Utrecht University

Master’s Thesis

Cyber Threats of Shadow IT in Dutch Higher Education and Research

November 2022

Joost Gadellaa, 5609704

A thesis submitted in fulfilment of the requirements for the degree of Master of Science in

Business Informatics
Department of Information and Computing Sciences

First supervisor:	Dr. K. Labunets
Second supervisor:	Dr. S. Jansen
External supervisor:	Drs. C.H. van Genuchten

This document was originally submitted November 18, 2023. Since the submission, small formatting, grammar and other improvements have been made. In the process of making the research data public we made additional changes to the way quotes are made untraceable to the institutions and persons interviewed. URLs to research data are added after several checks to the anonymization. The current version was published February 3, 2023.

You can check for an updated version via https://edu.nl/683h9

ABSTRACT

Usage of IT (information technology) in higher education institutions (HEIs) is influenced by the sector's diverse user needs and culture of academic freedom, openness, and innovation. Besides the IT managed by the institution, this context gives rise to the phenomenon of shadow IT: *"hardware, software, or services built, introduced, and/or used for the job without explicit approval or even knowledge of the organization"* (Haag & Eckhardt, 2017). Existing research often mentions that this loss of control has severe cybersecurity consequences, but technical details of the consequences are seldom provided. This thesis aims to model the role of shadow IT in HEIs' threat landscapes, to gain deeper insights into the risks as a first step towards cybersecurity risk management for this phenomenon. In a sequential approach, we first conducted two literature reviews on definitions and known cybersecurity consequences. Then, we interviewed eleven information security professionals from HEIs about the shadow IT occurrences and threats they perceive in their organization. These interviews were qualitatively analyzed to systematically identify occurrences and threats until code saturation occurred. This method allowed us to provide a rich set of observed occurrences and threat components by experts, which could then be structured into an occurrence-vulnerability view and several threat diagrams, representing the most important attack paths associated with shadow IT. This specific modeling of threats also allowed us to relate these threats to countermeasures. We conclude that cyber problems related to shadow IT are very diverse, and highly dependent on the prevention, detection and mitigation measures already taken by institutions. The role of shadow IT in the threat landscape can be very manageable if the institution accounts for its existence. Based on the results, we provide guidance for prevention, but also recommend ways in which institutions can responsibly allow and account for shadow IT.

Table of Contents

1 Introduction

Knowledge institutions such as Universities have been around for hundreds or even thousands of years. Guided by oral tradition, patient handwriting monks and, eventually, the printing press, they developed into significant institutions within our societies. In the past century, a new way of working came at them with a staggering stride. Today, increasingly digitalized research and digital native students demand high-quality and extensive IT (information technology) services, and the pace of change is higher than ever. Not only do these institutions have a regular, corporate IT environment, but for their primary operations of education and research, they also provide collaboration software for students and researchers, innovative IT equipment for research, and tools for blended learning.

IT Security for HEIs (higher education institutions) is an interesting research context. The complex, open, diverse and high-stake environment of higher education IT had Bongiovanni (2019) pick *"The least secure place in the universe?"* as the title for their literature review on the information security management of this sector. In the Netherlands alone, there are numerous examples of large incidents, such as a full-scale ransomware attack on Maastricht University in 2019, the Hogeschool and University of Amsterdam having their network security compromised in, or multiple servers with personal information being hacked at Leiden University and the Hogeschool Arnhem Nijmegen, all in 2021[1].

With a large amount of personal data, intellectual property, and computational power present at these institutions, not even mentioning the financial means to pay ransoms, HEIs are an attractive target for malicious actors. Like other organizations, their digital presence is ever-increasing. Since the Covid-19 pandemic forced them to adapt to online education and research quickly, risks have further increased (SURF, 2022).

In the introduction to their review of security risks in higher education, Ulven & Wangen (2021) identify two main factors that separate HE from 'normal' industry: *academic freedom* and *openness*. The valued academic freedom makes the end-users in this sector wary of interference or restrictions from the law, regulations or public pressure, something that carries over into a tension between IT security regulations and the day-to-day IT usage of HEI professionals. The second factor of openness is described as an emphasis on transparency and collaboration (Peter & Deimann, 2013). There is a wish to make technology, knowledge and other resources as accessible as possible, also to the outside world

[1] https://nos.nl/artikel/2316120l, https://nos.nl/artikel/2369091, https://nos.nl/artikel/2387732, https://nos.nl/artikel/2400478

(Schlagwein et al., 2017). This creates another tension with IT security, which often follows the principle of limiting access to resources to a minimum.

