in the interview questionnaire (see Appendix C) may not be appropriate for describing protection goals/needs in an Archistar configuration UI.

However, the interview respondents were also asked to select a type of data threat that they considered the most severe (i.e., most important to be protected against). The majority of respondents selected data loss over data breaches – or *vice versa*. One could argue that this serves as a prioritization of data availability and confidentiality, since the former should be maintained to prevent data loss and the

latter can be kept high by protecting storage nodes/data centres from data breaches. In other words, the interviews indicated that it might be feasible to ask the user *specifically* to make a prioritization between the two security factors.

Accordingly, it was suggested in the UI proposal that the user should make a prioritization between *"Data Loss Prevention – High Data Availability"*, *"Data Protection – High Data Confidentiality"* as well as *"Cost Minimization"*. Although these prioritization options were accepted by most participants during the user walkthroughs, they appeared to lack some clarity. In fact, it was argued that the term "high availability" was rather ambiguous since it may refer to either *24/7 uptime* or *quick access time*. This represents two different types of service availability guarantees which, in turn, may imply dissimilar restrictions when it comes to the geographical distribution of data chunks. That is, the former may involve requirements in terms of the *minimum* distance between data centres (so that multiple storage nodes will not be hit by the same disaster), whereas the latter may place constraints on the *maximum* distance between employed data centres and *the customer* (so that data chunks do not have to travel too far to reach its destination).

It might not be possible to combine both of the geographical restrictions mentioned above, so one may argue that a distinction could be made between "uptime" (reliability) and "access time" (latency/performance) in the UI. However, configuration settings for high uptime or quick access time may otherwise be very similar, making it difficult for users to select one over the other. In other words, even though the interviews had suggested that a prioritization procedure may be feasible, the set of prioritization options presented in the UI proposal may be oversimplified. And at the same time, it might be difficult to expand the list of options without compromising the user's ability to place them in order by importance. Furthermore, the walkthroughs of the UI prototype indicated that the prioritization procedure may (wrongfully) give the impression that the Archistar solution will improve *either* confidentiality *or* availability of data, although both will be enhanced in comparison to a single cloud solution.

Rather than using a High/Medium/Low data classification scheme or making a prioritization of security aspects, the discussion above concludes that an alternative method should be utilized to indicate protection goals/need related to a backup/archiving project. For instance, a more data-driven approach could be utilized where the user is asked to describe what type of data the project will hold. (The user knows what kind of data their back/archiving project will involve, but it should perhaps not be assumed that they can translate the data characteristics into requirements or a prioritization that will result in appropriate configuration settings. This could be taken care of by the system itself.) In a data-driven approach, the user could e.g. describe the characteristics of data in the backup/archiving project by selecting options such as:

- **"Sensitive personal data"** – chunks should not be distributed to data centres outside of EU (so long as an equivalent level of protection is not provided). To reduce the risk of collusions, the recommended value on k should be high in relation to N and at least one chunk in a private cloud should be required for reconstructing the data.
- **"Business-critical data"** – the organization's survival depends on the information and high uptime is highly important. The value on k should be low in relation to N. A minimum distance between data centres may be applied.
- **"Time-critical data"** – the information may not be frequently used but when it is needed, one should be able to obtain it quickly. The value on k should be low in relation to N. A maximum distance between data centres and the customer may be applied.
- **"Non-sensitive data"** – if none of the other options applies. No extraordinary precautions are

necessary. A low value on N (for cost minimization) and k.

If the information is crucial in multiple various ways or scenarios, the user could have the option to select multiple options (without putting them in any particular order). To reduce the user's cognitive load, the system itself could subsequently be in charge of finding a balance between security aspects. (The options may inherently be in a certain pecking-order/hierarchy. For instance, if the user selects both "sensitive personal data" and "time-critical data", the system may automatically prioritize the former first for compliance with GDPR.)

In accordance with the discussion above, default values were provided in the UI proposal for the Secret Sharing parameters. Some walkthrough participant argued that they would like to stick with these defaults, rather than manually selecting or changing values for N and k. It was pointed out that frequent/expert users may desire the option to manually configure the Secret Sharing parameters but as a first-time user, default settings would be desired as it is difficult to recognize which values suit their goals/needs. This indicates that the notion behind the parameters was insufficiently explained in the UI at hand.

As previously mentioned, when respondents were asked to select suitable values for the Secret Sharing parameters during the interviews, they mainly reflected on the concept of dividing data into a certain number of chunks (N) but the notion of a threshold for reconstructing the information into its original state (k) appeared to be less easy to picture. During the walkthroughs of the UI prototype, one participant seemed to be confused by the term "threshold". Thus, the request for a value on k may have to be worded differently. Alternatively, the user could be able to create a configuration for a backup/archiving project without being forced to manually select a value for k. The option to do so could still be available - but in an "Expert view" with advanced configuration settings.

5.1.2 Determining Factors: Cost and Trustworthiness of CSPs

As earlier mentioned, some interview respondents needed to know how the cost will be affected by different values on the Secret Sharing parameters. A prerequisite for estimating the *cost* is that CSPs have been selected *before* the Secret Sharing parameters are configured. Although a higher value on N may result in greater storage overhead (see Section 2.3.1), the cost may still depend on which cloud storage providers and service offerings are utilized. For instance, 4 data chunks in public/external clouds may be less expensive than 3 chunks in private/internal clouds.

Even though the Archistar solution would (in theory) allow for less trustworthy CSPs to be utilized without putting the data privacy at risk, some interview respondents indicated that the extent to which CSPs can be trusted may still be crucial for the user when selecting values for N and k. During the interviews, it was indicated that Secret Sharing as a security measure may be perceived as less sufficient when public/external CSPs are employed. Nearly half of the respondents were concerned that CSPs would collude and reconstruct the data behind the user's back. As described by Petcu (2013), a "federated cloud" implies that CSPs collaborate by combining computing resources to jointly serve and meet the needs of customers (i.e., a service may be purchased from Provider A, but the storage space may reside in the cloud of Provider B), while the conditions around a "multi-cloud" is different. That is, it is typically the *user/customer* that is in charge of creating a multi-cloud infrastructure. The CSPs are less strongly linked and lack knowledge that they (and the other providers) are part of a multi-cloud solution. Thus, one could argue that the likelihood of collusions is rather low.

However, each CSP may have to enter a contract with the company behind Archistar in order for their services to be mediated through the proposed solution. There might be a risk that providers will be

aware of other cloud storage providers that also have a contract agreement with the Archistar company and thereby be able to deduce which clouds are part of the user's personalized multi-cloud. Thus, the UI should somehow provide the user with assurance or proof that CSPs will not (be able to) collaborate.

Out of the respondents who stated that they *would* consider utilizing providers that they normally would not trust in a single cloud solution, a significant proportion still acknowledged the importance of trust factors such as *high trust ratings*, *trust/privacy seals*, and *compliance with privacy laws*. (EU laws were mentioned most frequently, but a significant portion of the respondents also mentioned local/national laws of countries such as Germany.) The trustworthiness of CSPs may typically be evaluated *before* it is decided which providers to employ in a multi-cloud. Furthermore, as indicated during the walkthroughs of the UI prototype, (organizational) users may not be allowed to utilize CSPs that their institution does not already have a contract agreement with. In other words, apart from a contract with the company behind Archistar, the user/customers may also need some sort of agreement with the providers whose cloud storage services are mediated through the Archistar UI. It was indicated that an assessment of information about (1) service offerings, (2) ways of handling and reporting incidents, and (3) precautions against natural disasters may be performed before organizations enter contracts with CSPs.

Even if the organization has a contract with a particular CSP and receives a negotiated price on the service offering, the cost may still be relative to the amount of storage that is utilized. In order to decide whether a solution is affordable, one may have to relate to a specific budget. The UI proposal suggested that a *budget* could/should be specified before the user selects configuration settings that will affect the final cost of the backup/archiving project (i.e., size of data, total number of chunks, and selection of CSPs). This may be appropriate for the final version of the system as well, but the order in which different configuration settings are selected should be altered/changed to suit large organizations (further discussed in Section 5.4).

5.2 Geographical Distribution of Data Chunks

Administrators within an organization may not be able to choose freely whichever CSPs they like. However, the majority of participants partaking in the walkthroughs of the UI prototype seemed to believe that it could be feasible to allow administrators to select specific data centre locations (of contracted CSPs) to which data chunks can be distributed. However, there were mixed opinions in regards to the level of abstraction that data centre locations should be specified in.

During the interviews, respondents were asked to make geographical restrictions for the geographical distribution of data chunks in various different ways, i.e.: (1) Minimum distance between data centres in terms of *kilometres* and/or *administrative level*; (2) Trusted *climate zones*; as well as (3) Trusted *countries* for preventing data loss/availability issues, breaches of data confidentiality and collusion respectively. Only half of the respondents were able to specify a minimum distance in kilometres, while the majority (i.e., 14 out of 16) was able to select a minimum distance in terms of administrative level. This suggests that the user/customer may not have a precise constraint (in kilometres) to comply with and an abstract restriction may constitute a more feasible configuration option.

However, those who did select the same amount of kilometres did not necessarily select the same administrative level. The reasons might be several:

- Continents, countries, counties/provinces, and cities all differ greatly in land size. Moreover, the number of urban areas may also be significantly different from one region to another, meaning that the distance between cities is not identical in all areas of the world.
- Countries may not use the same form of administrative division. For instance, Argento et al. (2009) make a comparison between Italy and Sweden where it is described that Italy has four levels of government – i.e., 1 *central government*, 20 *regions*, 103 *provinces* and nearly 8100 *municipalities*. In Sweden, on the other hand, the public sector is organized into a *central* and a *local* level. On the local level, Sweden is further divided into 21 *county councils* and 290 *municipalities*. Thus, terminology such as "county" or "province" may not apply to all nations.

In other words, a distance in terms of administrative level is not a uniform measurement of geographical distance. One could argue that it might be more appropriate to select data centre locations in a more flexible manner (without applying a fixed minimum/maximum distance between them).

Climate zones were selected by 9 out of 16 interview respondents, all of which argued that data should be distributed to *cold* or *temperate* zones in order to be stored safely. However, a "safe" climate may not guarantee that data will not be subject to (natural) disasters. For instance, during the walkthroughs of the UI prototype, one participant described that the data centres of his organization were based in a city that had suffered from floods in the past – even if it was located in a cold/temperate climate zone. Thus, restricting the distribution of chunks to a specific climate may not be relevant either.

When asked to select specific countries/regions that would be most trusted for safeguarding information against data loss and unauthorized disclosure due to breaches or collusions, *laws/regulations* were frequently mentioned as a determining factor. EU/Europe was the most frequently selected followed by Canada and Australia/New Zealand. Some interview respondents explained that they did *not* select certain nations because they simply did not have adequate knowledge about the general conditions in these areas. In the UI proposal, the user was provided with details about available data centre locations (in an information box/container beside the map) so that they would be able to make an informed decision about its reliability/trustworthiness. However, during the walkthroughs, some participants argued that there were details that were unnecessary to know (e.g., the *Corruption Perception Index* and *Government Debt* of the country in which the data centre resides).

Although some details in the UI proposal may be excessive, there might be other more relevant information that is missing. That is, some walkthrough participants indicated that they were not fully aware of the laws that would apply to a solution like Archistar (see Section 2.3.4). Therefore, one could argue that the UI should communicate privacy laws that the (organizational) users need to follow. For instance, personal information protected with Secret Sharing will turn into "pseudonymous data", meaning that it would still classify as personal information pursuant to GDPR - and should be protected accordingly.

In the UI proposal, it is suggested that data centre locations should be selected from a map with layers/views illustrating the (1) likelihood of natural disasters (i.e., earthquakes, floods, wildfires), (2) the Internet backbone infrastructure, (3) political relationship between countries. The map also provided a visual demonstration of the distance between available data centre locations (in relation to the user/customer).

During the user walkthroughs, it was questioned how accurate the map layers/views for natural disasters could be. Even if employed data centres are based in cities with risks for e.g. floods, the facilities may still reside high above sea level and therefore be in a safe distance from danger. The map may be misleading as it does not communicate the altitude on which data centres are located. Other natural disasters such as fire may also be less easy to predict through statistics. It was suggested that it would be more appropriate to select data centre locations from a list (containing information about risk values).

5.3 Perceived Adequacy

While some interview respondents acknowledged that Secret Sharing could constitute better protection against *data loss issues*, Encryption appeared to be generally seen as a stronger security measure against *breaches of data confidentiality*. Secret Sharing did not seem to be perceived as an adequate replacement of Encryption when it came to storage of *sensitive* data. A combination of both security measures was regarded as the best solution, but some respondents argued that Encryption on its own would also suffice as protection for information of a sensitive nature. This indicated that the option to add encryption should be available for data that has high confidentiality and privacy requirements.

When conducting the walkthroughs of the UI prototype, all participants argued that the option to add a layer of encryption should be available for *all* type of data, regardless of sensitivity. During both the interviews and the walkthroughs, comments were received indicating that the level of security would be difficult to prove in the Archistar solution and that Encryption may serve as an extra assurance that adversaries will not have access to the secret information.

As earlier mentioned, all people were not convinced that unauthorized individuals would be unable to reconstruct the information. The scepticism against the proposed solution may partly stem from mistrust towards the cloud as a storage medium. For instance, during the interviews, Secret Sharing appeared to be perceived as less adequate in the context of public/external clouds, in comparison to storage facilities that are on-premise. Furthermore, a significant proportion of the respondents were concerned about the risk of collusions between CSPs. This stresses the importance of providing the user with concrete evidence that data will be adequately protected.

5.4 User Groups with Different Skills and Needs

In the analysis of previous writings about Archistar and its use case (see Appendix A), an assumption was made that the solution would be utilized by (users whose competence corresponds to) system administrators. However, during the walkthroughs of the UI prototype, there were indications that some parts of the configuration process would require the involvement of other types of stakeholders. It was pointed out that *assessments of CSPs* (and risks/vulnerabilities) are not necessarily performed by an administrator. Furthermore, establishing *contracts* with service providers and making *budgeting decisions* is usually not the administrators' responsibility either. A "project leader" may have admin rights for a backup/archiving project and therefore be able to specify a budget within that particular project. However, the Archistar configuration may also need to comply with cost constraints pertaining to the organization/company as a whole. (The project budget might be flexible, whereas the organizational cost restrictions may perhaps serve as an upper limit for the capital expenditure.)

Unless the Archistar customer constitutes a private individual or a small-scale company (where employees may have multiple roles), it should not be assumed that a single individual will be able to make all decisions related to the configuration process. Consequently, one could argue that the user interface should be divided into two parts – i.e., one part where a manager creates a set of "global settings" that applies to all projects within an organization/company, and another part where

configuration settings associated to a specific project are made by an administrator.

In the Manager UI, an organizational cost restriction is specified. Furthermore, information about service offerings and guarantees of various CSPs could be presented in the UI. The manager reviews the information and enters a contract agreement with the providers whose services seems appropriate for the user to utilize. The Admin UI could, in turn, be divided into the following steps:

(1) The size and characteristics of the backup/archiving data is described (e.g., “100 GB” storage space is needed and the project includes "sensitive personal data").
(2) Furthermore, data centre locations to which data chunks should be distributed is selected (either from a list or a map). The default is that one data centre should be chosen for each of the contracted CSPs, and that one chunk should be transferred to each data centre location (i.e., the number of providers that the organisation has a contract agreement with represents the recommended value on N). (However, if cost minimization is important, the UI might advise against using all available CSPs. On the contrary, if data loss prevention is critical, the UI could inform the user that a higher number of locations may reduce the risk of data loss issues as he/she will be less reliant on individual data centres. In the latter scenario, the user may desire to use multiple data centres of a particular provider, increasing the value on N. If the project includes “sensitive personal data”, the user should not be able to select data centre locations whose level of security does not correspond to the standard of EU). When a data centre is chosen, the corresponding service will be added to the multi-cloud solution, the value on N is increased by 1 and the estimated *cost* is updated in the UI.
(3) Subsequently, the UI selects a suitable value for k, based on what type of data the project holds. The UI could provide the user with information that indicates the extent to which data availability and confidentiality will be protected/maintained (e.g., *"Availability Rate: 99.9%"*, *"Protected against 2 simultaneous Data Breaches"*). The user can enter an "Expert view" to change the selected threshold.
(4) Lastly, the user can apply a layer of encryption to each data chunks, before it is distributed to the clouds.

6. Conclusion

RQ1. What are suitable configuration options and guidelines for organizational or private users with different security requirements?

Scenarios of private and organizational use may involve different preconditions for creating a configuration of a multi-cloud storage solution based on Secret Sharing. A private user may single-handedly decide which configuration settings to use for a backup/archiving project, whereas the responsibility of making such decisions in an organization might be allocated between different roles/stakeholders. In other words, the UI proposal presented in this thesis might be suitable for private users. But in order for the proposed solution to be utilized in organizations, its UI should be divided into two *parts* targeting different types of stakeholders – i.e., one part for *managers* (responsible for budgeting decisions and establishing contract agreements with service providers) and another for *administrators* (in charge of handling specific backup/archiving projects).

The option to add a layer of encryption may be desired for all types of backup/archiving projects for trust reasons or for legal compliance. In contrast to the UI proposals, it may be more appropriate to select CSPs and data centre locations *before* the Secret Sharing parameters are configured. Thereby, the final cost could be estimated as the administrator selects a value for N (i.e., the total number of chunks). When selecting CSPs to employ in the multi-cloud solution, organizational users may have more limited options to choose from since they might be restricted to contracted providers only. The decision of entering a contract agreement with a CSP may be established by a manager, while the choice between various data centre locations of that particular provider could subsequently be made by an administrator. However, the level of abstraction in which data centre locations are to be selected is yet to be determined. Furthermore, the proposed map view could be replaced or supplemented with a simple list of data centre locations.

The UI could inform the user that a higher number of locations (i.e., a higher value on N) may reduce the risk of data loss issues (as he/she will be less reliant on individual data centres), but that it typically comes with a higher cost. While the concept of dividing data into a certain number of chunks appeared to be easy to picture, users might have difficulties to comprehend the notion of a threshold for data reconstruction (k). The UI could provide a suitable default value based on the type of data that the backup/archiving project will contain, allowing the user to create a configuration without manually selecting a value on k. The option to do so could, however, be available in an "Expert view" for experienced users.

Asking the user to find a suitable security trade-off (with a High/Medium/Low data classification scheme or through a prioritization procedure of security aspects) may not be feasible. Instead, a data-driven approach could be utilized where the user simply describes the type of data that their project will include and the system itself finds an appropriate balance by recommending a value on k and communicating the implications on both availability *and* confidentiality.

The legal aspects related to the proposed solution may not be clear to prospective users. If the original data is classified as *sensitive personal information* according to GDPR, then the same classification will be given to data chunks generated through the Secret Sharing mechanism even if they would not (individually) reveal any details about the information. Thus, the user should be provided with guidelines on how to sufficiently protect personal data (i.e., data should only be distributed to locations inside of EU or locations that provide an equivalent level of security, and k chunks should preferably not be distributed to public clouds).

RQ2. What are relevant trust factors, unique advantages, and risks of a multi-cloud storage solution based on Secret Sharing that should/could be communicated to the users?

Although the Secret Sharing mechanism may allow for less trustworthy CSPs to be employed without compromising the confidentiality of data, the interviews suggested that trust in providers may still be critical for prospective users. *High trust ratings*, *trust/privacy seals*, and *compliance with privacy laws* may be essential trust factors. Moreover, information about service offerings, ways of handling and reporting incidents, and precautions against natural disasters may also be crucial to know before contracts with CSPs are entered.