There is a constant balancing act between the need for institution-wide solutions and central control, versus the autonomy of departments and individuals, which have traditionally operated in a highly decentralized manner (Coen & Kelly, 2007). Besides the presence of IT managed by the institution, this context gives rise to the phenomenon of shadow IT: ***"hardware, software, or services built, introduced, and/or used for the job without explicit approval or even knowledge of the organization"*** (Haag & Eckhardt, 2017). Shadow IT can be deployed by departments, end-users (including students), or research groups to get work done with no malicious intent.

Research groups might set up their own messaging environment while the university procured one centrally. Freemium tools for integrating digital whiteboards and voting into lectures are used by teaching staff as experiments before they are procured. Researchers spin up virtual cloud machines or host web apps on their own hardware for research without the IT department knowing what is happening.

Besides lack of compliance, lack of integration with other IT and loss of synergy, cybersecurity threats receive the most attention in the literature about shadow IT (Klotz et al., 2019). In industry reports, shadow IT is described using terms such as *"an unprecedented risk"* (D'Arcy, 2011), with research and consulting firm Gartner predicting that in the near future, shadow IT will be involved in one-third of successful attacks experienced by enterprises (Gartner, 2016). A vulnerable (shadow) server could, for instance, be an initial access point for a hacker looking to deploy ransomware, an unencrypted hard drive could get lost and data leaked, or a malicious browser extension could steal company credentials. It has been suggested that the relatively open IT environment diversity of HEIs user groups makes shadow IT not only more prominent but also a bigger risk compared to 'ordinary' companies (Candia, 2021).

This open environment and how valuable it is to the sector is also one of the reasons a total ban on shadow IT is not the preferred solution to the shadow IT problems HEIs are facing. Identifying and managing all shadow IT instances would not be possible or pragmatic, and there is evidence that sanctioning is not an effective solution to shadow IT occurrence (Haag et al., 2015). Moreover, shadow IT systems can be innovative and a useful response to a lack of organizational agility. We refer to Klotz et al. (2019) for a review of causal factors and positive consequences.

Instead, many institutions take a risk-based approach to cybersecurity. The goal is to get a grip on the risks of shadow IT instead of banning the phenomenon. But empirical research on cybersecurity risks in this sector is scarce, with large gaps in the literature (Ulven & Wangen, 2021). One of our interview participants called shadow IT an 'unknown unknown'. For a risk-based approach, it is essential to understand the possible cyber threats related to shadow IT. Therefore, our main research question (MRQ) is: ***"What is the role of shadow IT in the cyber threat landscape of Dutch higher education institutions?".***

For this thesis project, we work together with SURF, the "*collaborative organization for IT in Dutch education and research*"[2]. They are the main community for cybersecurity policymakers and practitioners from HEIs in the Netherlands. Their network enables us to collect the empirical data needed. Later, the results will be disseminated via them and put into practice by their members.

The aim of this research is to gain empirical insights into shadow IT presence and the related cyber threats within higher education organizations as the first step towards comprehensive cybersecurity risk management on this topic. For the scientific community, the contribution consists of an overview of the occurrence of shadow IT and an explicit mapping of cyber threats and possible attack paths. We will also contribute a problem definition of this phenomenon in the unexplored context of education and research institutions. This will enable researchers to work on methods and theories for a more secure organizational IT ecosystem.

For the Dutch HEIs, as the main stakeholders on the practitioner's side, the gained insights can be an essential part of the problem definition and assessment. In the context of risk-based cybersecurity policy, it is necessary to understand the ways in which shadow IT can cause problems. Research results should enable SURF to suggest policies and approaches for their members and assess whether this problem requires new services from their organization. For industry in general, it should provide a tangible example of how shadow IT risks can be written up, and how specific problems can be combatted. Many of the general insights on IT security management for shadow IT will also be applicable outside of higher education.

1.2 Thesis Outline

This thesis has an extensive, sequential multi-method setup. After this introduction, Chapter 2 presents the research approach, subquestions, threat modelling approach and explains how the different parts of the thesis interact. Chapters 3 and 4 are literature reviews on the

[2] https://www.surf.nl/en

phenomenon of shadow IT and its cybersecurity aspects, respectively. Both start with a methods section before presenting review results and relevant outcomes for the further chapters. The main part of this research consists of interviews with experts from HEI, for which the research setup and results will be presented in Chapter 5. These results will be further analyzed in Chapter 0, which structures the results to gain further insights into how occurrences, vulnerabilities and attack paths relate. The discussion of Chapter 7 evaluates these analyses and combines insights from the interview research with previous chapters to discuss results, limitations and contributions to science and practice. Chapter 8 contains the conclusion and suggestions for future work.

2 RESEARCH APPROACH

The overarching aim of this thesis is to understand in what way the occurrence of shadow IT within HEIs contributes to cybersecurity problems. This is an essential first step for further work on risk-based governance of shadow IT-related threats in these institutions. Although the assumption is that shadow IT is a common phenomenon in higher education, the exact shape it takes had to discovered during the research. We synthesized the occurrences and threats mentioned by experts into an overview of shadow IT occurring in HEIs. Through systematic identification, we created a complete mapping of occurrences to shadow IT threats. To make these threats actionable and point towards areas of risk related to shadow IT, general patterns and possible attack paths were identified. This provides a first step towards the mitigation of risks with insights that are broadly applicable.

2.1 MAIN RESEARCH QUESTION AND SUBQUESTIONS

The research goals above are represented in the main research question (MRQ).

MRQ: *"What is the role of shadow IT in the cyber threat landscape of Dutch higher education institutions?"*

We evaluate the MRQ using five subquestions. The first two aim to describe the phenomenon at hand and the current state of cybersecurity research on shadow IT. This leads to a definition that can be used in the later research steps and an indication of how we can expect shadow IT to be a part of the threat landscape.

SQ1: *"How is shadow IT defined and differentiated from similar concepts in literature?"*

SQ2: *"What cyber threats are commonly associated with shadow IT in literature?"*

The next two subquestions aim to elicit empirical data on the occurrence of shadow IT in HEIs from domain experts using extensive qualitative, interpretative methods. Using concepts from threat modelling (see Section 2.3), we evaluate to what extend problems related to, caused, or worsened by shadow IT are perceived.

SQ3: *"What types of shadow IT are observed in Dutch higher education?"*

SQ4: *"Which cyber treats related to shadow IT are perceived by experts?"*

The last subquestion is explorative and aims to structure the results, linking types of shadow IT and the different threat components as a first step to evaluate possible mitigations of these identified risk factors.

SQ5: *"Which recurring patterns can be identified in these occurrences and cyber threats?"*

2.2 Research Methods

We answer the subquestions of the previous section sequentially, so results from early research questions can serve as input for the research setup of later subquestions. Explanation and justification of the methods precede results for each chapter, in line with this sequential research design.

TABLE 1: OVERVIEW OF METHODS

Question	Method	Results
MRQ	All of the below (2.2)	8
SQ1	Literature review, tertiary study (3.1)	3.2
SQ2	Literature review, mapping study (4.1)	4.2
SQ3	Expert interviews (5.1)	5.2.1
SQ4	Expert interviews (5.1)	5.2.2
SQ5	Synthesis of SQ3 and SQ4 (6.1)	6

2.3 Cyber Threat Modelling

To operationalize the main question, the concept of a cyber threat and how it will be operationalized should be explained. These cyber threats are what the threat landscape in the research question is composed of and what will be used as the ‘unit’ elicited for the intended results. A cyber threat, as defined by the US National Institute of Standards and Technology, is *“Any circumstance or event with the potential to adversely impact organizational operations, organizational assets, individuals, other organizations, or the nation through a system via unauthorized access, destruction, disclosure, modification of information, and/or denial of service.”* (Computer Security Resource Center, 2022).

To aid the modelling of elicited threats, we will use threat diagrams from CORAS, a model-driven risk analysis method by Lund et al. (2011). The CORAS method is composed of eight steps, aimed to develop a complete analysis of a specific subject’s risks, from scoping to making a treatment plan. From this elaborate method and notation language, we will only use the method and notation of *threat diagramming*. This approach is recommended in a systematic review by Tuma et al. (2018) for small-scale projects because of its intuitiveness and tool availability. Besides, it allows for unintentional human threats. This is something we expect to find in the context of shadow IT, where, by definition, the users do not intend to do harm, but where shadow IT might allow vulnerabilities to unintentionally be present. Their threat diagramming method follows the structure of a (1) threat actor (accidental or deliberate) who, via a (2) vulnerability, initiates a (3) threat scenario leading to an (4) unwanted incident with damage done to an (5) asset. These five elements, shown in Figure 1, compose the main structure in their notation and will be the way we operationalize the concept of a cyber threat. A diagram containing all five of these concepts can be traversed to understand what cyber-attacks or incidents can happen. This traversal from threat to asset damaged is called an 'attack path', another central concept to our threat modelling approach. Throughout this thesis, the five different concepts that form a threat model are referred to as 'threat components'.