The UI proposal at hand may give the impression that *either* confidentiality *or* availability of data will be improved. Even if there may be a trade-off between these two factors, the UI should clarify that both aspects will still be enhanced in comparison to a single cloud solution.

The interviews indicated that prospective users may be concerned about the risk of collusions between CSPs. Thus, the UI should somehow provide the user with assurance or proof that they will not (be able to) collaborate and reconstruct the user's data behind his/her back.

6.1 Limitations of Study

The UI(s) presented in this thesis should be seen as preliminary proposals rather than a final product of a multi-cloud system based on Secret Sharing. Due to the study's exploratory nature and small sample size, the author cannot claim generalizability. As described by Bryman (2012), in exploratory research it is typically difficult to determine whether the selected sample is representative over a larger population. Furthermore, although the qualitative data collected in the study was analysed thematically (i.e., emphasis on comments/themes mentioned by multiple respondents), the correlation between user's acceptance of the Archistar solution and trust factors is not confirmed through statistics. Thus, the proposed configuration options and UI solutions should be seen as a suggestion rather than a definite rule.

In future research, quantitative data can be collected by conducting a survey with more specific questions. Moreover, while the focal point of this study was the creation of configurations, future research could scrutinize other features of the proposed solution. For instance: How should the splitting/fragmentation and reconstruction of data be visualized in a UI to accomplish sufficient user trust and transparency? How can it be assured in UI that CSPs are unaware of each other and will not collaborate?

References

Aazam, M., & Huh, E. N. (2014). Inter-cloud architecture and media cloud storage design considerations. *the proceedings of 7th IEEE CLOUD, Anchorage, Alaska, USA, 27*.

Alaqra, A., Fischer-Hübner, S., Pettersson, J.S. (Eds. 2017) *PRISMACLOUD Deliverable – D3.2 HCI Guidelines.*

Alter, S. (2006). *The work system method: connecting people, processes, and IT for business results*. Work System Method.

Argento, D., Grossi, G., Tagesson, T., & Collin, S. O. (2009). The'externalisation'of local public service delivery: experience in Italy and Sweden. *International journal of public policy*, *5*(1), 41-56.

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., ... & Zaharia, M. (2010). A view of cloud computing. *Communications of the ACM*, *53*(4), 50-58.

Arockiam, L., & Monikandan, S. (2014, January). Efficient cloud storage confidentiality to ensure data security. In *Computer Communication and Informatics (ICCCI), 2014 International Conference on* (pp. 1-5). IEEE.

Babbie, E. R. (2012). *The practice of social research, 13th Edition*. Nelson Education.

Balasaraswathi, V. R., & Manikandan, S. (2014, May). Enhanced security for multi-cloud storage using cryptographic data splitting with dynamic approach. In *Advanced Communication Control and Computing Technologies (ICACCCT), 2014 International Conference on* (pp. 1190-1194). IEEE.

Ball, C. (2009). What Is Transparency? *Public Integrity, 11*(4), 293-308,

Bauer, E., & Adams, R. (2012). *Reliability and availability of cloud computing.* John Wiley & Sons.

Bellare, M., & Rogaway, P. (2016). *U.S. Patent No. 9,407,431*. Washington, DC: U.S. Patent and Trademark Office.

Benyon, D. (2014). Designing Interactive Systems: A comprehensive guide to HCI, UX and interaction design, 3/E.

Bessani, A., Correia, M., Quaresma, B., André, F., & Sousa, P. (2013). DepSky: dependable and secure storage in a cloud-of-clouds. *ACM Transactions on Storage (TOS)*, *9*(4), 12.

Bhowmik, S. (2017). *Cloud Computing.* Cambridge University Press.

Blakley, G. R. (1979). Safeguarding cryptographic keys. In *Proc. AFIPS 1979 National Computer Conference* (pp. 313-317).

Blasius, J. (2012). Comparing ranking techniques in web surveys. *Field Methods*, *24*(4), 382-398.

Bostrom, A., Anselin, L., & Farris, J. (2008). Visualizing seismic risk and uncertainty: A review of related research. *Annals of the New York Academy of Sciences*, *1128*(1), 29-40.

Brinck, T., Gergle, D., & Wood, S. D. (2002). *Designing Web sites that work: Usability for the Web*. Morgan Kaufmann Publishers.

Brocca, M.D., Facchinetti, S., Länger, T., Lorünser, T., & Happe, A. (n.d.). *Cloud storage solution for hybrid cloud in eGovernment context: A distributed cloud storage solution for e-Government can help optimize resources usage, data security and privacy, service reliability.*

Brocca, M.D., Facchinetti, S., Ferrari, G., Galbis, A.Z., Cáceres, S., Perez Perez, M. M., Gallico, D., & Biancani, M. (2016). *PRISMACLOUD Deliverable – D2.3 Use Case Specification v.2.1.*

Brodies LLP (n.d.). *What does the GDPR mean for…Public Authorities?* [Online]. Available at: https://brodies.com/sites/default/files/brodies_handy_guide_gdpr_-_public_authorities.pdf [2018-12-07].

Bryman, A. (2012). *Social research methods*. Oxford university press.

Carey, M., Lanyi, M. M., Longo, D., Radzinski, E., Rouiller, S., & Wilde, E. (2014). *Developing quality technical information: A handbook for writers and editors*. Pearson Education.

Chaffey, D. (2015). *Digital business and e-commerce management*. Pearson Education Limited.

Chandrasekaran, K. (2014). *Essentials of cloud computing.* CRC Press.

Chen, H. C., Hu, Y., Lee, P. P., & Tang, Y. (2014). NCCloud: A network-coding-based storage system in a cloud-of-clouds. *IEEE Transactions on computers*, *63*(1), 31-44.

Cloud Security Alliance (2017). *The Treacherous 12: Top Threats to Cloud Computing – Industry Insights.* [Online]. Available at: https://downloads.cloudsecurityalliance.org/assets/research/top-threats/treacherous-12-top-threats.pdf [2018-01-20].

Colborne, G. (2011). *Simple and usable web, mobile, and interaction design*. New Riders.

Coleman, C. V. (2017). *Visual Experiences: A Concise Guide to Digital Interface Design*. Chapman and Hall/CRC.

Conrad, F., Tourangeau, R., Couper, M., & Zhang, C. (2017, April). Reducing speeding in web surveys by providing immediate feedback. In *Survey Research Methods* (Vol. 11, No. 1, pp. 45-61).

Cooper, A., Reimann, R., Cronin, D., & Noessel, C. (2014). *About face: the essentials of interaction design*. John Wiley & Sons.

Dransch, D., Rotzoll, H., & Poser, K. (2010). The contribution of maps to the challenges of risk communication to the public. *International Journal of Digital Earth*, *3*(3), 292-311.

Erl, T., Puttini, R., & Mahmood, Z. (2013). *Cloud computing: concepts, technology & architecture*. Pearson Education.

Fabian, B., Ermakova, T., & Junghanns, P. (2015). Collaborative and secure sharing of healthcare data in multi-clouds. *Information Systems*, *48*, 132-150.

Fadeyev, D. (2009). User interface design in modern web applications. *The Smashing Book*, 15.

FIPS Publication 199 (2004). Standards for Security Categorization of Federal Information and Information Systems.

Firdhous, M., Ghazali, O., & Hassan, S. (2012). Trust management in cloud computing: a critical review. *arXiv preprint arXiv:1211.3979*.

Ford, J.L., Jr. (2015). *Programming for the absolute beginner*. Nelson Education.

Freedom Information Act 2000, c. 36. Available at: https://www.legislation.gov.uk/ukpga/2000/36/pdfs/ukpga_20000036_en.pdf [2018-06-21].

Fugini, M. G., Maggiolini, P., & Valles, R. S. (2014). *E-government and employment services: a case study in effectiveness*. Springer.

Galitz, W. O. (2007). *The essential guide to user interface design: an introduction to GUI design principles and techniques*. John Wiley & Sons.

Gao, Y. (Ed.). (2005). *Web systems design and online consumer behavior*. IGI Global.

Garrett, J. J. (2010). *Elements of user experience, the: user-centered design for the web and beyond*. Pearson Education.

General Data Protection Regulation (GDPR) 2016/679. Available at: https://gdpr-info.eu/ [2018-11-20].

Gharehchopogh, F. S., & Hashemi, S. (2012). Security challenges in cloud computing with more emphasis on trust and privacy. *International Journal of Scientific & Technology Research*, *1*(6), 49-54.

Goyal, S. (2014). Public vs private vs hybrid vs community-cloud computing: a critical review. *International Journal of Computer Network and Information Security*, *6*(3), 20.

Grauer, Y. (2018). *British and Canadian Governments Accidentally Exposed Passwords and Security Plans to the Entire Internet.* [Online]. Available at: https://theintercept.com/2018/08/16/trello-board-uk-canada/ [2018-12-07].

Grimmelikhuijsen, S. G. (2012). *Transparency and trust. An experimental study of online disclosure and trust in government* (Doctoral dissertation, University Utrecht).

Gu, Y., Wang, D., & Liu, C. (2014). DR-Cloud: Multi-cloud based disaster recovery service. *Tsinghua Science and Technology*, *19*(1), 13-23.

Hagen, R., & Golombisky, K. (2017). *White space is not your enemy: A beginner's guide to communicating visually through graphic, web & multimedia design*. 3rd Edition. Focal Press.

Halpert, B. (2011). *Auditing cloud computing: a security and privacy guide* (Vol. 21). John Wiley & Sons.

Happe, A., Wohner, F., & Lorünser, T. (2017). The archistar secret-sharing backup proxy. In *Proceedings of the 12th International Conference on Availability, Reliability and Security* (p. 88). ACM.

Hodgson, G. (2015). Breaking Encryption and Gathering Data: International Law Applications. *J. Tech. L. & Pol'y*, *20*, 39.

Hoekman Jr, R. (2010). *Designing the obvious: A common sense approach to Web & Mobile Application Design*. Pearson Education.

Human Rights Act 1998, c. 42. Available at: https://www.legislation.gov.uk/ukpga/1998/42/pdfs/ukpga_19980042_en.pdf [2018-06-21].

Hurwitz, J. S., Bloor, R., Kaufman, M., & Halper, F. (2010). *Cloud computing for dummies*. John Wiley & Sons.

INSPIRE Directive 2007/2/EC. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32007L0002&from=EN [2018-06-21].

Ioannidis, C., Pym, D., & Williams, J. (2012). Information security trade-offs and optimal patching policies. *European Journal of Operational Research*, *216*(2), 434-444.

Jarrett, C., & Gaffney, G. (2009). *Forms that work: Designing Web forms for usability*. Morgan Kaufmann.

Johnson, J. (2008). *GUI bloopers 2.0: common user interface design don'ts and dos*. Elsevier.

Johnson, J. (2013). *Designing with the mind in mind: simple guide to understanding user interface design guidelines*. Elsevier.

Kalof, L., Dan, A., & Dietz, T. (2008). *Essentials of social research*. McGraw-Hill Education (UK).

Kamara, S., & Lauter, K. (2010, January). Cryptographic cloud storage. In International Conference on Financial Cryptography and Data Security (pp. 136-149). Springer Berlin Heidelberg.

Karth, S. T. (2011). *A Comparison of a Traditional Ranking-Task and a Drag-and-Drop Ranking Task* (Doctoral dissertation, University of Dayton).

Khan, K. M., & Malluhi, Q. (2010). Establishing trust in cloud computing. *IT professional*, (5), 20-27.

Krug, S. (2014). Don't make me think revisited: A common sense approach to web and mobile usability . Berkeley.

Krutz, R. L., & Vines, R. D. (2010). *Cloud security: A comprehensive guide to secure cloud computing*. Wiley Publishing.

Kunz, T. (2015). *Rating scales in Web surveys. A test of new drag-and-drop rating procedures* (Doctoral dissertation, Technische Universität).

Lal, R. (2013). *Digital Design Essentials: 100 ways to design better desktop, web, and mobile.*

Li, X., Hess, T. J., & Valacich, J. S. (2008). Why do we trust new technology? A study of initial trust formation with organizational information systems. *The Journal of Strategic Information Systems*, *17*(1), 39-71.

Li, C., Liu, Y., Xie, T., & Chen, M. Z. (2013). Breaking a novel image encryption scheme based on improved hyperchaotic sequences. *Nonlinear Dynamics*, *73*(3), 2083-2089.

Lidwell, W., Holden, K., & Butler, J. (2010). *Universal principles of design, revised and updated: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design*. Rockport Pub.

Linthicum, D. S. (2009). *Cloud computing and SOA convergence in your enterprise: a step-by-step guide*. Pearson Education.

Lior, L. N. (2013). *Writing for interaction: crafting the information experience for web and software apps*. Newnes.

Liu, F., Tong, J., Mao, J., Bohn, R., Messina, J., Badger, L., & Leaf, D. (2011). NIST cloud computing reference architecture. *NIST special publication*, *500*(2011), 292.

Lopuck, L. (2012). *Web design for dummies*. John Wiley & Sons.

Lorünser, T., Slamanig, D., Länger, T., & Pöhls, H. C. (2016, August). PRISMACLOUD tools: a cryptographic toolbox for increasing security in cloud services. In *Availability, Reliability and Security (ARES), 2016 11th International Conference on* (pp. 733-741). IEEE.

Lorünser, T., Happe, A., Krenn, S., Hudic, A., Riemer, K., Länger, T., Zambrano, A., Moffie, M., Palomares, A., Sfyrakis, I., Martinez, A., & Pöhls, H.C. (2017*). PRISMACLOUD Deliverable D7.6 Guidelines and Architecture for Secure Service Composition.*

Lowdermilk, T. (2013). *User-centered design: a developer's guide to building user-friendly applications*. " O'Reilly Media, Inc.".

Marinescu, D. C. (2017). *Cloud computing: theory and practice*. Morgan Kaufmann.

Martin, K. M. (2008). Challenging the adversary model in secret sharing schemes. *Coding and Cryptography II, Proceedings of the Royal Flemish Academy of Belgium for Science and the Arts*, 45-63.

Mather, T., Kumaraswamy, S., & Latif, S. (2009). *Cloud security and privacy: an enterprise perspective on risks and compliance*. " O'Reilly Media, Inc.".

McKay, E. N. (2013). *UI is communication: How to design intuitive, user centered interfaces by focusing on effective communication*. Newnes.

Mell, P., & Grance, T. (2011). The NIST definition of cloud computing.

Mishra, U. (2009). *Improving Graphical User Interface using TRIZ*. Umakanta Mishra.

Moran, J. (2015). *File Management Made Simple, Windows Edition*. Apress.

Moran, K. (2017). *The Aesthetic-Usability Effect.* [Online]. Available at: https://www.nngroup.com/articles/aesthetic-usability-effect/ [2017-11-12].

Mourby, M., Mackey, E., Elliot, M., Gowans, H., Wallace, S. E., Bell, J., ... & Kaye, J. (2018). Are 'pseudonymised'data always personal data? Implications of the GDPR for administrative data research in the UK. *Computer Law & Security Review*, *34*(2), 222-233.

Muehlenhaus, I. (2013). *Web cartography: map design for interactive and mobile devices*. CRC Press.

Neumann, S., Kulyk, O., & Volkamer, M. (2014, September). A usable android application implementing distributed cryptography for election authorities. In *Availability, Reliability and Security (ARES), 2014 Ninth International Conference on* (pp. 207-216). IEEE

Neuman, W. L. (2013). *Social research methods: Qualitative and quantitative approaches*. Pearson education.

Nielsen, J. (1995). *10 usability heuristics for user interface design.* [Online]. Availability at: https://www.nngroup.com/articles/ten-usability-heuristics/ [2017-11-12].

Nielsen, J. (2006). *F-Shaped Pattern For Reading Web Content (original study).* [Online]. Available at: https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content-discovered/ [2017-11-12].

Nissenbaum, H. (1999). Can trust be secured online? A theoretical perspective.

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Constellation.

Noyes, J., & Baber, C. (1999). *User-centred design of systems*. Springer Science & Business Media.

Oshri, I., Kotlarsky, J., & Willcocks, L. P. (2015). *The Handbook of Global Outsourcing and Offshoring 3rd Edition*. Springer.

O'Sullivan, D. (2017). *Black Box, Red Disk: How Top Secret NSA and Army Data Leaked Online.* [Online]. Available at: https://www.upguard.com/breaches/cloud-leak-inscom [2018-12-07].

Ren, K., Wang, C., & Wang, Q. (2012). Security challenges for the public cloud. *IEEE Internet Computing*, *16*(1), 69-73.

Rittinghouse, J. W., & Ransome, J. F. (2010). *Cloud computing: implementation, management, and security*. CRC press.

Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: how to plan, design and conduct effective tests*. John Wiley & Sons.

Patel, R., & Davidson, B. (2011). *Forskningsmetodikens grunder. Att planera, genomföra och rapportera en undersökning*. Studentlitteratur.

Pearson, S. (2013). Privacy, security and trust in cloud computing. In *Privacy and Security for Cloud Computing* (pp. 3-42). Springer London.

Petcu, D. (2013, April). Multi-Cloud: expectations and current approaches. In *Proceedings of the 2013 international workshop on Multi-cloud applications and federated clouds* (pp. 1-6). ACM.

Preece, J., Rogers, Y., & Sharp, H. (2015). *Interaction design: beyond human-computer interaction*. John Wiley & Sons.

Quick, D., Martini, B., & Choo, R. (2013). *Cloud storage forensics*. Syngress.

Saffer, D. (2010). *Designing for interaction: creating innovative applications and devices*. New Riders.

Salman, T. (2015). On Securing Multi-Clouds: Survey on Advances and Current Challenges. *Semantic Scholar*, 1-16.

Schulz, G. (2011). *Cloud and virtual data storage networking*. Auerbach Publications.

SFS 2003:460. Swedish Ethical Review Act. Stockholm: Utbildningsdepartementet.

Shamir, A. (1979). How to share a secret. *Communications of the ACM*, *22*(11), 612-613.

Singhal, M., Chandrasekhar, S., Ge, T., Sandhu, R., Krishnan, R., Ahn, G. J., & Bertino, E. (2013). Collaboration in multicloud computing environments: Framework and security issues. *Computer*, *46*(2), 76-84.

Sitaram, D., & Manjunath, G. (2011). Chapter 7 – Designing Cloud Security. In *Moving to the Cloud: Developing Apps in the New World of Cloud Computing*. Elsevier.

Sloan, R. H., & Warner, R. (2013). *Unauthorized access: The crisis in online privacy and security*. CRC press.

Social Research Association (2003). *Ethical quidelines.* [Online]. Available at: http://the-sra.org.uk/wp-content/uploads/ethics03.pdf [2017-03-27].

Sosinsky, B. (2010). *Cloud computing bible* (Vol. 762). John Wiley & Sons.

Surbhi, S. (2015). *Difference Between Public Sector and Private Sector.* [Online]. Available at: https://keydifferences.com/difference-between-public-sector-and-private-sector.html [2018-12-19].

Sunyaev, A., & Schneider, S. (2013). Cloud services certification. *Communications of the ACM*, *56*(2), 33-36.

Swanson, M. (2001). *Security self-assessment guide for information technology systems* (No. NIST-SP-800-26). BOOZ-ALLEN AND HAMILTON INC MCLEAN VA.