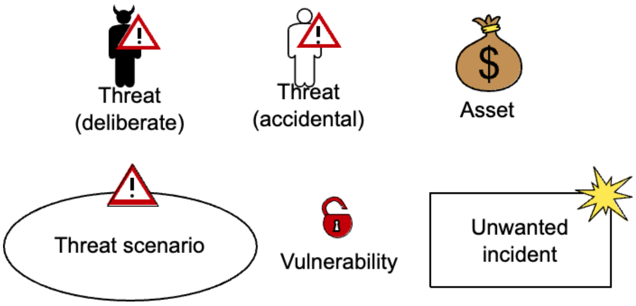


FIGURE 1: CORAS THREAT COMPONENTS

Although the CORAS threat diagramming language also provides components to estimate likelihood and consequence using a quantitative or qualitative scale, this was not the goal of the current research, and these components are not used. In this research, we use CORAS as a source of vocabulary and ontology for threat modelling, without adopting all related methods and techniques as well.

Modelling with threat diagrams should lead to the clear and actionable '*how*?' perspective intended by our research approach. Linking the different steps as attack paths should make the usually abstract notion of 'shadow IT poses a large risk' tangible. Combining these with the identified occurrences provides leads for action: occurrences can be prevented, vulnerabilities mitigated, and threat scenarios detected.

2.4 Limitations and Threats to Validity

Several threats to validity can be named for this research setup. For the sections of this thesis that use systematic literature reviews as a method, validity is dependent on the external validity of the papers selected. To mitigate this risk, we took a conservative approach to selection, for instance, by disregarding all grey literature and feature articles and by working with existing guidelines for study selection, this is elaborated upon in sections 3.1 and 4.1.1.

For the part of the research that uses interview data, an incomplete overview of the threat landscape can be a limitation to research results. By definition, not all shadow IT is known by the interviewees, so the findings on occurrences of shadow IT and the related threats will never be complete. The interviewees perceive only a subset of actual shadow IT present and combine actual observations with suspicions. We think that interviewing experts until code saturation occurs (see Section 5.1.5) is an effective strategy to mitigate the risk of incomplete observations from interviewees. Furthermore, the analysis steps of SQ5 will generalize and summarize results to an abstraction level that should lead to widely applicable insights, where possible missed observations of shadow IT can be anticipated.

Besides this challenge with the data sample, there is also a limitation in the analysis method, where the coding of interview data can be subjective. The identified risks can be influenced by the researchers' frame of reference, causing the coding of data to be less generalizable to the domain as a whole. On the one hand, researcher subjectivity allows for more reflection during the research process, which can lead to a deeper understanding of the topic (Garcia & Quek, 1997). To ensure the researcher's frame of reference did not influence results too much, we coded the data with two researchers, and the codebook was discussed together with a third researcher, both during development as well as after coding. Besides, we assessed the construct validity of codes by calculating inter-coder agreement (ICA) measures throughout the process (see Section 5.1.4).

3 The Phenomenon of Shadow IT

To get acquainted with the research area of cybersecurity for shadow IT and to define the concept, we conducted two literature studies. This chapter presents the first part and serves to get a clear definition and demarcation of shadow IT as a concept. Since there were several recent secondary studies on this topic, a tertiary review approach was chosen, where existing systematic literature reviews (SLRs) were assessed and synthesized. This review answers SQ1: *"How is shadow IT defined and differentiated from similar concepts in literature?"*.

3.1 Tertiary Study Research Protocol

An elaborate and strict methodological framework ensured complete coverage of relevant literature. The selection of electronic databases was shared between the two parts and will be elaborated upon in Section 3.1.1. Sections 3.1.2-3.1.4 continue with the Identification and Selection, Quality Assessment, and Knowledge Extraction and Synthesis steps for this review.

During the preliminary research for this project, it became clear that several recent SLRs on the different aspects of shadow IT already exist in the academic literature. To review these SLRs and extract the relevant definitions of concepts for our purposes, a tertiary review was conducted. This ensured consistent conceptualization throughout the thesis, an essential step in IS (information systems) research, known to lead to validity problems when not performed properly (Lund-Jensen et al., 2016). The specific questions, in line with SQ1, to be addressed by this review are:

- How is shadow IT commonly defined?
- Which parts of an organization's IT are (not) covered by this concept?
- What related concepts exist?
- Which research directions are present for this concept?

3.1.1 Sources to be Searched

For both parts of the literature review, we searched the same electronic databases. Using the overview and selection criteria by Gusenbauer & Haddaway (2020), four services were selected, all with enough features and accuracy to gain the status of 'principal search system' in their comparison. In order to cover some research specific to IS research, AIS Electronic Library was used as a supplementary resource, as recommended by Bandara et al. (2011). Only curated databases of peer-reviewed work that support complex queries and searching specific fields (title, abstract, keywords, *etc.*) were used. Crawler-based search engines like Google Scholar and Microsoft Academic were avoided because they include grey literature and can be inconsistent and limited in their search results (Orduna-Malea et al., 2015). Based on these requirements, we selected Scopus as the primary database, while Web of Science, ACM Digital

Library, and AIS Electronic Library were searched with the same terms to find potentially missed documents.

3.1.2 Identification and Selection

To begin the systematic search for literature, relevant keywords had to be selected. Based on Huber et al. (2016), the following was used as a start: *"shadow system(s)", "shadow IT", "feral system(s)", "grey IT", "hidden IT", "rogue IT", "end user computing", or "EUC"*. These terms were then combined with *"literature review" or "meta-analysis"*. We chose these keywords because their review was one of the few with an explicit list of keywords available to us before we conducted this literature study. Testing revealed that EUC had to be removed both because the abbreviation is used very often in the medical domain for unrelated concepts, as well as never used in relevant literature without also writing the unabbreviated term. All used databases either included the plural of *system* by default or supported the use of wildcards so it could be specified manually. This resulted in the following query, as formatted for Scopus: *TITLE-ABS-KEY(("shadow system" OR "shadow it" OR "feral system" OR "grey it" OR "hidden it" OR "rogue it" OR "end-user computing") AND ("literature review" OR "meta-analysis")).*

Since the goal was to gain a complete but state-of-the-art overview, we only included items from the past 20 years. The main inclusion criterium was that an article had to be a systematic literature review with a defined research question, search process, data extraction and data presentation (Kitchenham et al., 2007), a criterium mostly enforced in the quality assessment, although obvious 'un-systematic' reviews could be disregarded in the first selection phase already. Furthermore, the review had to have shadow IT as one of its main topics and not just mention it.

3.1.3 Quality Assessment

For the appraisal of quality of the selected SLRs, guidelines for systematic review quality from the *Database of Abstracts of Reviews of Effects (DARE)* were followed (Centre for Reviews and Dissemination (CRD), 1995). Their method includes five checks, of which a potential systematic review must pass at least four. We assessed the quality requirements in a lenient way, since gaining a complete picture was important. In contrast to tertiary reviews with quantitative synthesis, differences in setup could be accounted for in the summary. For each check, the way it was assessed is described:

- **Were inclusion/exclusion criteria reported?** It should be clear how many items were identified in the search phase, but not included in the synthesis. For this process to be transparent, numbers should be provided at each selection stage.

- **Was the search adequate?** Several databases should be searched, or the limitations of a smaller scope should be clear. Crawler-based search engines are only to be used for snowballing or when grey literature is explicitly included.
- **Were the included studies synthesized?** There should be some sort of metrics computation or qualitative summarization of results. Enumeration is different from synthesis, so some form of trend or pattern identification between items is necessary.
- **Was the quality of the included studies assessed?** It should be clear to what standard included works are held. Limiting the study to only peer-reviewed works published in selected sources is sufficient for this check.
- **Are sufficient details about the individual included studies presented?** A list of all included studies should be available, and many of them should appear in the synthesis.

3.1.4 Knowledge Extraction and Synthesis

For the qualitative synthesis of the reviews, we turned to Petticrew & Roberts' (2008) three-step method of narrative synthesis, described as follows: *"(i) organizing the description of the studies into logical categories; (ii) analyzing the findings within each of the categories; and (iii) synthesizing the findings across all included studies.*". As logical categories, the research questions underpinning this review, as stated in Section 3.1 were used. For each document, we took notes for each of the categories to gain insights into commonalities and differences between the selected papers. We present the results in Section 3.2.

3.2 Tertiary Study Results

For the first part of the literature review, we found a total of 25 SLRs after disregarding duplicates and papers published in multiple languages. The specification of sources can be found in Table 2. The search was conducted in March and April of 2022. A list of all works, from all stages of this literature research can be downloaded (see Section 8.2). This file also includes notes on inclusion, exclusion, and knowledge synthesis.

TABLE 2: NUMBER OF RESULTS PER DATABASE FOR THE TERTIARY STUDY

Source	**No. of results (of which unique)**	**No. of results after selection and quality assessment**
Scopus	22 (21)	7
ACM Digital Library	3 (2)	0
AIS Electronic Library	6 (2)	1
Web of Science	5 (0)	0

We disregarded fifteen results as irrelevant. In case of doubt if the topic of shadow IT was central enough to the SLR, the paper's possible contributions to answering the guiding questions in Section 3.1 was leading. Papers were removed for the following reasons: Eight papers were either about end-user computing satisfaction (EUCS) or just end-user applications; four were reviews specifically on integrating shadow IT, not the concept itself; two did not have shadow IT as the main topic, just as a sidenote for its results, and the last one was disregarded because it simply had "shadow. It" in its abstract.