Swedish Research Council (2002). *Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning.* [Online]. Available at: http://www.codex.vr.se/texts/HSFR.pdf [2017-03-27].

Zambrano, A., Cáceres, S., Martínez, A.I., Pérez, M., Palomares, A., Decandia, M., & IRT Team Projects (2017). *PRISMACLOUD Deliverable D8.1 – Specification of test-bed configurations for validation phase.*

Ulutas, M., Ulutas, G., & Nabiyev, V. V. (2011). Medical image security and EPR hiding using Shamir's secret sharing scheme. *Journal of Systems and Software*, *84*(3), 341-353.

United Nations (1948). *Universal Declaration of Human Rights.*

Uusitalo, I., Karppinen, K., Juhola, A., & Savola, R. (2010, November). Trust and cloud services-an interview study. In *Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on* (pp. 712-720). IEEE.

Velte, A. T., Velte, T. J., Elsenpeter, R. C., & Elsenpeter, R. C. (2010). *Cloud computing: a practical approach* (pp. 1-55). New York: McGraw-Hill.

Vukolić, M. (2010). The Byzantine empire in the intercloud. *ACM Sigact News*, *41*(3), 105-111.

Vu, K.P.L., & Proctor, R.W. (2011). *Handbook of human factors in Web design*. CRC Press.

Wang, C., Chow, S. S., Wang, Q., Ren, K., & Lou, W. (2013). Privacy-preserving public auditing for secure cloud storage. *IEEE Transactions on computers*, *62*(2), 362-375.

Warren, S. D., & Brandeis, L. D. (1890). The right to privacy. *Harvard law review*, 193-220.

Weinschenk, S. (2011). *100 things every designer needs to know about people*. Pearson Education.

Westin, A. F. (2003). Social and political dimensions of privacy. *Journal of social issues*, *59*(2), 431-453.

Wieczorek, A. M., Klyszejko, Z., Sarzynska, J., Szóstek, A., Chmiel, K., Soluch, T., & Brzezicka, A. (2014). Mode of text presentation and its influence on reading efficiency: scrolling versus pagination. *Studia Psychologica*, *56*(4), 309.

Wilson, C. (2009). *User experience re-mastered: your guide to getting the right design*. Morgan Kaufmann.

Wilson, M. L. (2011). Search user interface design. *Synthesis lectures on information concepts, retrieval, and services*, *3*(3), 1-143.

Wilson, C. (2013). *User interface inspection methods: a user-centered design method*. Newnes.

Wong, K. S., & Kim, M. H. (2016). An enhanced user authentication solution for mobile payment systems using wearables. *Security and Communication Networks*, *9*(17), 4639-4649.

Xu, T., & Zhou, Y. (2015). Systems approaches to tackling configuration errors: A survey. *ACM Computing Surveys (CSUR)*, *47*(4), 70.

Yin, Z., Ma, X., Zheng, J., Zhou, Y., Bairavasundaram, L. N., & Pasupathy, S. (2011, October). An empirical study on configuration errors in commercial and open source systems. In *Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles* (pp. 159-172). ACM.

Yuk, M., & Diamond, S. (2014). *Data visualization for dummies*. John Wiley & Sons.

Özpolat, K., & Jank, W. (2015). Getting the most out of third party trust seals: An empirical analysis. *Decision Support Systems*, *73*, 47-56.

Appendix A. Analysis of the e-Government Use Case used in the Study.

1. e-Government Use Case

The e-Government use case defined in PRISMACLOUD is used in the present study. It is based on the following real-life scenario: In the Lombardy region in northern Italy, public authorities currently have to self-manage data storage in their own data centres. The IT infrastructure within these data centres lacks the flexibility to adapt to public authorities' needs. Thus, the aforementioned infrastructure is commonly over-dimensioned, leading to cost issues, or under-dimensioned, resulting in performance or service level issues. In accordance with regional and national regulations, the aim is to reduce costs by rationalizing data centres of public authorities and subsequently moving data and services to a cloud-based environment. Lombardia Informatica SpA (LISPA), a regional ICT company, is tasked with developing a secure distributed storage solution called "Archistar" to enable this data migration to the cloud (Brocca et al. 2016; n.d.).

1.1 Purposed Use of Service

Brocca et al. (2016; n.d.) state that the proposed solution should provide *backups* of data in a secured, distributed manner. Moreover, according to Brocca et al. (n.d.:6), the focal point of the overall use case is "the data backup need[s] of public bodies and authorities". In Lorünser et al. (2016:735-736), on the other hand, it is described that a "secure object storage tool" is developed, which will be customized for the e-Government use case to serve as a secure *archiving* service. Happe et al. (2017) also explain that Archistar will be used for archiving data.

"Data backups" and "archives" may sometimes be confused with each other, but the terms should not be used interchangeably as they refer to rather different operations (Sosinsky 2010:327). Archives typically contain data that is rarely or no longer in use, while backups represent data that is still operational (Hurwitz et al. 2010:275; Sosinsky 2010). Archives imply that a set of data is *moved* from its original location to cheaper, secondary storage, whereas backups entail that a data set is *copied* and stored in multiple locations (Sosinsky 2010). Archives can e.g. be established for legal compliance or historical record keeping (Hurwitz et al. 2010; Sosinsky 2010). The aim of archives is to *preserve* old data for potential future use and it is commonly retained for a longer period of time. On the other hand, backups serve as a means for *protecting* current information by enabling data recovery in case of an incident (Schulz 2011). Archives cannot be utilized to restore a current set of information (Sosinsky 2010).

Interpretation 1: In the context of this thesis, ARCHISTAR will be seen as a solution for both backup and archiving of data.

1.2 Active Stakeholders

Archistar is being developed by LISPA in collaboration with PRISMACLOUD. Entities with an interest in PRISMACLOUD-enabled solutions can generally be divided into two main categories – i.e., "Actors/Active Stakeholders" and "Inactive Stakeholders". The former interact directly with the solution and may consist of the following (Brocca et al. 2016):

- **Cloud Provider** – a stakeholder that offers an Infrastructure as a Service (IaaS). Storage and computing resources are provided via virtualization to entities that want to offload their own IT infrastructure (Cf. Section 2.1.4.1).
- **Service Provider** – a stakeholder that hosts the PRISMACLOUD-enabled solution (either on its own IT infrastructure or by leveraging on a Cloud Provider's IaaS) and offers it to other entities.

- **Customer/end-user** – a stakeholder that ultimately uses the PRISMACLOUD-enabled solution.

As indicated above, the PRISMACLOUD-enabled solution represents Archistar in the context of the e-Government use case. This service offers means for increasing the data security in the cloud by dividing the data into *chunks* and distributing them to separate servers/storage nodes. From a customer/end-user's perspective, the cloud resource relevant to the e-Government use case is *storage*. Thus, a cloud provider is hereon referred to as a **Cloud Storage Provider (CSP)**.

While Archistar is provided by LISPA, storage resources on which data chunks can be stored are offered by various different CSPs. In Zambrano et al. (2017:39), CSPs that could be employed in the e-Government use case are exemplified. That is, a *public* provider may be Amazon AWS[17], whereas *private* providers mentioned by Zambrano et al. (2017) are LISPA and Interoute SpA (IRT)[18]. In other words, LISPA may serve as a provider of both Archistar and of a subset of cloud storage utilized *within* the Archistar solution.

Interpretation 2: Both private and public CSPs can be employed in the Archistar solution.

1.2.1 Who is the Customer?

The term "e-Government" could be used to refer to technologies for electronic transactions of information, services or financial assets, which are applied to government- and public services (Chaffey 2015). Through the use of the Internet and the World Wide Web, government information or services are electronically exchanged between different *governmental actors*, or between a government and *citizens/businesses* (Fugini et al. 2014). In the PRISMACLOUD e-Government use case, the customer of LISPA (and, therefore, also of Archistar) is referred to as "public authorities" (Brocca et al. 2016).

According to the INSPIRE Directive 2007/2/EC and the UK Human Rights Act (1998), a "public authority" may refer to a (natural or legal) person who serves a public administrative function. The INSPIRE Directive 2007/2/EC also describes that it may represent a government or public administration at a local, regional or national level. On the other hand, the definition by the Human Rights Act (1998) suggests that public authorities include courts/tribunals – but *not* the Parliament or individuals exercising a function in connection with parliamentary proceedings. In Chapter 36 of the UK Freedom of Information Act (2000), public authorities are broadly defined to comprise: Government departments, local governments (such as county councils), national health services, the police, and educational institutions (such as schools, colleges and universities).

In other words, public authorities typically *also* provide services to other entities. Accordingly, it is described in Brocca et al. (2016) that the public authorities may in some cases act as a service provider and mediate PRISMACLOUD-enabled solutions to citizens (or businesses).

Interpretation 3: The customer of Archistar may be public authorities – or citizens/private businesses. In the context of this thesis, Archistar will not be seen as a solution limited to only organizations and private individuals in the Lombardy Region.

[17] https://aws.amazon.com
[18] https://www.interoute.com/office/italy

1.2.2 Who is the End-user?

If Archistar is mediated to citizens, private individuals would represent both the customers *and* the end-users. However, if the Archistar customer constitutes *organizations/companies*, the scenario would be less certain. Customer organizations may contain multiple types of stakeholders, meaning that the potential end-user of the PRISMACLOUD-enabled solution may be various different entities.

In Lorünser et al. (2016), it is described that the archiving/backup service (i.e., Archistar) is fundamentally an Infrastructure as a Service (IaaS). Cloud service models (discussed in Section 2.1.2) differ in terms of required level of knowledge and skillsets (Bhowmik 2017) as well as target audience: A Software as a Service (SaaS) is aimed at users who are free from all maintenance responsibilities; a Platform as a Service (PaaS) is intended for software developers; and an Infrastructure as a Service (IaaS) typically targets IT architects (Chandrasekaran 2014).

In a single cloud solution, the underlying infrastructure is controlled by the CSP (Chandrasekaran 2014). In the case of Archistar, the user will select and employ *several* CSPs, meaning that they will be part of forming the infrastructure of a "multi-cloud" (described in Section 2.3.2). Moreover, while the CSPs are responsible for the security within their data centres, the Archistar users will themselves apply security measures upon the data *before* it is transferred to a cloud-based environment. That is, the users will create *configurations* where it is established how many chunks a particular data set should be divided into and how many of these chunks should be required in order to reconstruct the information into a legible state.

System configurations are typically more complex than applications operated by users with low or moderate computer skills. Thus, configuration tasks are generally performed by "system administrators". Such a stakeholder has more technical expertise than ordinary computer users, but still not the same level of understanding as the developer behind the system. That is, they are less capable than the developers to debug an application if issues are encountered, since they did not write the code behind it (Xu & Zhou 2015). Nevertheless, this suggests that Archistar users need to be technically knowledgeable in order to select a configuration that will provide a suitable amount of protection.

Interpretation 4: The end-user of Archistar is an individual with greater/deeper IT knowledge.

1.3 Inactive Stakeholders

According to Lorünser et al. (2016), Archistar is built upon a secure object storage tool that contains the following components:

- *Dealer (Client)* – divides the data into chunks and distributes them to different servers.
- *Reader (Client)* – gathers chunks from servers and reconstructs the data into its original state.
- *Servers (Cloud Storage Provider)* – storage nodes that constitute the destination for data chunks distributed by the Dealer.
- *Verifier (Auditor)* – makes remote checks of the integrity of data chunks stored on the Servers, without knowing its content (see Figure 5).

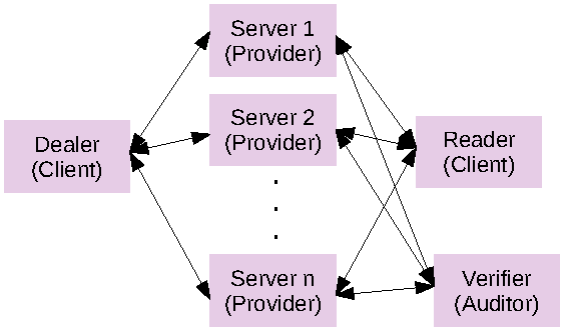


Figure 5. Components of the secure object storage tool.

Lorünser et al. (2016) state that the Dealer and Reader functionality is held by a client (i.e., the customer) with read and write access, whereas the Verifier usually represents a third-party auditing service. Brocca et al. (2016) describe that an *External Auditor* is the main form of "Inactive stakeholder" in the e-Government scenario. As an inactive stakeholder, this entity does not personally deploy a PRISMACLOUD-enabled solution or play an *active* role in any use case. The auditor should have high technical knowledge and may be employed to validate that the Archistar solution fulfils functional and security-related requirements.

However, the secure object storage tool (on which Archistar is built) can be customized for more than one purpose. In the context of the e-Government use case, the tool is tailored to function as a secure archiving (or backup) service, but in a Smart City-related scenario, it may rather be designed for secure data sharing (Lorünser et al. 2016). In Lorünser et al. (2017:39), it is indicated that the aforementioned components (i.e., Dealer, Reader, Servers and Verifier) are arranged differently when the secure object storage tool is adapted to the e-Government use case. In contrast to the descriptions in Brocca et al. (2016) and Lorünser et al. (2016), it is suggested the verification might be made by the client rather than a third-party auditor. Furthermore, the customer may also employ *in-house* servers/storage nodes when distributing data chunks (see Figure 6).

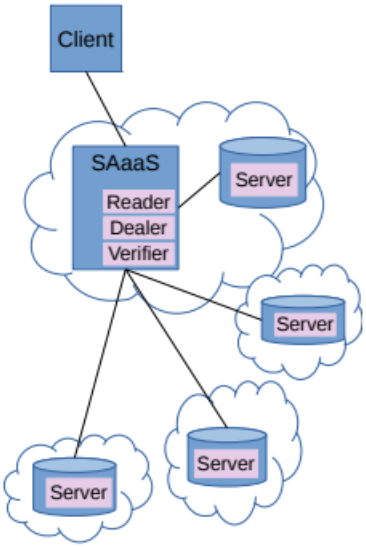


Figure 6. Components of the secure object storage tool customized to function as a Secure Archiving service (SAaaS).

Interpretation 5: Archistar should provide means for users to evaluate the cloud solutions themselves. Thus, the user may need a "high degree of technical knowledge" equivalent to an external auditor.

Appendix B. Written consent form utilized in the Interviews and Walk-throughs.

CONSENT FORM

Within the scope of the EU H2020 project PRISMACLOUD on "Privacy and Security Maintaining Services in the Cloud", Karlstad University is conducting interviews/walk-throughs to derive trust factors, suitable configurations and guidelines that are crucial for the usage of 'Secret Sharing' in organizations or by private users. This concept implies that data is divided into "chunks" which are distributed to numerous Cloud Storage Providers to decrease the risk of data loss and data breaches.

If you consent, the interview will be recorded electronically or manually. Results will be aggregated and analyzed on a non-individual level, and will be published in a Master's thesis and possibly in research papers and/or reports. No personal data will be published at any time. You have the right to withdraw your consent at any time during data-collection, request access to your data and demand the deletion or correction of the material, recording your interview or any other personal data relating to you.

☐ I consent that data about collected during the interview can be processed under the conditions described above.

Name, date

Signature.

☐ I consent that my participation can be recorded electronically.

Name, date

Signature.

Contact details:

Erik Framner erik.framner@kau.se
John Sören Pettersson john_soren.pettersson@kau.se
Simone Fischer-Hübner simone.fischer-huebner@kau.se
Alaa Alaqra alaa.alaqra@kau.se

Data Protection Officer dpo@kau.se

Appendix C. Interview Questionnaire.

Questions on Secret Sharing

Gender: ◯Man ◯Woman ◯Undefined **Background:** (3min)

Age:

Occupation: (optional)

Country:

Background Questions (7min)

1a. What are the different types of applications/data that you currently store/backup – or intend to securely store/backup in the Cloud?* (For example: Documents/Publications, Photos, GitHub, Music, Games etc.)

Application/Data type 1:

Application/Data type 2:

Application/Data type 3:

* If you do not store/backup more than one or two types of application/data, leave the fields for "Application/Data type 2" and/or "Application/Data type 3"empty. This principle applies to the rest of the questions as well.

1b. How would you rate these application/data types in terms of *Sensitivity*?

	Low	Medium	High
Application/Data type 1:	◯	◯	◯
Application/Data type 2:	◯	◯	◯
Application/Data type 3:	◯	◯	◯

1c. How would you rate these application/data types in terms of *Confidentiality requirements*?

	Low	Medium	High
Application/Data type 1:	◯	◯	◯
Application/Data type 2:	◯	◯	◯
Application/Data type 3:	◯	◯	◯

1d. How would you rate these application/data types in terms of *Availability requirements*?

	Low	Medium	High
Application/Data type 1:	◯	◯	◯
Application/Data type 2:	◯	◯	◯
Application/Data type 3:	◯	◯	◯

1e. How would you rate these application/data types in terms of *Integrity requirements*?

	Low	Medium	High
Application/Data type 1:	◯	◯	◯
Application/Data type 2:	◯	◯	◯
Application/Data type 3:	◯	◯	◯

(If you have any additional thoughts on question 1a-e, these can be written below)

1f. Which of the following risks are (most) severe for you?

- ☐ Availability issues
- ☐ Data Loss
- ☐ Breach of Data Confidentiality

1g. Have you experienced any of the threats listed above?

- ◯ Yes:
- ◯ No

Geographical distribution (10min)

2. Data and backups could be lost in a Cloud Storage Provider's data centers due to power outages or natural disasters such as fire, earthquakes or floods. If you were to adopt a Secret Sharing solution to minimize the risk of data loss, what are the requirements in terms of:

a. Number of shares in total (m) and number of chunks needed for reconstruction of data/backup (n)?

	m:	n:
Application/Data type 1:		
Application/Data type 2:		
Application/Data type 3:		

Minimum
m=3, n=2

b. The minimum distance between storage units?

Application/Data type 1:		km
Application/Data type 2:		km
Application/Data type 3:		km

(If located in the same country.)

The storage units should be located in different…

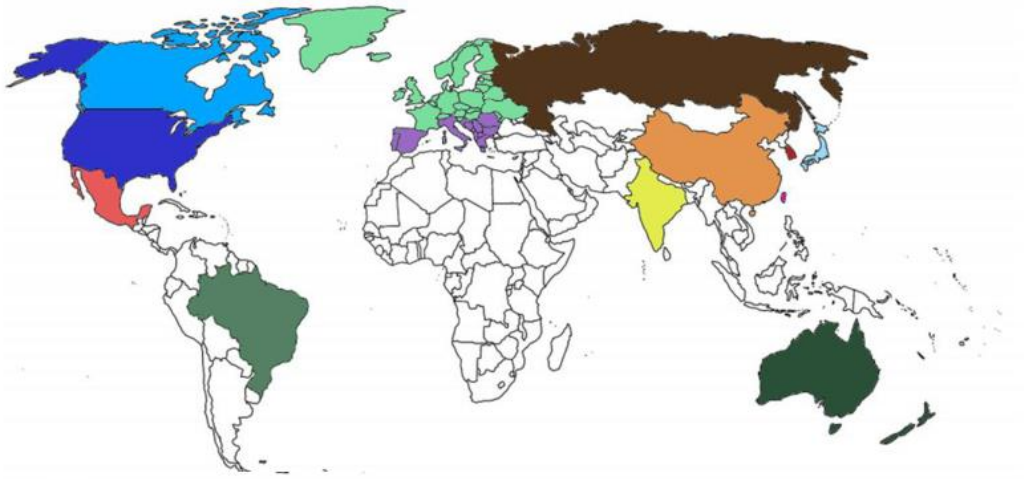
	Cities	Counties	Countries	Continents
Application/Data type 1:	◯	◯	◯	◯
Application/Data type 2:	◯	◯	◯	◯
Application/Data type 3:	◯	◯	◯	◯

The storage units could be located in the following climate zones:

	Tropical	Subtropical	Temperate	Cold
Application/Data type 1:	☐	☐	☐	☐
Application/Data type 2:	☐	☐	☐	☐
Application/Data type 3:	☐	☐	☐	☐

(If you have any additional thoughts on question 2a-b, these can be written below)

2c. Imagine that you were to select *different* countries/regions in which chunks of your backups should be distributed to *prevent data loss*. Which are the top 5 countries/regions that you would choose from the list below? (Leave fields for not trusted countries/regions empty.)