After this selection on relevance, ten SLRs were checked for quality as described in Section 3.1.3. Two of the selected works failed the quality assessment, although it should be noted that several failed on the 'transparent inclusion criteria' check. Therefore, we included eight SLRs in our analysis. A complete list can be found in Appendix A: Sources Used in the Literature Reviews.

Looking back, all these results would have been identified with just '*shadow IT*' in the subject part of the query. Especially '*end user computing*' resulted in a lot of irrelevant results. We discovered that this concept is one of the first to be explicitly defined as *not* being related to shadow IT, because research on the topic focusses of a type of system that is officially initiated and supported (Rentrop & Zimmermann, 2012). The fact that a paradigm enables non-IT staff to build their own applications does not make it shadow IT, since it is known, and boundaries are set by the IT department.

3.2.1 The Concept of Shadow IT

Going through the selected SLRs chronologically, an evolution of shadow IT definitions becomes clear. Much of the early work on shadow IT focusses on extensions to big information systems like CRM or ERP systems (Kretzer & Maedche, 2014), often taking the form of spreadsheets or databases. The SLR by Kretzer and Maedche interprets existing definitions purely as 'unofficial IT extending an existing IS' to align with their research on generativity, a property of such IS. They are, to the best of our knowledge, the first to write an SLR with shadow IT as a central topic. They focus mainly on its benefits and downsides, which we will discuss in Section 4.1.2. They use an early definition that reappears in several later SLRs, by Zimmermann & Rentrop (2012): *"The term **Shadow IT** describes business process supporting IT systems, IT service processes, and IT staff. They are deployed autonomously within business departments and by IT users. Thereby, Shadow IT entities are involved neither technically nor strategically in the IT service management of the organization ..."* translated from German by Zimmermann et al. (2014). The literal translation from German requires some extra emphasis: the 'business process supporting' refers to the IT systems, and 'organization' refers to the central IT department. Although Zimmermann et al. contributed much of the early work on shadow IT, fast developments in enterprise IT kept adding new forms of IT to the enterprise, calling for clearer delineations to properly structure observations in the following years.

Shadow IT as an area of research seems to mature with the work of Kopper & Westner (2016b), whose taxonomy provides a clear demarcation of the concept, which was the basis for many later papers (e.g., Furstenau et al., 2017; Klotz et al., 2019; Magunduni & Chigona, 2018)). Although later extended and refined, this taxonomy does not have fundamental problems, and many later definitions can be fitted to it. Their main contribution is a concept hierarchy shown in Figure 2. It defines '**feral practices**' as the overarching concept, covering both misuses of official IT (**workarounds**) and unofficial IT. The research streams on unofficial IT are split on perspectives. Research on the creation of unofficial IT artefacts is called **un-enacted projects** or **shadow sourcing**, where the latter covers sourcing done by an individual user. The main perspective taken in research up until then was on the outcomes, as opposed to the sourcing, of artefacts, where 'shadow IT' covers both devices and applications, and '**shadow systems**' only applications, with a focus on bigger information systems like CRM or ERP systems. Besides a taxonomy, Kopper & Westner (2016a) also provided one of the first state-of-the-art analyses into the causes, consequences and governance of shadow IT.

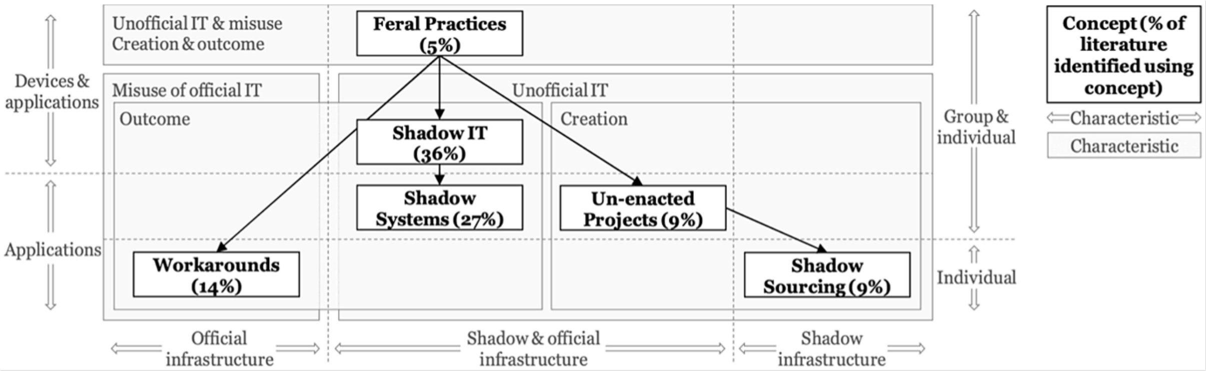


FIGURE 2: TAXONOMY OF SHADOW IT (KOPPER & WESTNER, 2016A)