☐ Southern/southeastern Europe	☐ Brazil	☐ Japan
☐ Rest of Europe	☐ Russia	☐ Singapore
☐ Canada	☐ China	☐ South Korea
☐ US	☐ Australia & New Zealand	☐ Taiwan
☐ Mexico	☐ India	

2d. What influenced your rating?

Distribution of Cloud Storage Providers

3. To what extent do you agree with the following statements?

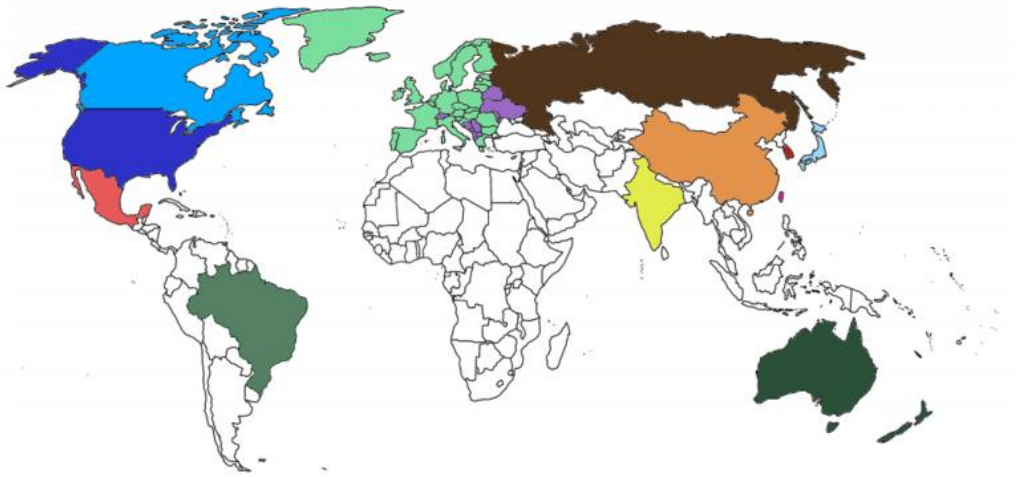
a. I believe that using this solution increases the trustworthiness of Cloud Storage Providers.

	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
Application/Data type 1:	○	○	○	○	○
Application/Data type 2:	○	○	○	○	○
Application/Data type 3:	○	○	○	○	○

b. I'm concerned that Cloud Storage Providers will collaborate and reconstruct my data behind my back.

	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
Application/Data type 1:	○	○	○	○	○
Application/Data type 2:	○	○	○	○	○
Application/Data type 3:	○	○	○	○	○

4a. Imagine that you were to select different countries/regions in which chunks of your data should be distributed to ensure *high confidentiality*. Which are the top 5 countries/regions that you would choose from the list below? (Leave fields for not trusted countries/regions empty.)



☐ EU (incl. EES)	☐ Brazil	☐ Japan
☐ Other countries in Europe	☐ Russia	☐ Singapore
☐ Canada	☐ China	☐ South Korea
☐ US	☐ Australia & New Zealand	☐ Taiwan
☐ Mexico	☐ India	

4b. What influenced your rating?

4c. Imagine that you were to select *three* specific countries/regions in which chunks of your files should be distributed to prevent *illegal cooperation* between Cloud Storage Providers. Which would you choose?

Location 1:	Location 2:	Location 3:
Select a Country/Region ▾	Select a Country/Region ▾	Select a Country/Region ▾

4d. Why were these specific countries/regions chosen?

What type of Cloud Storage Providers would you trust? (5min)

5a. How important is it that the Cloud Storage Providers follow privacy legislations?

	Very Unimportant	Unimportant	Neutral	Important	Very Important
Application/Data type 1:	◯	◯	◯	◯	◯
Application/Data type 2:	◯	◯	◯	◯	◯
Application/Data type 3:	◯	◯	◯	◯	◯

5b. Which country's/region's legislations should the Cloud Storage Providers follow (if any)?

5c. How important is it that the Cloud Storage Providers have a trust or privacy seal (i.e. certification/stamp of approval)?

	Very Unimportant	Unimportant	Neutral	Important	Very Important
Application/Data type 1:	◯	◯	◯	◯	◯
Application/Data type 2:	◯	◯	◯	◯	◯
Application/Data type 3:	◯	◯	◯	◯	◯

5d. How important is it that the Cloud Storage Providers have a high trust rating?

	Very Unimportant	Unimportant	Neutral	Important	Very Important
Application/Data type 1:	◯	◯	◯	◯	◯
Application/Data type 2:	◯	◯	◯	◯	◯
Application/Data type 3:	◯	◯	◯	◯	◯

(If you have any additional thoughts on question 5a-c, these can be written below)

Applicability of Secret Sharing (5min)

6a. Does Secret Sharing seem secures enough for storing data in data centers *within* your organization (i.e. Private Cloud)? Why?

6b. Does Secret Sharing seem secure enough for storing data in data centers that are located *outside* of your organization (i.e. Public Cloud)? Why?

6c.Imagine that you use Secret Sharing and need to access a data backup; chunks will then need to be gathered and combined in order to reconstruct the data to its original state. For how long would you be ready to wait until this process is completed?

6d. Are you familiar with Hybrid Clouds?

◯ Yes

◯ No

Are you familiar with Community Clouds?

◯ Yes

◯ No

6e. Imagine that you were to use Secret Sharing in a Hybrid Cloud (i.e., combination of Public and Private Cloud). How many chunks would you keep internal and external?

6f. Imagine that you use a Community Cloud, where private individuals share parts of their disk space with others within the community to get backup capacities for personal data. Would Secret Sharing be an adequate security measure in this scenario? Why?

6g. Imagine that you use a Community Cloud, where public bodies like province governments and city councils share storage capacities with each other. Would Secret Sharing be an adequate security measure in this scenario? Why?

Security Measures (10min)

7a. Do you see any advantages of the keyless nature of Secret Sharing?

7b. Do you see any disadvantages?

7c. Are there any cases or types of data that could require encryption instead of - or on top of - Secret Sharing?

7d. Which security measures would you like to use to protect your applications/data in the Cloud?

	Standard encryption	Secret Sharing	Neither	Both
Application/Data type 1:	◯	◯	◯	◯
Application/Data type 2:	◯	◯	◯	◯
Application/Data type 3:	◯	◯	◯	◯

8a. For how long would you like your data to be (securely) stored in the Cloud?

	1-5 years	5-10 years	10-25 years	25+ years
Application/Data type 1:	◯	◯	◯	◯
Application/Data type 2:	◯	◯	◯	◯
Application/Data type 3:	◯	◯	◯	◯

8b. Data fragmentation on top of encryption can guarantee the secretly shared encrypted data is unbreakable and impossible for quantum computers* to crack in the future. Do you think it is important to be prepared for/be safe from possible future attacks by quantum computer? Why?

**A (hypothetical) computer that can break asymmetric cryptography.*

Security tradeoffs of Secret Sharing (5min)

9a. Imagine that you in your everyday work use a Cloud-based application containing a database that is protected with Secret Sharing. This application involves the aspects listed below. How would you prioritize these aspects? (From 1 to 5.)

	Security	*(i.e. number of shares)*
	Cost	*(for hosting backup storage)*
	Usability	*(e.g. number of steps involved in tasks)*
	Performance	*(e.g. latency)*
	Reliability	*(e.g. trustworthy performance and error rate)*

9b. What influenced your rating?

9c. Do you see any trade-offs between these aspects?

Appendix D. Introduction Script used in the User Walkthroughs.

BACKGROUND:
"This study is part of an EU project called *PRISMACLOUD* whose purpose is to develop "privacy and security maintaining technologies" for the cloud.

In the use case that the study falls into, the focal point is a security measure called "Secret Sharing". This solution implies that data (backups/archives) is divided into "chunks" which are then distributed to different cloud storage providers. If a data breach occurs in one of the clouds so that unauthorized individuals can get a hold of ONE chunk, they will still not be able to access any information from it. Because you need a certain number of chunks to recreate the data. Thus, data is protected in the cloud without having to handle an encryption key.

Such a solution is intended to be used in Lombardy (nothern Italy) to protect cloud data of municipalities and other regional organizations. Partly, chunks will be distributes to internal/state clouds, but you may also use external/commercial clouds such as Amason AWS."

INSTRUCTIONS:
"Today, you will try to use a possible user interface for the Secret Sharing solution, known as 'Archistar'.

Worth noting is that this interface does not represent the system where data backups/archives are created. Instead, it serves as a suplement to the backup/archiving system. You will specify how many chunks the data should be divided into and which cloud storage providers to use, etc. (A configuration file would then be saved locally on your device. Subsequently, this file would be inserted to the backup/archiving system where you also specify which set of data will be backed up/archived in the cloud.)

You will try to create a configuration for a "typical" data set that you may backup/archive in your everyday work. This process is divided into steps. We will pause at each step and ask you some questions. You will also have the opportunity to ask questions if there is anything that you are wondering about.

During the walkthrough, we would like to use screen and voice recording. Data that is collected may be used in writings related to the PRISMACLOUD project. Information about your identity will be treated with the highest possible confidentiality, and will not be disclosed in or outside the confines of these writings. Whenever you want, you can withdraw your participation and demand that collected information will not be utilized in the study. The purpose of the study is *not* to put you to the test. If something is difficult to understand, it will be seen as a sign that the user interface is unclear and needs to be improved."

Access Key:	Karlstad
Password:	Test
N:	3
k:	2
Chunks in External Clouds:	k-1

Appendix E. Prepared questions for the User Walkthroughs.

STEP 1:

- Thoughts on the proposed prioritization procedure?
- Any aspect that feels less important?
- Any important aspect that is missing?
- Based on the user's prioritization of these aspects, some recommended settings will be automatically selected by the system. Do you think automation is appropriate in this scenario or would you prefer to select all settings manually?

STEP 2:

- Is the difference between "Service Credentials" and "Encryption" apparent?
- How is data protected in your business/organization today?
- Would a layer of encryption be necessary if Secret Sharing were to be used in your business/organization?

STEP 3:

Monthly budget:

- What is your thoughts on the notion of entering a budget/cost restriction first in the configuraiton form?

Estimated Size of Data:

- You should estimate how large the data volume will be. Do you see any problems with this?

Number of Chunks (N), Restore Threshold (k), Chunks in External Clouds:

- Is the information about "availability rate" and "downtime per year" helpful or misleading?
- Is it clear that "external" and "internal clouds" refers to?
- What do you think "k-1" means?
- What alternatives would you have chosen?
- Is there any relevant option that is missing? Would you have preferred other options here? (For instance, to specify a number of chunks that should be in the EU)

- Is there any input field in the configuration form that should not be mandatory?
- Is there any input field in the configuration form that should be hidden and automatically selected by the system?

Map:

- Is it clear how to select CSPs in the UI?
- Is it clear which clouds are "internal" and "external" on the map?
- Ponder that you would create a "community cloud" with county councils and municipalities, and subsequently allow public bodies to distribute data chunks to different data centres.
    - Does it sound reasonable to allow customers/users to choose where the data will be stored?
    - How "detailed" should the choice be? (e.g., county, municipality, city, datacentre)
- Thoughts on the different map views/layers?
- Are they helpful or misleading?

- Does the map miss any relevant view/layer?

STEP 4:

- Does this step provide a clear overview of selected CSPs?
- Some information that is missing?
- Would you have preferred another type of presentation?

STEP 5:

- What do you think of the Configuration screen?
- Now that a configuration has been created, is there any information ou think is missing?
- Is it appropriate to send a summary report to other users (e.g., clients) or do you prefer another solution?

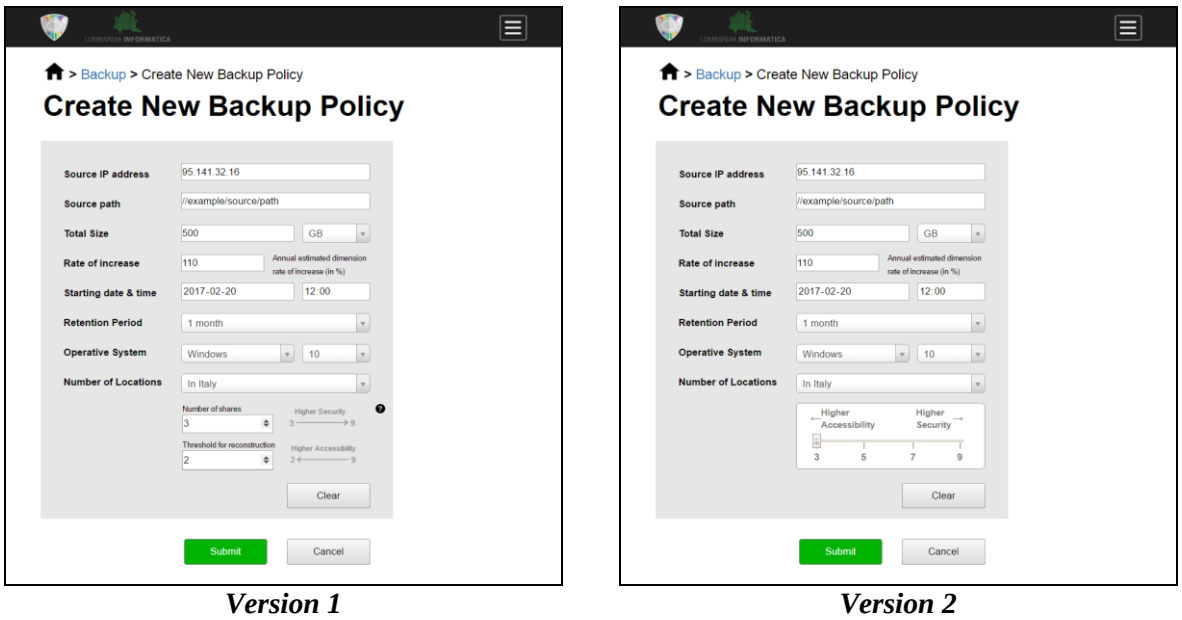
Appendix F. Description of previous User Interface (UI) proposals.

1. Pilot Study

The pilot study prototype included screens for a start page as well as subpages with functionality corresponding to the Dealer, Reader and Verifier components (described in Appendix A, Section 1.3). However, the UI did not include screen elements for performing the actual data splitting/fragmentation or the distribution of chunks. Instead, it allowed users to create "policies" (or *configurations*) for future backups/archiving projects. This served as a description/plan of how the Secret Sharing mechanism and multi-cloud infrastructure should be arranged to safeguard data. In other words, the user was provided with the options to (1) create, modify or delete configurations for future backup/archiving projects; (2) restore data from backups; and (3) execute a data integrity check on-demand or schedule periodical checks.

The prototype was rather primitive and simply provided an abstract illustration of the kind of information that the user may have to specify in order to complete the aforementioned tasks. The process of actually dividing data and distributing chunks to different CSPs, or the process of accumulating chunks from different providers during the data restore operation, was not simulated in the prototype.

Individual walk-throughs with 5 backup experts were conducted to evaluate each screen in the UI prototype. During the assessment, the subpages for creating backup configurations (see Figure 7 and Figure 8) received most comments from participants (e.g., the confirmation received once a configuration had been created was perceived as unclear and insufficient). This indicated that this part of the UI was in greatest need of revision and improvements.



Version 1 ***Version 2***

Figure 7. Page for Creation of New Backup Policy (i.e., Configuration) in the Pilot Study Prototype.

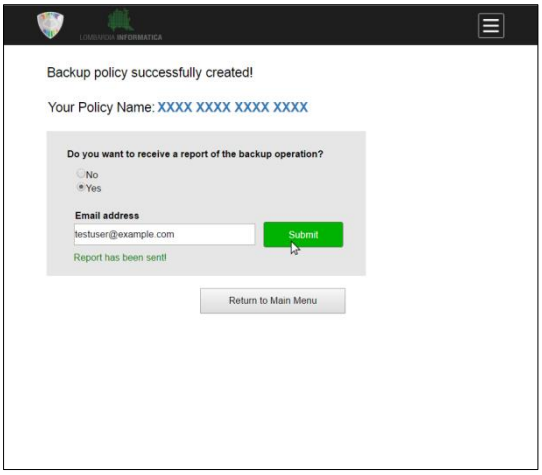


Figure 8. Confirmation Screen in the Pilot Study Prototype.

2. Mock-up by LISPA

In the alternative UI solution, presented during a PRISMACLOUD plenary meeting, it was proposed that the page for creating configurations for future backups/archiving projects should allow users to calculate the cost (i.e., "calcola costo") before data chunks are generated and distributed to different CSPs. The intention was that the users should be able to change values back and forth in the configuration form and simultaneously see how different settings affect the overall price. Thus, the user would be able to recognize that the selected configuration settings will exceed a prospective budget *before* the configuration is completed (see Figure 9).

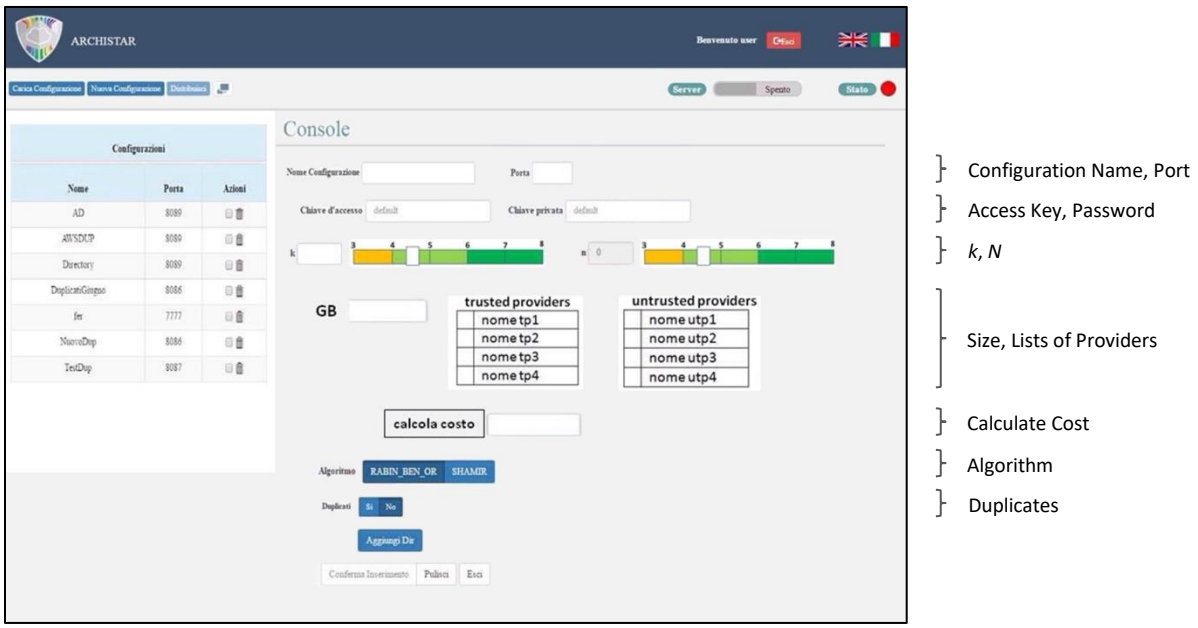


Figure 9. Mock-up presented by LISPA.

3. Observations from Previous Prototypes/Mock-ups Compared with Interview Findings

The following section describes observation/shortcomings taken into consideration (regarding *settings* for the data splitting/fragmentation and distribution of chunks) prior to the development of the new UI proposal.

3.1 Secret Sharing Parameters

Pilot study prototype: Two versions of the page for creating backup/archiving configurations were presented during the walkthroughs. In the first version, the total number of chunks that data should be divided into (N) was specified using a range slider, while the second version had a numeric input field. Only the latter provided the user with means for also selecting a threshold for data reconstruction (k). Both versions failed to properly communicate the relationship between the Secret Sharing parameters due to its complexity.

In the user interface, proposed in the pilot study, arrows and text were utilized to indicate what different values on the Secret Sharing parameters would imply. The first version suggested that a *low* number of chunks (N) would lead to "higher [data] accessibility"[19], while a *high* number would result in "higher security"[20]. However, this rule rather applies to parameter k (which was not considered in the first version of the pilot study prototype). That is, a lower threshold (k) typically makes it easier to reconstruct the data into its original state, while a high threshold may have the opposite effect and thereby keeps the information better protected from intruders.

The second version described that a low number on k would increase the availability of data and (wrongfully) that a higher value on N would enhance the "security". This also constituted an inaccurate description, partly because the Secret Sharing parameters was not considered *in relation to each other*. That is, the contrast between the value on N and k (i.e., $N - k$) may be significant for both the availability and "security". A high contrast allows users to retrieve the information even if multiple chunks are inaccessible, lost or corrupted. A low contrast means that a larger proportion of chunks needs to be gathered in order to reconstruct the information, meaning that unauthorized individuals can less easily get a hold of it.

LISPA's mock-up: Values for the Secret Sharing parameters (i.e., N and k) were both selected with sliders. As an UI input control, sliders have a bar that illustrates the full range of values that the user can choose from (Cooper et al. 2014; Galitz 2007). Sliders prevent invalid inputs because no value outside of its boundary can be specified, regardless of the user's action (Cooper et al. 2014). However, while a *minimum* value for the Secret Sharing parameters can be determined beforehand (i.e., $N = 3$ and $k = 2$), a *maximum* value cannot. Unless the user has a predefined budget for the backup/archiving project, there is no upper limit for the value on N. In order to fully benefit from the Secret Sharing mechanism, the threshold for data reconstruction should in turn be *lower* than the total number of chunks (see Table 2). Thus, the highest possible value on subset k can only be accurately displayed in the UI once the user has selected a value on N. In LISPA's mock-up, the sliders for N and k had the same scale (ranging from 3 to 8), even though the user should ideally have a more limited set of options when selecting a value for parameter k.

Colours were utilized to indicate the level of protection that different values would entail. That is, the sliders for N and k both had a background colour that shifted from yellow/amber (low value) to dark green (high value). Yellow and green – along with the colour red – are widely used as status indicators (McKay 2013) and to communicate risks. The hierarchy of aforementioned colours suggests that red is riskier than yellow, while yellow is riskier than green (Bostrom et al. 2008). The colour yellow/amber is commonly interpreted as a warning that something is about to go wrong, while green suggests that everything is in order (Yuk 2014). However, in the context of the Archistar UI, use of colours would give the impression that a high value on both Secret Sharing parameters is always the optimal solution

[19] Referring to Data Availability.
[20] Referring to Data Confidentiality.

which may not be the case. High value on N may imply higher costs, making such a configuration less suitable if the user has a constrained budget. High value on k (in relation to N) may entail a higher risk of data loss or inaccessibility if e.g. multiple CSPs were to experience an outage simultaneously. Thus, the latter configuration is less suitable if the user/organization is highly dependent on the information in question.

Interviews: In accordance with the previously created UI proposals, the interviews suggested that the users should be provided with an indication of what different values on the Secret Sharing parameters will imply as this may not be clear to them. Without knowing the implications of different values, potential *security trade-offs* may also not be considered or comprehended by the users. To avoid that the Archistar configuration will end up providing insufficient data availability *or* confidentiality, the users could specify their protection needs/goals in the UI and subsequently receive a recommendation of values on N and k. For addressing the trade-off issue, a prioritization needs to be made in the UI between confidentiality and availability (i.e., which factor is desired the most). Specifying needs with a data classification scheme such as the one utilized in the interview questionnaire (see Appendix C) may not indicate a priority, since equally "High" requirements could be selected for both factors. Furthermore, prospective needs for minimizing *cost* should also be specified in the prioritization as this may put restrictions on potential values on N.

3.2 Selection of Cloud Storage Providers and Geographic Restriction

Pilot study prototype: It was suggested that the user should specify geographical restrictions *before* selecting values for the Secret Sharing parameters.[21] Geographical restrictions were specified in terms of a region in which data chunks should be stored (e.g., "In EU"). It did not allow the user to apply different restrictions to different chunks (i.e., all chunks would be stored within the same region). Moreover, the proposed UI did not provide means for selecting specific CSPs.

LISPA's mock-up: It was suggested that the destination of data chunks should be selected *after* values for the Secret Sharing parameters had been specified. Rather than specifying geographical restrictions, the user would select specific CSPs that should be utilized in the multi-cloud solution. Information about the data centre location(s) of each provider was not communicated by the UI. Furthermore, CSPs were labelled as either "trusted" or "untrusted" in the UI without explaining why. The user was not provided with information that would allow them to make their own assessment of a certain provider's trustworthiness.

None of the previously proposed UIs provided a comprehensive description of what the multi-cloud's underlying infrastructure would look like. That is, the pilot study prototype indicated a broad geographic region in which employed data centres would reside, whereas LISPA's mock-up indicated which CSPs would be utilized. The distance of a particular data centre in relation to the user and other data centres could not be determined in either of the UIs.

Interviews: Whether CSPs and data centre locations are trusted by the users may depend on numerous factors (e.g., compliance with privacy laws, safety from natural disasters, and reliability of the infrastructure). Prior to decisions regarding the geographical distribution of data chunks, users may not be fully aware of the conditions around particular data centre locations as well as service offerings of individual CSPs. Thus, such information could be communicated by the UI.

From the interviews, it was concluded that locations (to which data chunks should be distributed) should be selected in a flexible manner, rather than applying a fixed minimum distance between all

[21] Given that the user fills the configuration form from the top to bottom.

data centres. The conditions around – and the distance between – available locations could still be displayed in the UI by using visual aid.

Users should be able to form a multi-cloud consisting of both private and public clouds. Private clouds may be perceived as more trustworthy, and users may, therefore, have a preference when it comes to the number of private and public clouds to employ in the Archistar solution. Hence, the UI should describe the deployment model of each individual cloud storage service.

Appendix G. Report about design decisions in the new User Interface (UI) proposal.

1. Creation of a New User Interface Proposal

As described by Garrett (2010), human-centred design is a practice for accomplishing an engaging and efficient *user experience*. The creation process of a user experience is divided into five layers/planes (see Table 4).

Table 4. Planes of User Experience described by Garrett (2010). They are presented in ascending order (i.e., bottom plane first).

Plane	Description
Strategy	Defines what the customer as well as the system provider should get out of the system.
Scope	Defines what features and functions should be included in the system.
Structure	Defines what pages the system should be divided into, and the path one should take to reach each individual page.
Skeleton	Defines which interface elements should be used on particular pages and how they should be arranged.
Surface	Brings everything together and defines what the final product will look like visually.

The *Strategy* for the Archistar solution is described in Appendix A (i.e., facilitating migration of sensitive data to the cloud by providing users with a secure distributed storage solution). The previously described interviews (see Section 3.6) can, in turn, be regarded as an elicitation of user requirements/needs which would be used to determine which features/options should be included in the system. The new UI proposal would allow for further elicitation and refinement of user requirements.

While the final product of Archistar should include features for (1) storing data in a distributed manner, (2) reconstructing fragmented data, and (3) performing data integrity checks, the *Scope* should be narrower in this thesis. The new UI prototype would constitute a decision-making support system, with a focus on means for creating a configuration for upcoming backup/archiving projects. Design decisions would concern steps that the Archistar configuration should be divided into (*Structure*) and interface elements that should be used on each step (*Skeleton*). The aesthetics would not be a focal point during the design process.

Design decisions in the new UI proposal would be based on:

- The result of conducted interviews (see Section 4.1).
- Observations and shortcomings in previous UI proposals (see Appendix F).
- Information provided on the official websites of popular CSPs (see Appendix H and I).
- General design principles mentioned in literature about e.g. HCI, User-centred design, UX design, Interaction design, Interface design, Web design, Usability principles, and Data visualization.

2. Dividing the Configuration into Steps

As suggested above, previous UI proposals may not consider all relevant factors and, therefore, include an insufficient set of input fields/controls. Increasing the number of screen elements *below* the previously proposed input fields would mean that all of them would not fit within a “screen length”

and scrolling would be required. Jarrett and Gaffney (2009) argue that UI forms should have a sufficient length to ask all relevant questions, but still be kept short to minimize the user's cognitive effort. Brinck et al. (2002) suggest that users often base their decision-making on information that is instantly perceptible in the UI, and important content should therefore be visible immediately on screen without the need of scrolling.

Long, scrollable pages are difficult to scan (Hoekman 2010), and people have a tendency to look more thoroughly at content that appears close to the top of webpages, while information presented low on the page is given less and less attention (Nielsen 2006). In the Archistar configuration form, crucial information may not only appear in the beginning, meaning that the user should remain attentive throughout the entire configuration process. Furthermore, the configuration form would include input fields/controls that are connected (i.e., changing the value of one configuration parameter may require other parameters to be updated as well). As argued by Galitz (2007), long pages that require scrolling often result in users losing a sense of *context*, because the spatial proximity of related information is increased. When scrolling is used, some of the information may even disappear entirely from the view. Often, information that is out of sight is also out of the user's mind, meaning that the user may fail to recognize/register the connection between the information elements (Galitz 2007). Thus, it was determined that a page layout that requires manual vertical scrolling should be avoided.

Jarrett and Gaffney (2009) describe that there are two options for shorting the length of a UI form, i.e.: (1) abbreviate labels and pack all input fields/controls together so that they fit onto as few screens as possible, or (2) divide elements into multiple smaller pages. The former option may be perceived as overwhelming and give the impression that all questions are being "shouted" at once (Jarrett & Gaffney 2009). Furthermore, a dense webpage layout may reduce the readability of content (Hoekman 2010), and a "cluttered" presentation may make it difficult for users to find and focus on the important information on the page (Colborne 2011; Krug 2014). Clutter also gives websites an unprofessional appearance, which could have a negative impact on trust among online consumers (Lal 2013).

Dividing information into several pages that are shown individually may, on the other hand, make the content easier for users to process (Wieczorek et al. 2014). Thus, it was proposed that the configuration should be divided into multiple steps which would be presented one at a time to reduce clutter. That is, input fields/controls belonging to a currently open configuration step should all be visible simultaneously on the screen, while screen elements corresponding to other steps should be hidden. The configuration steps would be represented in the UI by utilizing so-called "Steppers"[22]. Such elements can display the user's progress through a sequence of logical and numbered steps, but also serves as a means for navigating through each part of the configuration process (see Figure 10). Although the user would not be navigated to a new page when he/she proceeds to the next configuration step, each stage of the configuration would still be shown *separately* and manual scrolling would not be required.

[22] https://material.io/archive/guidelines/components/steppers.html

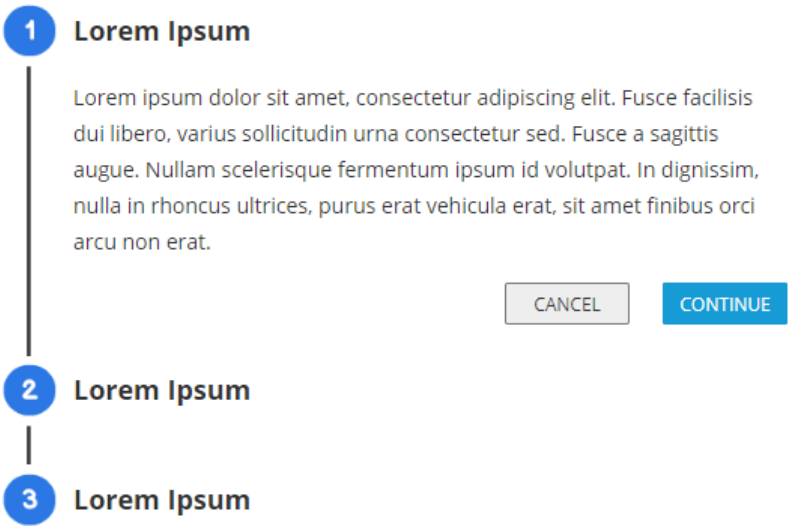


Figure 10. Steppers element.

While using a system, users do not tend to memorize the presented information or utilize their ability to plan in advance. Thus, when dividing information into pages/steps that are presented separately, one should ensure that the content on each page/step functions *independently* (Brinck et al. 2002). That is, users should be able to make decisions on a configuration step without having to go back and check selected settings on a previous step. The thesis author proposed that the configuration process should be divided into the following steps:

(1) **Priorities** – the user should indicate their priorities/needs before configuration settings are selected (this procedure would correspond to the data classification made in the beginning of the previously conducted interviews, see Appendix C). Based on specified priorities, the UI could assist the user in making decisions that suit their needs.
(2) **Name configuration** – corresponds to the first four input fields in LISPA's UI mock-up (see Figure 9), which could be presented on their own as they do not have an interconnection with the remaining fields. A name is specified which will be used to distinguish the configuration from other configurations (for previous backup/archiving projects). Furthermore, the user will have the option to add encryption as an extra level of protection to the Secret Sharing mechanism (depending on the prioritization made in the previous step, encryption may even be mandatory).
(3) **Secret Sharing details** – corresponds to the remaining controls/input fields in LISPA's UI mock-up. Configuration settings related to the data splitting/fragmentation and the geographical distribution of data chunks are made. Furthermore, some characteristics of the data are specified (e.g. size) in order for the system to estimate the cost.
(4) **Overview** – presents a summary list of chosen CSPs, service offerings and data centre locations. Furthermore, the total cost of the selected configuration settings is shown. (This step would resemble a summary of an online shopping cart.)
(5) **Confirmation** – provides the user with feedback that the configuration has been successfully completed.

Design decisions regarding interface elements on each configuration step will be described below.

2.1 First Configuration Step: User's Priorities

On the first step, users should indicate their needs by describing the order in which different protection goals are prioritized. During the previously conducted interviews, it was indicated that *data*

confidentiality, *availability*, as well as *cost* would be determining factors when finding a suitable configuration for the Secret Sharing parameters. The majority of respondents were also able to make a prioritization between confidentiality and availability issues (i.e., which of the two is most crucial to be protected against), suggesting that such a procedure would be feasible. Accordingly, the protection goals presented in the UI proposal would be: (1) *Cost Minimization – Low Cost*, (2) *Data Protection – High Confidentiality*, and (3) *Data Loss Prevention – High Availability*.[23]

Each of these factors could imply different values on the Secret Sharing parameters. Table 5 shows suitable configurations of the Secret Sharing parameters, depending on the user's protection goal. Trade-offs of such configurations is also indicated.[24]

Table 5. Trade-offs of Protection goals.

Protection Goal	Suitable Configuration	Negative Trade-off / Drawback
Cost Minimization	Low N to decrease storage overhead.*	Lower redundancy increases the risk of availability issues/data loss.
High Data Confidentiality	Low $N - k$ (i.e., the threshold for reconstruction should be high in relation to the total number of chunks). Thereby, data is less easily disclosed due to collusion/data breaches.	If a higher subset of chunks is needed for reconstruction (k), the risk of availability issues/data loss is also higher.
High Data Availability	High $N - k$ (i.e., the total number of chunks should be significantly higher than the threshold for reconstruction). Thereby, greater redundancy is achieved.	If the total number of chunks (N) is higher, adversaries have more subjects to choose from for targeted attacks.

** Assuming that a Perfect Secret Sharing (PSS) algorithm is utilized and that each chunk has the same size as the original data.*

The aforementioned goals can be prioritized in the UI by ranking them from the "most" to "least" important. (Next section describes the reasoning behind the selected method/screen elements for making such a ranking in the UI.)

2.1.1 Methods for Ranking Items in a UI

Blasius (2012) compare different *methods* for ranking items in a web-based interface – i.e., arrows, checkboxes, drag and drop, as well as numbering (see Table 6 for a description of each method). In Blasius' study, it is suggested that "drag and drop" is the most appropriate method when a small number of items are ranked. Other conclusions/observations were that:

- The "drag and drop" as well as the "arrows" method both visualize items in their intended order. On the other hand, the "checkboxes" and "numbering" methods do *not*, meaning that the actual order has to be imagined by the user which increases the difficulty of performing the ranking task.

[23] Apart from *Confidentiality* and *Availability*, information security also involves a third aspect known as *Integrity* (Bhowmik 2017). However, when it comes to Secret Sharing, high confidentiality and integrity requirements have similar implications – i.e., unauthorized individuals should not have access to k chunks (so that they can read/modify the original data). Thus, these two aspects will not be distinguished in the UI.

[24] This has been established after the interviews, through discussions with other researchers in the PRISMACLOUD project.

- When used in a web survey, the "arrows" method never produces incomplete answers, since the order presented by default is selected if no changes are made by the user. The "drag and drop" method was also less likely to generate incomplete responses, compared to the "checkboxes" and "numbering" methods.
- The "arrows" method is inconvenient for relocating items from the bottom to the top of the list (or vice versa) since the up/down arrow needs to be clicked for each position the item is to be moved. When utilizing the "drag and drop" method, a lower number of clicks is required since the item only need to be moved once.

Table 6. Methods for ranking items, compared by Blasius (2012).

Method:	Description:
Arrows	Each item in a list is accompanied with an up- and down-arrow which can be clicked to give them a higher or lower placement in the list.
Checkboxes[25]	Beside each item in the list, columns with checkboxes for "first" to "last choice" are presented. Only one checkbox can be checked on each row and column (i.e., items can only be given *one* ranking, and the ranking of each item should be *unique*).
Drag and drop	Items are listed in a column (i.e., the starting point). They should individually be dragged and dropped in another column (i.e., the drop area) in the order they are prioritized.
Numbering	Each item in the list is accompanied with an input field where the user should specify a number for how highly it is prioritized (e.g., 1 represents "first choice", 2 represents "second choice", etc.).

In Karth (2011), it is indicated that users prefer to utilize "drag and drop" over the "numbering" method in ranking tasks. It was *perceived* by users as easier and faster to use, although in reality there was no significant difference in completion time.