An important development after this taxonomy, driven by the work of Kopper (2019), was an increased focus on the concept of **business-managed IT** as deliberate shadow IT by business units. This is clearly described in a later SLR by Klotz et al. (2019), who separate the previous notion of Shadow IT as follows: when instances are covert, it is called shadow IT, and when it is overtly managed by business units, it is called *business managed IT*. This distinguishes a more decentralized type of IT management that can be in alignment with the business and IT organization. This was an important addition since, before the concept's introduction, the definitions did not allow for IT that was both overt and managed by any other business unit besides the corporate IT department.

Besides the increased role of decentralized IT, other trends in society called for the incorporation of new concepts into the definition of shadow IT. Although self-made solutions like spreadsheets were still present, new trends emerged. Increased adoption of Software as a Service (SaaS) via cloud-solutions and self-acquired devices has become more accessible. This led Mallmann et al. (2019) to suggest a new typology of shadow IT consisting of a division into four types, see Table 3. This division of the concept makes it tangible and can likely be used when working on classifications of shadow IT that aid governance.

TABLE 3: SHADOW IT TOPOLOGY BY MALLMANN ET AL. (2019)

Unapproved cloud services	Use of Internet-based Software and Software as a Service (SaaS) that are not approved or unknown by the IT department. These systems are also called Mobile Shadow IT once they can be accessed outside the workplace.
Self-made solutions	Use of solutions developed by employees on the company's computers to perform their work tasks. For example, an excel spreadsheet or an application developed by employees.
Self-installed applications	Use of software installed by employees to perform their work tasks on the company's computers. For example, downloading and installing software available free of charge on the internet.
Self-acquired devices	Use of devices owned by employees. These devices are purchased directly from retail rather than being ordered through the official catalogue of the IT department. It includes the use of applications in the employee's personal devices at the workplace.

From all SLRs identified, the impression emerges that after 2014 many new research directions were explored around what Käss et al. (2021) generalize as 'IT outside the IT department'. According to them, researchers and practitioners have created multiple concepts for the various aspects and trends, creating an overlap. At the same time, no single concept has become dominant, although both taxonomy-developing SLRs identify 'shadow IT' as either the most or second most common term (Käss et al., 2021; Kopper & Westner, 2016b). In an effort to delineate the newly emerged concepts once more, Kass et al. worked towards a taxonomy that was still in line with Kopper and Westner's, but with 'IT outside the IT department as a broader overarching concept. Apart from the already mentioned shadow IT and business managed IT (BMIT), they included Lightweight IT, IT Consumerization and BYOD as well. Since a clear distinction will be important for the rest of this thesis, we shortly discuss them.

- **Lightweight IT**, a concept introduced by Bygstad (2015), can be viewed as the idea of subsuming business-managed IT; a positive view on what smaller-scale IT initiatives undertaken by business units can contribute. Lightweight IT supports front-end work and aims to enable quickly adapting new innovations to later integrate them with heavyweight IT (Godefroid et al., 2021).
- **IT consumerization** is delineated as the trend where consumer-oriented devices, applications and services are used for business purposes (Yan et al., 2016). Although it can lead to Shadow IT, it mainly refers to the trend and enabling/causing factor, not the specific devices or applications that can be observed in an organization.
- **BYOD** (Bring Your Own Device), according to Käss et al. (2021), is a specific form of IT consumerization where employees bring private devices to work instead of using company-owned devices (French et al., 2014). It is solely about hardware, and the term is mostly used to describe the policy that allows the overt use of these devices. While the policy in place then actually acknowledges and allows these devices, this can still lead to (more) shadow IT in the form of unapproved applications or peripherals.