Regardless, finishing quickly is not necessarily an advantage since it may imply that the user puts insufficient thoughts and effort into the task completion. Satisficing (i.e., taking mental shortcuts in decision-makings) may entail that answers or entered values in the inquiry form will have both poor quality and accuracy (Conrad et al. 2017). When it comes to the similar task of *rating* items, Kunz (2015) suggests that "drag and drop" leads to a higher burden on respondents and, therefore, longer response time than a conventional scale with radio buttons (equivalent to the "checkboxes" method). As a result, users also tend to be more attentive and careful when using "drag and drop" to perform the rating task, and less susceptible to shortcuts in the task completion.

Out of the aforementioned methods, "drag and drop" appeared to include most benefits (i.e., visualizes the actual order of items, easy to rearrange the order if needed, user acceptance, keeps users attentive, and lower risk of inappropriate shortcuts in the task completion). Thus, it was determined that protection goals should be prioritized by using drag and drop (see Figure 11).

[25] Or more appropriately: Radio buttons.

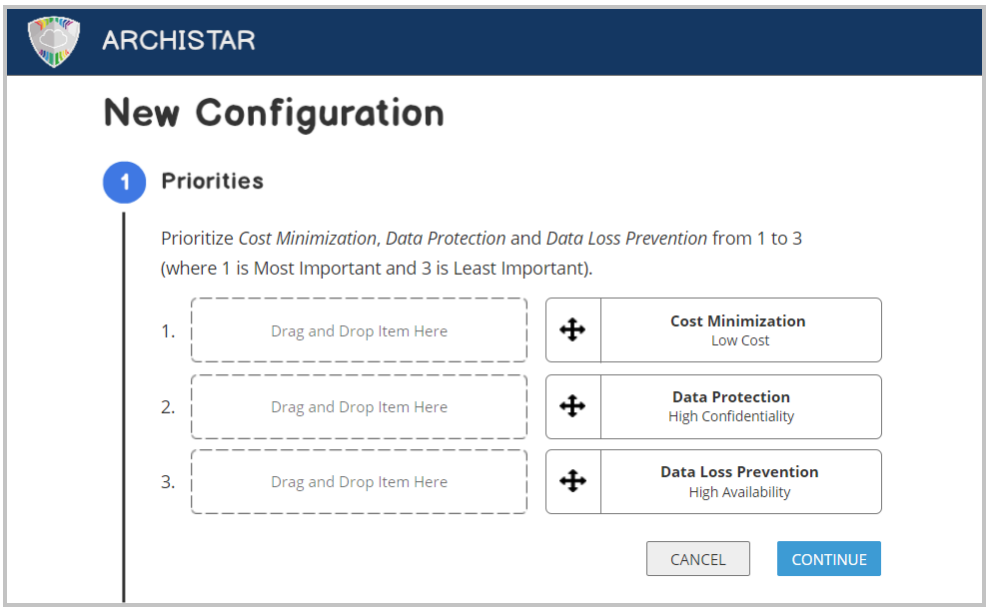


Figure 11. The first configuration step in the UI proposal.

2.1.2 Automation

Based on the user's prioritization of "protection goals", the UI can assist in making decisions during the configuration process. The thesis author made the assumption that the user may find more than one protection goal important for their backup/archiving project. The two goals that are not ranked as the user's lowest priority would therefore be combined to subsequently establish suitable default settings.

System configurations are often complex and suitable setting may be difficult for users to select. A way to prevent misconfigurations caused by human error is to *automate* configuration settings (Yin et al. 2011; Xu & Zhou 2015). By providing default values automatically in the UI form, the amount of work and efforts needed to complete the task is also reduced. However, default values are sometimes overlooked by the user and may be accepted even if he/she would prefer to utilize other settings. If it would involve extensive efforts to *change* the result from a wrongful acceptance, then pre-selected values should be avoided (Weinschenk 2011).

Given that the user is provided with an overview of selected settings before reaching the final configuration step in the UI (i.e., the confirmation), the user would have the opportunity to double-check and potentially reconsider some of the (automatically or manually) selected settings before the configuration is completed. If the user wishes to change a setting, he/she can simply click on the Stepper's "Back" button to return to the previous configuration step. Moreover, previous UI proposals (see Appendix F) included a screen for "modifying" (or editing) already created configurations. Although the new prototype would not represent such a procedure, its design would not prevent such features from being added to a final product. In other words, the effort to change mistakes during or after the configuration process is assumed to be non-exhaustive.

However, not all settings can be selected in an automated manner since it is difficult for systems to self-determine configurations that require external information (e.g., details from the user) (Xu & Zhou 2015). If there is no reasonable basis for making a particular value the default, then pre-selected values should not be utilized (Johnson 2008).

Despite the prioritization made in the first configuration step, there would still be certain aspects in later steps for which suitable configuration setting cannot be predicted (e.g., attributes such as "data size"). Fields for aspects that may vary significantly depending on the circumstances of the

user/organization – as well as the information being backed up or archived – would not be populated with default values. Default settings would, on the other hand, be provided for the Secret Sharing parameters (see Table 8), and the UI would also provide a recommendation for whether or not a layer of encryption should be added to the protection (see Table 7).

2.2 Second Configuration Step: Name Configuration

On the second step, a name is specified for the configuration. The user would also have the option to add encryption as an extra level of protection to the Secret Sharing mechanism. However, as indicated in Section 2.3.1 and 2.3.3, the use of encryption can potentially increase the risk of data loss. When encrypted data is reconstructed from k chunks, it would still be unreadable without the cryptographic key. If the user forgets or loses the key, he/she would no longer possess the means for converting the information back from "cipher-text" to legible plain-text. This means that encryption may not be appropriate for all type of data.

Allowing users to pick and choose *freely* in the user interface might be a desirable solution, but could potentially lead to unwanted results due to mistakes or unsuitable selections (Mishra 2009). Rather than presenting the user with an error message once an issue occurs, Nielsen (1995) argues that one should prevent problems from happening in the first place. This can be accomplished by eliminating error-prone conditions (Nielsen 1995) or the source from which errors/issues may originate (Colborne 2011). Thus, in order to reduce the likelihood of errors/issues, *restrictions* may be applied to the options that the user can choose from. One solution for adding such constraints is to *hide* irrelevant, inappropriate or potentially harmful features in the user interface (Lidwell 2010; Mishra 2009).

Mistakes and errors/issues can also be prevented by providing the user with a *confirmation* option (Nielsen 1995). A dialogue that asks whether the user is sure about going through with a certain action is a common approach. However, although this may cause the user to reconsider a selection, it may also have a disruptive effect on the user's concentration and increase his/her cognitive load in the performance of a task (Colborne 2011). Another solution for preventing mistakes and errors/issues is to *disable* and grey-out input fields/controls for features that are inappropriate at a given time (Carey et al. 2014; Mishra 2009). The users are sometimes provided with means for enabling a disabled feature, allowing them to proceed with an action that is not available by default. As described by Ford (2015) and Lior (2013), checkboxes are often utilized in user interfaces as a "toggle" for turning system features on or off.

Disabling and greying-out the encryption option by default could signal to the user that he/she is not required (or even advised) to use it. Even if the user is allowed to manually enable the feature, this procedure would require the user to spend an extra thought before a layer of encryption is added to the protection. This may be appropriate in scenarios when there is a trade-off situation (i.e., the benefits from using it does not necessarily outweigh the drawbacks). Hiding the encryption feature, on the other hand, would constitute a more drastic solution since it would no longer be visible in the user interface and the user may, therefore, be unaware that the function exists. This may be suitable in scenarios where all potential risks of data loss should be eliminated.

The severity of the "damage" from losing the encryption key (and subsequently the information protected by it) would be indicated by the user's prioritization of protection goals. In other words, the prioritization made in the first configuration step would determine whether the user should be required or even have the option to add encryption (see Table 7 and Figure 12).

Table 7. Provision of the encryption option depending on users' priorities of protection goals.

Top two priorities	Implications	Encryption?
High Data Confidentiality & ***Cost Minimization***	The data is sensitive/personal and needs the highest possible protection.	Yes. A layer of encryption is *required/mandatory*.
High Data Confidentiality & ***High Data Availability***	The user/organization has to compromise since there is a trade-off between availability and confidentiality.	Yes/No. The option to add encryption is *disabled* by default in the UI. However, it can be enabled if the user desires.
High Data Availability & ***Cost Minimization***	The user/organization is (highly) dependent on the data's existence. Precautions should be taken against any incident that could cause data loss.	No. The option to add encryption should be hidden in the UI, to eliminate the possibility of losing data due to key loss issues.

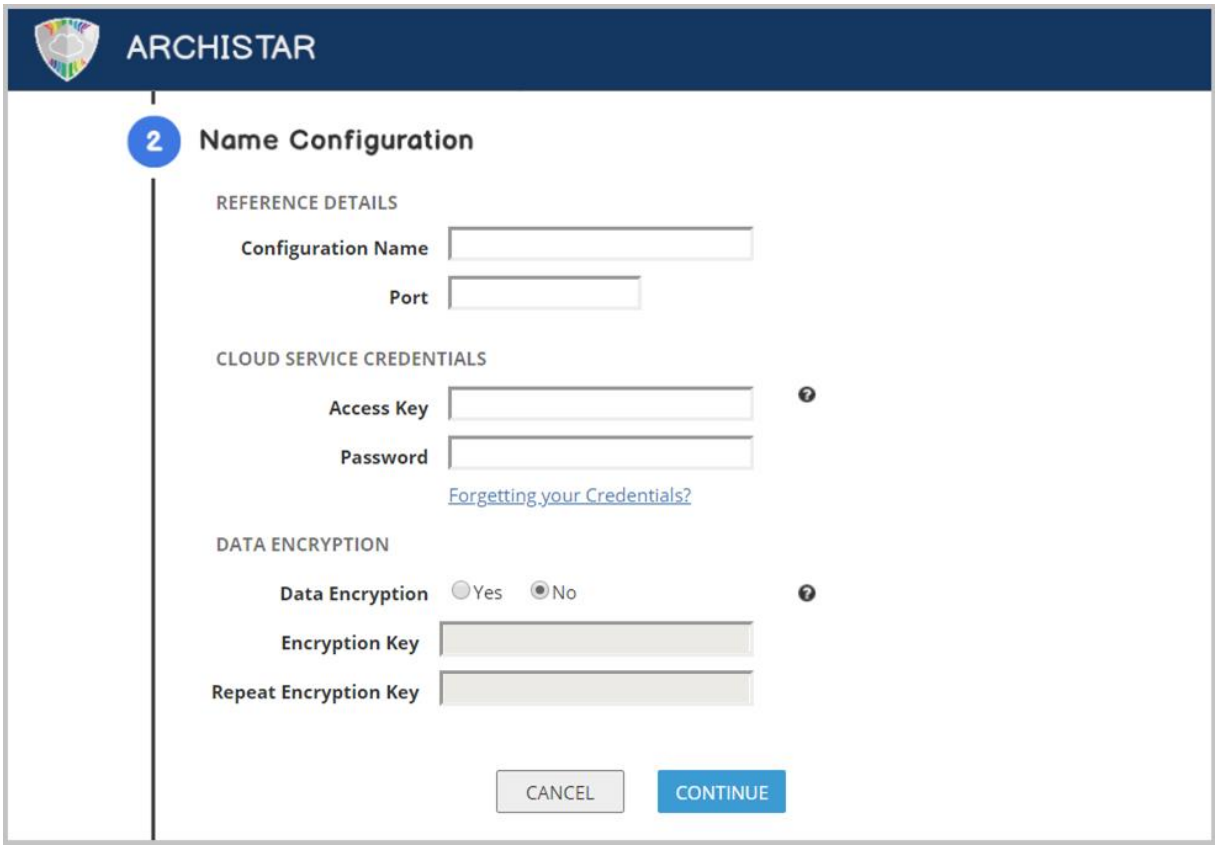


Figure 12. The second configuration step in the UI proposal (if High Data Confidentiality and High Data Availability is the top two priorities).

In similarity to LISPA's UI proposal (see Figure 9), each configuration entry would be made in regular text fields, allowing the user to freely choose a configuration name and encryption key of his/her liking.

2.3 Third Configuration Step: Secret Sharing Details

On the third step (see Figure 13), the user would specify the attributes of the data that is to be backed up/archived (see 2.3.1 in Appendix), and enter settings related to the data splitting/fragmentation (see

2.3.2 in Appendix) as well as the geographical distribution of data chunks (see 2.3.3 – 2.3.4 in Appendix).

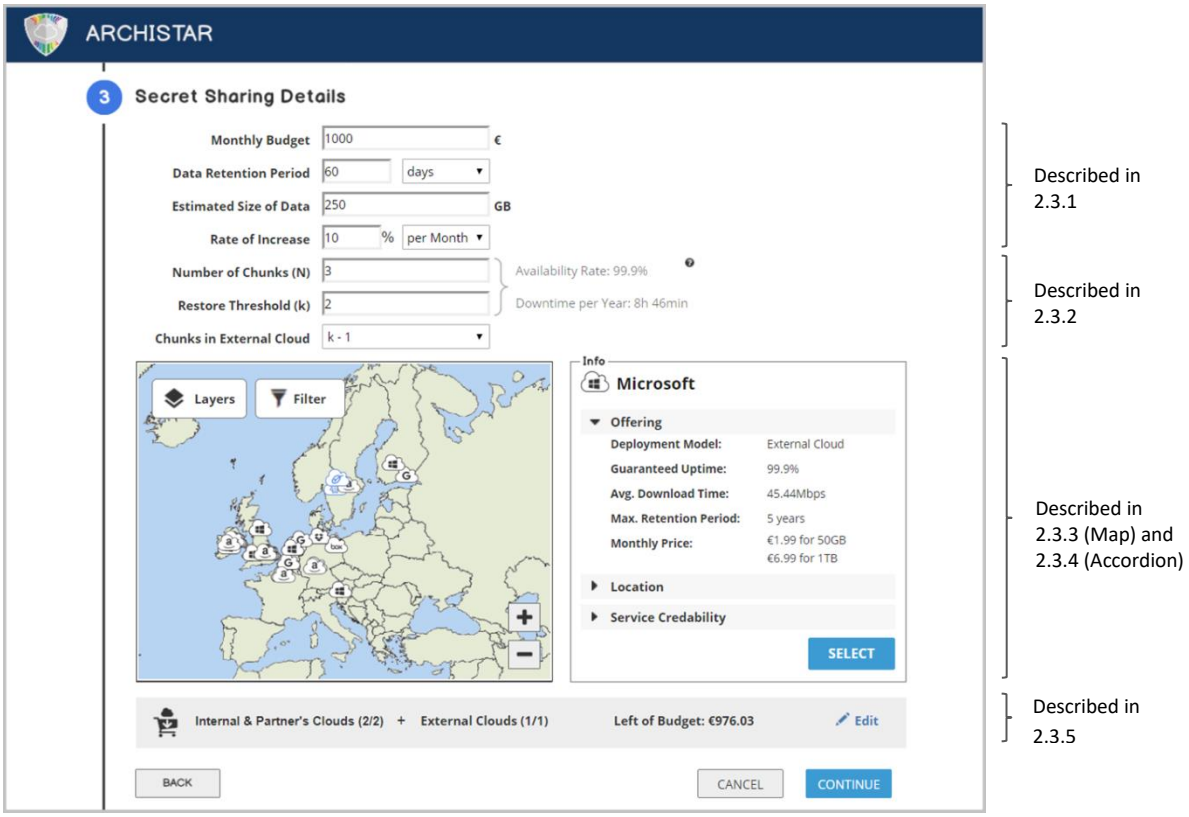


Figure 13. The third configuration step in the UI proposal.

2.3.1 Data attributes

While the first configuration step indicated how critical or sensitive the data is, the third step would indicate how much storage space the data would consume. In accordance with the Pilot study prototype (see Figure 7), numeric input fields would be provided for the following aspects:

- "Estimated Size of Data" – indicates how much storage space would be needed initially.
- "Retention Period" – indicates how long the data (chunks) should be stored in the user's personalized multi-cloud.
- "Rate of Increase" – indicates how much the user's needs in terms of storage space would change over this time period.

Furthermore, if "Cost Minimization" is highly prioritized, the user would have the option to specify a "Monthly budget". The UI would assist the user in keeping track of how much is left of the budget once the user starts selecting service offerings (see 2.3.5 in Appendix).

2.3.2 Indication of relationship between Secret Sharing Parameters

As suggested in Appendix F, sliders would not constitute a feasible control for the selection of values on the Secret Sharing parameters (i.e., N and k). The reason being that neither of the parameters has a predetermined upper limit and a full range of possible values cannot be visualized in the UI. In similarity to the Pilot study prototype, numeric input fields would therefore be utilized instead. Although a predefined maximum value would not exist for either N or k, default configurations would be suggested by the UI based on the user's priorities of protection goals (see Table 8).

Table 8. Suitable values suggested in the new UI proposal for N and k, depending on the user's priorities of the protection goals.

Top two priorities	**Relatively Low N** (Less Storage Overhead and Lower Cost)	**Higher k than the Minimum** (More Effort to Reconstruct Data)	**Higher Service Uptime** (Greater Availability)	**Example**	
				N	k
High Data Confidentiality & ***Cost Minimization***	Yes	Yes	No (99.9%)	**4**	**3**
High Data Availability & ***Cost Minimization***	Yes	No	Yes (99.999%)	**4**	**2**
High Data Confidentiality & ***High Data Availability***	No	Yes	Yes (99.9999%)	**6**	**3**

Different combinations of values on N and k affect the availability of data (chunks) in a multi-cloud setting (see Table 9). The availability of cloud services is commonly expressed in percentage of uptime – or number of "leading 9s" (Bauer & Adams 2012; Happe et al. 2017). Services/data should ideally have an uptime of 99.999% or "five 9s" in the cloud, but such an availability rate is hardly reached by any individual provider. Well-reputed CSPs may promise an availability of 99.9% or "three 9s" (Bhowmik 2017) which constitutes a downtime of 8.76 hours per year (see Table 10). However, generally speaking, a more common guarantee in service level agreements is 98% (Happe et al. 2017). While individual providers may fail to accomplish a desired rate of service availability, the risk of downtime can be lowered when multiple providers are employed (assuming that cloud services do not become unavailable simultaneously).

In order for users to make an informed decision about (whether the default values would be a) suitable configuration of N and k, the relation between the two parameters should be indicated by the UI. In the Pilot study prototype and the LISPA mock-up this was done by using arrows and colours respectively (see Figure 7 and 9) – both of which failed to properly communicate the connection. Thus, another form of representation would be proposed in the new UI proposal. Changing the value of one Secret Sharing parameter may imply that the value on the other will have to be altered as well to accomplish a suitable configuration. Rather than using a *fixed* representation, it would be suggested that the UI should provide information that *adjusts* as the user makes modifications to entries.

Table 9. Number of "leading nines" achieved from different combinations of N and k (Happe et al. 2017).

N	k										
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
3	3	1	0	-	-	-	-	-	-	-	-
4	5	3	1	0	-	-	-	-	-	-	-
5	6	4	2	1	0	-	-	-	-	-	-
6	8	6	4	2	1	0	-	-	-	-	-
7	9	7	5	4	2	1	0	-	-	-	-
8	11	9	7	5	3	2	1	0	-	-	-
9	13	10	8	6	5	3	2	1	0	-	-
10	14	12	10	8	6	5	3	2	1	0	-
11	16	14	11	9	8	6	4	3	2	1	0
12	18	15	13	11	9	7	6	4	3	2	1
13	19	17	15	12	11	9	7	5	4	3	2
14	21	18	16	14	12	10	8	7	5	4	3
15	23	20	18	16	14	12	10	8	7	5	4

Table 10. Service uptime achieved from different numbers of "leading nines".

Number of "Leading Nines"	**Percentage of Service Uptime**	**Service Downtime per Year**		
1	90 %	876 hours	=	52 560 minutes
2	99 %	87.6 hours	=	5 256 minutes
3	99.9 %	8.76 hours	=	525.6 minutes
4	99.99 %	0.876 hours	=	52.56 minutes
5	99.999 %	0.0876 hours	=	5.256 minutes
6	99.9999 %	0.00876 hours	=	0.5256 minutes

Being reactive and providing users with immediate feedback on their inputs can ensure that they are kept informed about the implications of modifying entries in the UI (Scott & Neil 2009). The users can instantly see whether or not a particular action will facilitate his/her goal. If a certain input will lead to an unwanted result, immediate feedback can point the user back to the right direction. It may quicken the user's learning (Galitz 2007), and ensure that the finally selected values/entries will be based on informed decisions rather than guesswork (Scott & Neil 2009). Moreover, it helps the user to correct inaccuracies in his/her mental model and to build confidence in that the right decisions are made (Mathis 2016). Thus, instant feedback on user's inputs can also minimize the risk of errors/issues due to mistakes and unintended actions (Scott & Neil 2009).

Besides the input fields for N and k, information about the "Availability Rate" (i.e., percentage of uptime) and "Downtime per Year" would be presented. As the user changes the values in the input fields, this information would be automatically updated to indicate the direction in which the user is headed (see Figure 14).

Figure 14. Immediate feedback when the user changes the value on the default values on N and k.

2.3.3 Map-mediated Risk Communication

CSPs such as Amazon[26], Box[27], Google[28] and Microsoft[29] all provide some information on their official website about *where* their corresponding data centres reside. As evidenced by Appendix H, the level of detail in which locations are described in text varies from a specific city to a particular country. Each of the aforementioned providers also communicates data centre locations with other means than words – i.e., by pointing them out on some form of visual *map*. For data centre locations whose description in text is merely country-specific, one can assume that the map marker has an arbitrary placement within the country's confines and does not describe the exact location either.

While language is one of the oldest tools of communication and is crucial for informing about risks, visual displays may be more effective and can have a great impact on people's perception of potential threats (Bostrom et al. 2008). Different forms of visual representations may be suitable for different purposes. For instance, bar charts may be mainly utilized for making comparisons, pie charts may be used to represent parts of a whole, and maps may be employed to illustrate *geographical areas* – as well as statistical and geological factors associated with them (Hagen & Golombisky 2017). Natural disasters (e.g., earthquakes) are spatial in nature. Thus, systems with geographic visualization features are commonly used in decision-making regarding the prevention of such hazards (Bostrom et al. 2008). Maps can be effective in assessments of risks with spatial dimensions, since they can illustrate the magnitude of a threat and how widespread it is. Furthermore, they enable a comparison of the same risk at different locations, as well as a comparison of various risks at a single location. In interactive maps, different information components can be joined and linked to a particular geographical location. Thereby, it can help to organize information in a manner that is not overwhelming (Dransch et al. 2010). Although maps are more complex than traditional graphs/charts (Hagen & Golombisky 2017); Yuk (2014) argues that they may be easier for people to grasp. It is suggested that Google has with its widely popular web mapping service (i.e., Google Maps) made many people familiar and comfortable with the use of geographic information systems.

Accordingly, the Archistar UI would present available locations (to which data chunks can be distributed) on an interactive map. While Amazon, Box, Google and Microsoft use traditional pin icons, dots or circles to point out data centre locations on their respective maps, a different type of location marker would be used in the Archistar UI. Given that the infrastructure of *various* cloud storage services should be highlighted at the same time, location markers for different providers needed to be visually distinguishable. Furthermore, while the aforementioned providers present their

[26] https://aws.amazon.com/about-aws/global-infrastructure/
[27] https://cloud.app.box.com/s/u65ydojm3lxcn3hfmbowg0t7f53ukbdd
[28] https://www.google.com/about/datacenters/inside/locations/index.html
[29] https://products.office.com/en-us/where-is-your-data-located?geo=All

own maps on a web page specifically dedicated to describing the locations of their data centres, the Archistar map would be shown on a page with *multiple* purposes. Thus, the intention with the Archistar map (and the subjects being highlighted on it) may not be equally clear at a first glance. Location markers would therefore be given an appearance that was assumed to make the map more self-explanatory. That is, a cloud icon with the provider's company logotype would be utilized. The cloud symbol would indicate that a cloud service is associated with the location, whereas the logotype would clarify *which* specific service the marker is referring to. Markers representing private and public providers would be differentiated by giving them a blue and black outline respectively.[30]

Even though maps may be beneficial in certain scenarios, they could also be confusing if too much data is shown simultaneously, and there may be times when the map presentation needs to be simplified by hiding parts of the visualized information. Furthermore, the user may need means for showing data in a different hierarchy or with another base map in the background (Muehlenhaus 2013). Alternative map views can be utilized if multiple features cannot be shown at the same time, or if the user needs to temporarily view certain data from a different perspective to gain insight about a particular issue (Tidwell 2011). Dividing information into *layers* would allow users to change what is highlighted on the map (Muehlenhaus 2013). Moreover, *filters* that affect the display of content are typically utilized on webpages to enable users to search and to express what they are looking for in a quick, effective manner (Wilson 2011).

Apart from a stripped-down view where country borders are outlined, the map in the Archistar UI would also include alternative views/layers to illustrate aspects that may impose a threat towards the data availability or confidentiality. Examples of map views available online would be used to illustrate the risk of (1) *Wildfires*[31], (2) *Floods*[32], and (3) *Earthquakes*[33]. The vulnerability towards power outages or "cable cut-offs" would also be indicated with an example view of the (4) *Internet Backbone Infrastructure*[34]. Furthermore, the risk of political or legal issues would be communicated with a map layer where (5) *Trade Blocs* (e.g., EU and EFTA) are highlighted (see Appendix J). (The map views/layers were simply used as *examples* for the sake of demonstrating the proposed UI solution. Their accuracy was not analysed/contemplated by the author.)

The map would also feature a "Filter" option, allowing the user to hide/show different types of CSPs on the map. As suggested by the interviews, the perceived adequacy of the Archistar solution may depend on the deployment model of individual clouds and whether or not they comply with EU legislation. Thus, the map would allow the user to filter out CSPs whose services are private (internal) or public (external), as well as data centre locations within or outside the European Union[35] (see Figure 15).

[30] A similar distinction was made in an illustration in PRISMACLOUD *deliverable D8.7 Specification of test-bed configurations for validation phase* (Zambrano et al. 2017).
[31] *Global Wildfire Information Systems (GWIS): Fire Danger Forecast.* http://gwis.jrc.ec.europa.eu/static/gwis_current_situation/public/index.html
[32] *Flood Map: Water Level Elevation Mao (Beta).* http://www.floodmap.net/
[33] *USGS World Earthquake: Heat Map [Unofficial].* https://fusiontables.google.com/DataSource?snapid=S327323orMC
[34] *Example Internet backbone network topology in Europe.* https://www.researchgate.net/figure/Example-Internet-backbone-network-topology-in-Europe-6_fig2_304410350
[35] Although providers outside of EU may abide to the union's privacy laws, they may not be forced to.

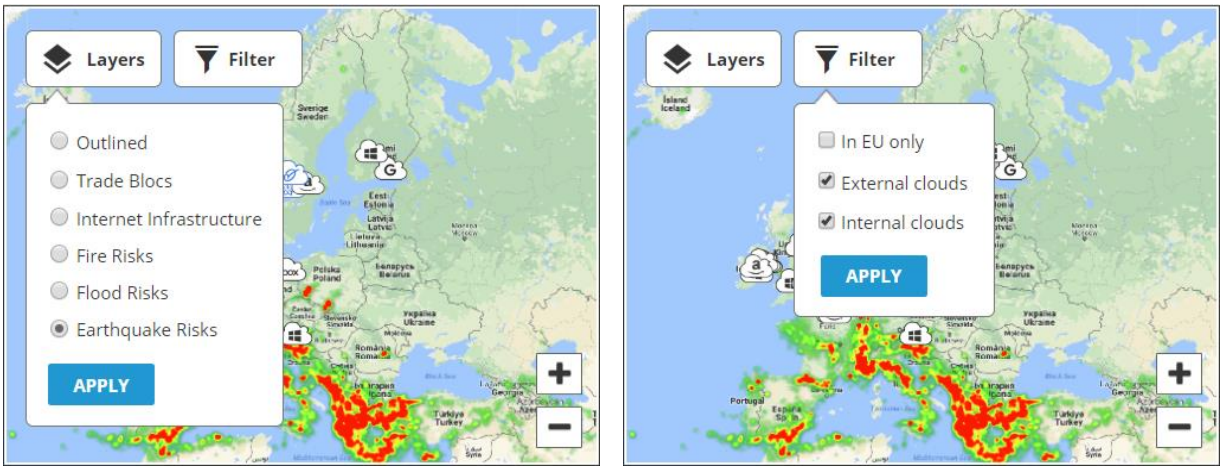


Figure 15. Map view in the UI proposal.

2.3.4 Accordion

While maps may be appreciated by users, information in text form is still necessary in the user interface (Jarrett & Gaffney 2009). Complementing the map with text may be a requisite in order for it to be sufficiently usable. For instance, Galitz (2007) argue that it is sometimes not self-explanatory which areas on an interactive map that are clickable (i.e., so-called "hot spots"), and such instructions could therefore be given to the user with words.

Furthermore, a graphical representation such as a map would not be able to communicate all crucial factors mentioned during the previously conducted interviews (e.g., local privacy laws, trust ratings of the service provider). Therefore, when the user selects a location marker on the map, additional information (in text form) about the corresponding data centre and cloud storage service would automatically appear in the UI. Rather than embedding the information into the map itself, it would be listed in a box/container next to the map to reduce clutter. Before a location marker has been selected, the box/container would provide a hint about which elements are clickable on the map so that the user will be aware of how to use it (see Figure 16).

It would not be possible to display the entire list of information at the same time in the box/container and inline scrolling would therefore be required, making the content less easy to overview. However, various solutions exist for minimizing the length of lists (and thereby also the need for scrolling). Tidwell (2011) as well as Scott and Neil (2009) describe patterns for doing so by relocating details about the items in the list – i.e., *two-panel sector*, *different page*, *overlay*, an *inlay* (see Table 11 for description of each patterns).

Table 11. Different patterns for minimizing the length of a list.

Method:	Description:
Two-panel sector	When selecting an item/category, details are presented in a panel next to the list. This pattern is commonly used in file directories (Tidwell 2011).
Different page	When selecting an item/category, the user is navigated to a separate page on which details are presented. This pattern is commonly used in e.g. email services (Tidwell 2011).
Overlay	When selecting an item/category, details are presented in a dialogue box that appears *over* the list (or other screen elements) (Scott & Neil 2009).
Inlay	Items/categories serve as toggles in the list. When clicking on one of them, corresponding details are presented or hidden *inside* the list (Tidwell 2011).

The first two patterns (i.e., two-panel sector and different page) did not seem appropriate in the Archistar UI. Since the list would already be located beside a map, there would not be enough horizontal space for a two-panel sector layout. Redirecting the user to a new page instead would, on the other hand, imply that the remaining information on the third configuration step would be hidden. In other words, the user would not be able to see the map and details about list items/categories at the same time, making this solution less suitable. When it comes to the third pattern (i.e., overlay), Johnson (2013) and Scott and Neil (2009) argue that dialogue boxes (or pop-ups) are beneficial in situations where the user's work needs to be interrupted to show urgent information. For instance, it can be utilized to bring users' immediate attention to a critical error message (Johnson 2013). However, dialogue boxes (with non-urgent information) are often perceived as annoying (Coleman 2017; Johnson 2013; Mathis 2011), and users have a habit of ignoring or dismissing them without reading its content (Mathis 2011). In the Archistar UI, details about the items listed in the box/container should not be presented in such a manner that it will be mistaken for an error message or be deliberately avoided by the user. Thus, the list inlay pattern would be utilized.

The inlay pattern is employed by so-called "accordions" that groups list content into expandable/collapsible panels (Scott & Neil 2009). It is useful for when there is more content than can be comfortably presented at the same time on one screen (Tidwell 2011; Hagen & Golombisky 2017). It serves as an effective way to hide details until it is needed and for preserving screen space (Scott & Neil 2009). Furthermore, grouping and hiding content can be a very effective technique for de-cluttering a user interface (Tidwell 2011).

The information in the list would be grouped into three categories – i.e., "[Service] Offering", "Location" and "Service Credibility". Each category would serve as an expandable/collapsible panel that would hold related information. (See Appendix K for a complete list of the details presented in each category. The information would correspond to factors brought up in the previously conducted interviews.)

A "Select" button was subsequently placed below the list of information. Thus, if the user makes the assessment that the attributes of the highlighted service/data centre meet their needs and requirements, they can press the button to add it to their personalized multi-cloud infrastructure.

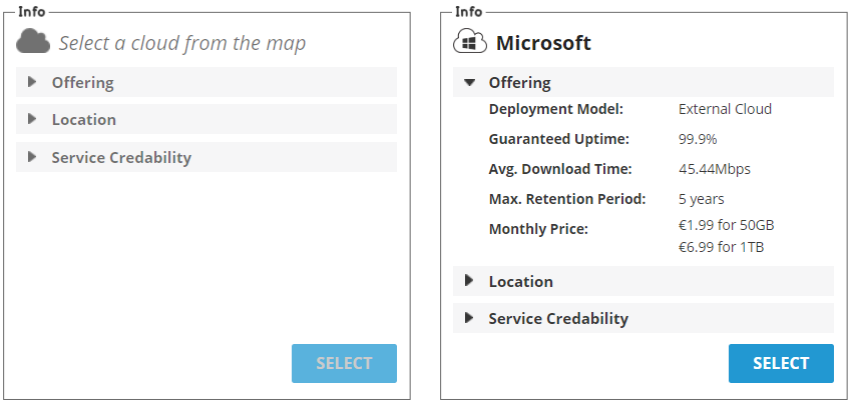


Figure 16. Information box/container, before and after a cloud icon has been clicked on the interactive map.

2.3.5 Shopping Cart Metaphor

Creating a multi-cloud infrastructure for the backup/archiving project would involve procurements/purchases of cloud storage services. Selected services would be treated as products/items in an online shop. As described by Lopuck (2012), a central part of e-commerce websites is the *shopping cart*.

In similarity to a traditional supermarket, customers in online shops are provided with a repository to hold products/items that they intend to purchase. The metaphor of a shopping cart is widely used in online shops and due to its high familiarity, users typically understand the basic functionality and the way it operates (Gao 2005). In online shops, a quick preview of the shopping cart is commonly provided in the website's upper-right corner (Lal 2013; Lopuck 2012). As evidenced by examples provided in Appendix L, the preview may present information about the number of items in the shopping cart as well as the total cost of these items (without requiring any mouse clicks). The shopping cart metaphor would be utilized in the proposed Archistar UI as well. That is, selected cloud storage services on the third configuration step would be added to a shopping cart and a quick preview of it would indicate the total cost. However, if a "Monthly Budget" had been previously specified on the third configuration step, "Total Cost" would be replaced by "Left of Budget". Thereby, the user would not have to make the calculation him-/herself and his/her cognitive load would be reduced.

Moreover, unlike the common standard, the quick preview would *not* be placed in the page's upper-right corner. Instead, it would be placed below the map and the information box/container (see Figure 15 and 16). The reason being that it should be presented in proximity to the "Select [Service]" button so that the user can easily see/notice how the "Total Cost" or "Left of Budget" changes when a data centre location is added to the multi-cloud infrastructure.

As mentioned in Appendix A (Section 1.2), the multi-cloud solution could include both internal (private) and external (public) CSPs. The conducted interviews indicated that the user may have less trust in public clouds and may, therefore, not be willing to rely exclusively on such clouds in the solution. In other words, there might be a requirement that chunks in external clouds should be lower than k so that sensitive/confidential data cannot be reconstructed without at least one chunk from an internal cloud. Right after values has been selected for the Secret Sharing parameters (see Figure 17), the user would have the option specify such a requirement by selecting a value from a dropdown list labelled "[Number of] Chunks in External Clouds".

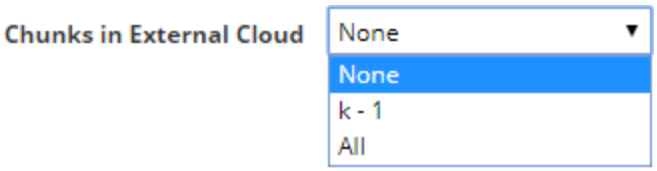


Figure 17. Drop down list for ”Chunks in External Clouds”.

Given that internal and external clouds will subsequently be selected from the same map, the UI would assist the user in keeping track of how many clouds of each deployment model has been chosen. This would be done by distinguishing internal and external clouds in the shopping cart quick preview (see Figure 18).

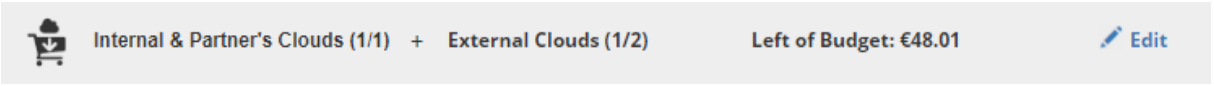


Figure 18. Shopping cart quick preview.

2.4 Fourth Configuration Step: Overview

Apart from a quick preview, the shopping cart on e-commerce websites is typically presented in its entirety on a dedicated page before the checkout of selected products/items (Gao 2005; Lal 2013; Lopuck 2012). On such a page, the following is traditionally summarized for each product/item in the shopping cart (Gao 2005):

- The *product name* and/or *code*.
- The *short description* (if the product name is non-descriptive or temporarily unavailable).
- The *quantity* (i.e., the number of copies of the product/item that the customer wishes to purchase).
- The *price* (per unit).

Moreover, the final cost of all products/items should also be presented (Gao 2005; Lal 2013; Lopuck 2012). The fourth step of the Archistar configuration process would provide an overview equivalent to such a shopping cart summary, presenting a list of the selected cloud storage services.

As evidenced by Appendix H, CSPs often have data centres in various locations. This means that the user can distribute multiple data chunks to the same cloud but still have a significant geographical distance between them (ensuring that they will not be hit by the same disaster). The significance of selected data centre locations would be considered in the presentation of the shopping cart list/summary. That is, apart from the abovementioned factors, the location of employed data centre(s) would be described for each cloud storage service. If multiple chunks are distributed to the same cloud, their geographical location would be presented individually in the list (see Figure 19).

Furthermore, as indicated by Appendix I, providers on the single cloud market may utilize different pricing models. Ponder that the user has selected three clouds in the Archistar UI and needs space for 100GB worth of data in each of them. “Provider A” usually offers storage in a pay-as-you-go manner (i.e., they charge per GB), while “Provider B” and “Provider C” offers storage packages with the size of 50 and 100 GB respectively. Unless a special contract arrangement is made between the CSPs and the company behind Archistar, a uniform pricing model would not exist in the context of the Archistar solution either. However, for the sake of consistency in the shopping cart summary, the pricing of Provider A can also be translated into package pricing. That is, one could argue that the notion of charging “per GB” essentially means that “packages” of storage are still offered – but with the small size of 1GB. In order to meet the user’s needs, the number of packages

needed from each provider may vary. Thus, the "quantity" parameter would become useful in the shopping cart summary (see Table 12).

Table 12. The "quantity" of storage packages that is needed from each CSP may vary depending on the size of offered packages.

	Item	Quantity	Total
Provider A	1 GB	**100**	100 GB
Provider B	50 GB	**2**	100 GB
Provider C	100 GB	**1**	100 GB

When hovering the mouse cursor over a cloud storage service in the list, some further details about it would be presented beside the shopping cart summary (see Figure 19). This would serve as the "short description" that complements non-descriptive product names.

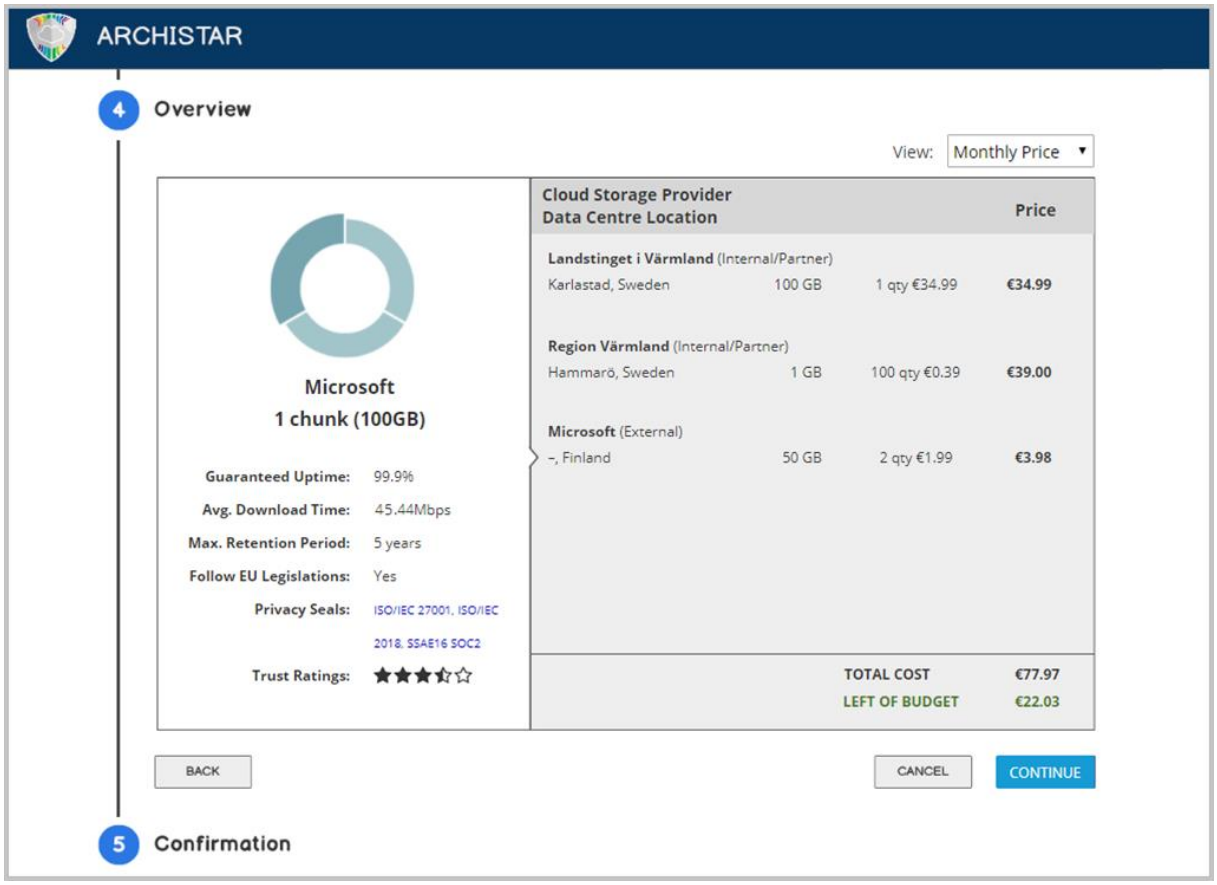


Figure 19. The fourth configuration step in the UI proposal.

8.2.5 Fifth Configuration Step: Confirmation

On the fifth step, the user is provided with a simple confirmation that the Archistar configuration has been successfully completed. In similarity with the pilot study prototype (see Figure 8), the user would have the option to send a confirmation/summary of the configuration to an email address of choice.

However, during the pilot study evaluation, participants expressed that they wanted to receive a summary directly on the confirmation page so that they would be able to more easily verify that everything is in order. Thus, selected configuration settings would also be listed on the fifth configuration step (see Figure 20).

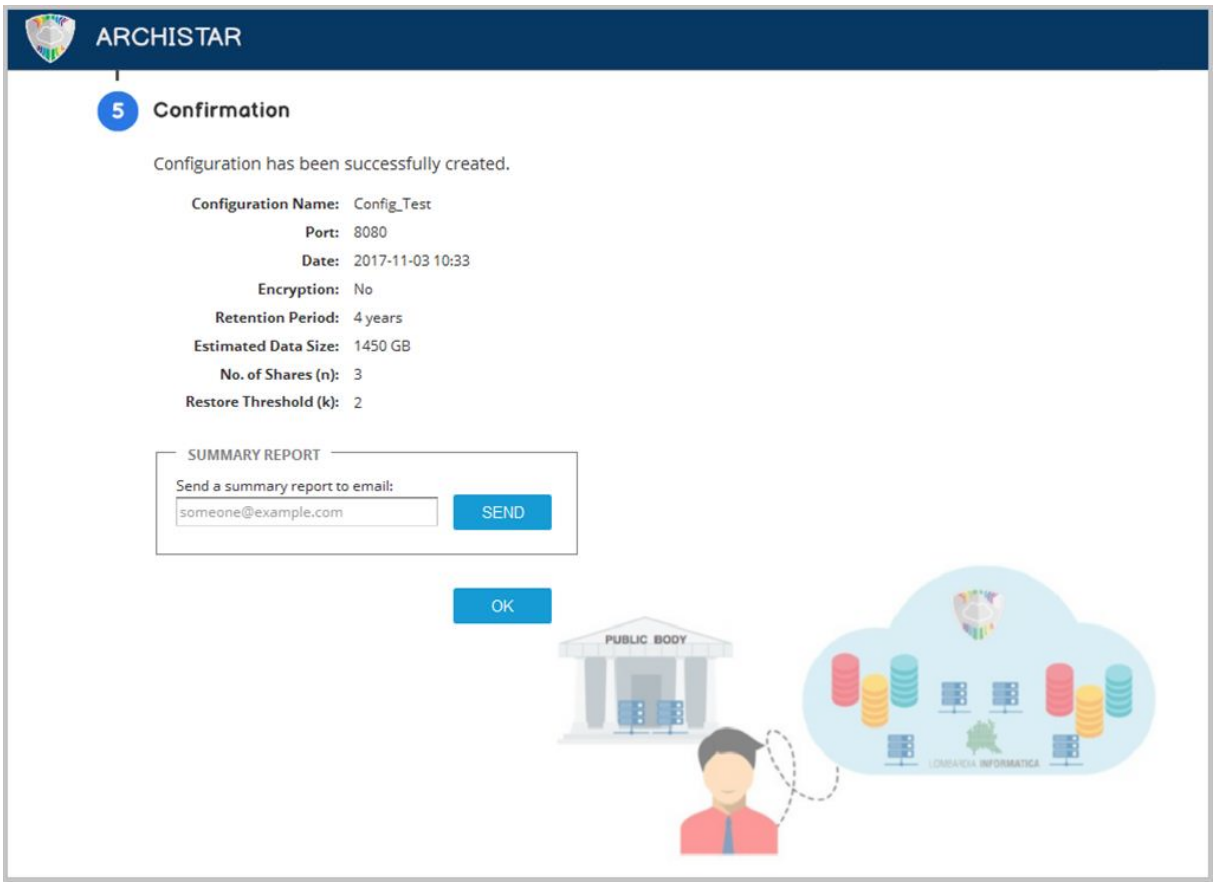
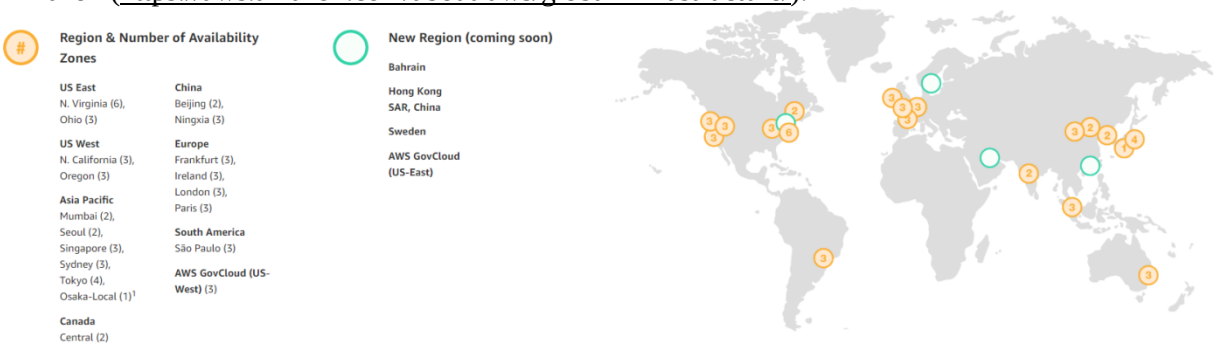


Figure 20. The fifth configuration step in the UI proposal.

Appendix H. Maps utilized on the official website of public cloud storage providers to communicate the location of data centres.

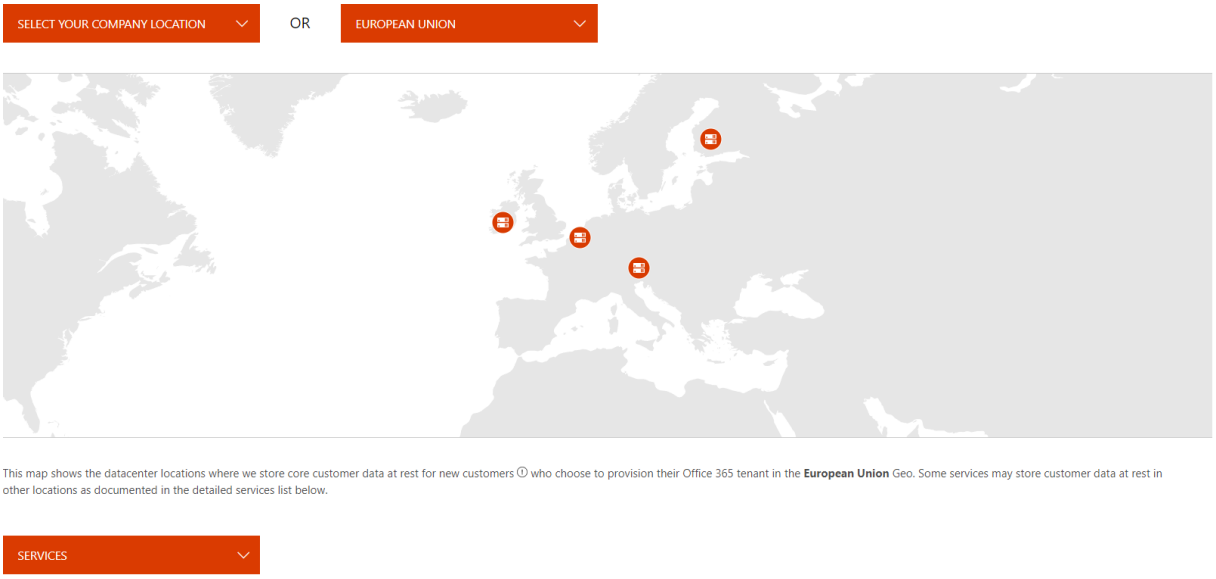
Amazon (https://aws.amazon.com/about-aws/global-infrastructure/):



Box (https://cloud.app.box.com/s/u65ydojm3lxcn3hfmbowg0t7f53ukbdd):

Google (https://www.google.com/about/datacenters/inside/locations/index.html):

Microsoft Office 365[36] (https://products.office.com/en-us/where-is-your-data-located?geo=All):



[36] The map highlights the data centre geography (Geo) that will be defaulted to the customer based on the location associated with their first subscription. However, the map only apply to *new* customers/tenants. Data of customers that are already subscribed to the service may be stored in other regions than the locations marked on the map.

Appendix I. Description of how the dissimilar pricing models have been considered.

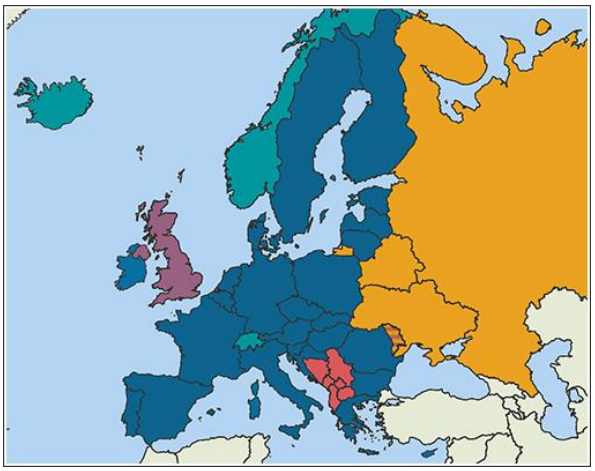
On the single cloud market, CSPs may use different pricing models. For instance, services such as *Google Drive* and *Microsoft OneDrive* offer "packages" with a fixed amount of cloud storage space. 15 and 5 GB are respectively provided by Google Drive and OneDrive for free as a first-time offer, while further storage space can be added for a certain price. Additional storage packages offered by Google Drive are 100GB, 1TB, 10TB, 20TB and 30TB.[37] As for OneDrive, 1 or 5TB is offered to Office 365 subscribers, whereas consumers who are interested in *storage only* can add 50GB at a time.[38] (This type of pricing model implies that the customer is likely to be assigned and charged for more space than their data is occupying. Extra storage space may serve as a buffert, allowing the data to grow a little in size without running out of space.)

On the other hand, *Amazon (AWS)* offers storage space on a pay-as-you-go basis by charging *per GB*. The unit price of Amazon's services varies depending on multiple different aspects, i.e.: (1) *The amount of storage utilized.* Price is reduced once the consumer uses more than 50TB, and decreased even further once the utilized storage is larger than 500TB. (2) *The location in which data is stored.* For instance, Standard Storage hosted in Ireland (EU) costs $0.023 per GB, while the equivalent storage service costs $0.026 per GB when it is operated in Northern California (US West). (3) *The storage class.* Amazon offers various categories/types of storage (i.e., "Standard Storage", "Infrequent Access Storage", and "Glacier Storage") which have different price tags as they are designed to accommodate different availability requirements. However, all classes of storage are not offered in all regions.[39] [40] (Given that it is not conventional for CSPs to offer different storage classes, the Archistar UI proposal would simply provide a single type of storage for each provider.)

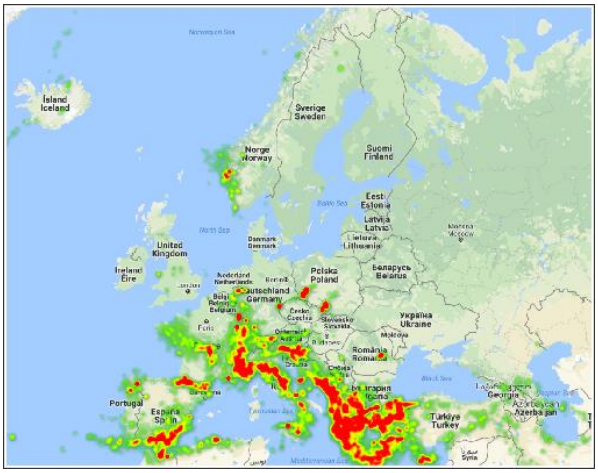
[37] https://www.google.com/drive/pricing/
[38] https://onedrive.live.com/about/plans/
[39] https://aws.amazon.com/s3/pricing/
[40] https://aws.amazon.com/s3/faqs/

Appendix J. Map views used as examples in the Archistar UI proposal.

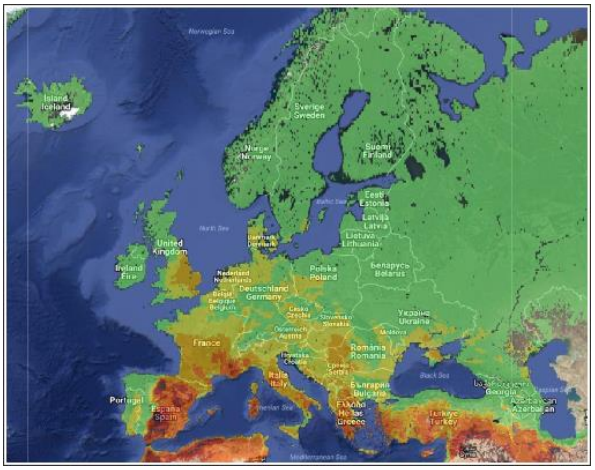
Outlined



Trade Blocs



Earthquake Risks
(https://fusiontables.google.com/DataSource?snapid=S327323orMC)



Wildfire Risks
(http://gwis.jrc.ec.europa.eu/static/gwis_current_situation/public/index.html)

Flood Risks
(http://www.floodmap.net/)

Internet Backbone Infrastructure
(https://www.researchgate.net/figure/Example-Internet-backbone-network-topology-in-Europe-6_fig2_304410350)

Appendix K. Factors presented in each panel of the accordion.

		Information in Panel	Influences…			Description of connections to trust factors mentioned during interviews
			Availability	Confidentiality	Cost	
Panels in the Information Box/Container	[Service] Offering	**Deployment Model**		x	x	The Secret Sharing mechanism may be perceived as less secure when combining it with ”Public” or ”Community clouds”.
		Guaranteed Uptime	x		x	“Availability (of individual clouds)” may be important to know when configuring Secret Sharing parameters.
		Avg. Download Time/Speed	x		x	”Connectivity/access time” may be important to consider when preventing data loss issues.
		Max Retention Period	x		x	*(A “data retention period” is specified in an input form, earlier on the third configuration step. This provides the users with an assurance that their needs are met.)*
		Monthly Price			x	The solution must be “affordable” and “cost-beneficial”.
	Location	**Data Centre Location**	x			*(Information about where the data centre resides in text form, supplementing the map.)*
		Local Privacy Laws		x		Compliance with “Privacy Legislations” may be important to trust CSPs. Local/national laws of e.g. Germany commonly mentioned.
		Trade Bloc Membership		x		”Political relationships” may be important to consider when preventing collusions.
		Government Debt	x			”Stability” may be important to consider when preventing data loss issues and breaches of data confidentiality. *(“Bankruptcy” mentioned in the introduction video shown to the respondents prior to the interview.)*
		Corruption Perception Index		x		
	Service Credibility	**Follow EU Legislations**		x		Compliance with “Privacy Legislations” may be important to trust CSPs. EU/European laws most frequently mentioned.
		Trust Ratings	x	x		“High trust ratings” may be important to trust CSPs.
		Privacy Seals/Certificate		x		Possession of “trust/privacy seals” may be important to trust CSPs.
		Breach Report	x	x		“Publicly known incidents” may be important to consider when preventing collusions.

Appendix L. Quick previews of the shopping cart on e-commerce websites.