Although many of the reviewed SLRs include a problem statement indicating a lack of uniformity in the terms used, it is certainly possible to see a consensus. While a multitude of directions is explored, with varying success, the overall definition stays consistent. The later distinction between shadow IT and business-managed IT introduced by Kopper et al. (2018) seems useful to indicate a situation where something that would be regarded shadow IT by previous definitions is overtly managed by a business unit other than the central IT department. From this review, we can conclude that the definition of Haag & Eckhardt (2017), which is mentioned in all but one of the SLRs written after it, covers the concept best:

> *"[Shadow IT is] hardware, software, or services built, introduced, and/or used for the job without explicit approval or even knowledge of the organization.".*

3.2.2 Current Research Directions

As summarized above, the identified SLRs often focus on the conceptual definition and taxonomy of the shadow IT phenomenon. Four out of eight papers had taxonomy development or a comparison of concepts as their main topic. This might not represent the whole research area of shadow IT since a literature review is especially suitable for this research direction. Besides, some literature reviews specifically on the topic of shadow IT integration, most notably Huber et al. (2017), were not selected for our review since their goal was too specific for answering our research question. This omitted line of work on shadow IT integration could also be seen as one of the main research directions.

In the remaining body of research three themes encompassing three research directions are often mentioned: Causal factors, outcomes, and governance. The categorization and review of these themes by Klotz et al. (2019) seem to be the most extensive and state-of-the-art available. Preceding work by Kopper & Westner (2016a) uses similar categories, only without explicitly incorporating business-managed IT. Other works by Kretzer & Maedche (2014) and Rakovic et al. (2020) take only one or two of these themes as their lens for reviewing. For this reason, Klotz et al. (2019)'s categorization will be summarized below, as it serves to guide further discourse in this thesis. **Causal factors** can be subdivided into enablers, motivators, or missing barriers, with motivators as reasons to introduce/use shadow IT seeing the most scholarly attention. In general, causing factors to have been widely studied. Especially at the level of the individual, there is a clear understanding of why employees would use shadow IT. The **Outcomes** theme is subdivided into benefits and risks/shortcomings. With security risks being the most prominent theme on the risk side, this will be the focus of the second part of our literature study in Section 4.2. The last emerging theme is **Governance**, which has seen the biggest increase in research interest in the last few years. Klotz et al. interpret this as a sign of progress in the research field. This research is either about the governance of uncovered Shadow IT instances or general governance, where the latter is split into policy setup, awareness training, IT systems gap resolution, and monitoring and identification. An interesting choice is made by them, classifying all uncovered shadow IT as 'business-managed', not leaving a category for shadow IT that is uncovered but not maintained by a business unit.

None of the reviews indicates any dominant theories that are used to analyze shadow IT occurrence. Although transaction cost economics, actor-network theory, the technology acceptance model and configuration theory appear multiple times throughout the literature, there is no dominant approach (Klotz et al., 2019; Kopper & Westner, 2016a). Much of the conducted research has an emphasis on observation and data gathering on a small scale such as case interviews without much attention to analysis or theory building (Klotz et al., 2019).

3.3 Takeaways for the Next Chapters

For the next literature review and the empirical research to follow afterwards, there are a few takeaways. Firstly, a widely adopted and complete definition was found. This definition by Haag & Eckhardt (2017) will delimit the scope of the next research phases. It can be checked with the interview participants to ensure conceptual consensus and to report on the applicability of this definition in the higher education sector. The work on topology and definitions provided the conceptual understanding necessary to keep the focus on actual shadow IT. Similarly, the topology of Mallmann et al. (2019) can serve as vocabulary to distinguish between different types of shadow IT.

Three concepts that kept showing up at the edges of the concept definition need further elaboration to not blur the scope later: spreadsheets, BYOD, and business-managed IT. Although we could disregard spreadsheets as 'just a file' or misuse of official IT in the case where spreadsheet software is provided by the institution, we follow Mallmann et al. (2019), who view it as a self-made IT solution. Although often simple by today's IT system standards, spreadsheets can be a versatile tool in the hands of some employees. For BYOD, we had to decide whether to include it or not in our research. We expect that many of the experts face similar problems in dealing with BYOD-devices even though they are permitted. We will therefore interpret these cases as the *'IT [...] not approved by the organization'* part of the definition. We will not take note of every BYOD case that comes up, but when security problems arise because of unapproved actions on the devices, it will be considered 'in scope'. The same will be done for business-managed IT if (and only if) there is no clear governance system in place. We will ask each interview participant how the roles between central and decentral IT are distributed. If, within an institution, there is a clear policy on what the departments can and cannot do themselves, the central IT department could be seen as 'in control' of these systems.

The last two choices are part of a pragmatic approach. Even though devices brought under BYOD policy or systems managed by other business units could be seen as self-inflicted vulnerabilities, we expect the challenges for IT professionals and the consequences for the organization to be similar to those caused by shadow IT brought into the institution without BYOD policy or other management layers' approval. Therefore, it is deemed relevant in answering our research question on what kinds of threats are perceived.

4 Cybersecurity Aspects of Shadow IT

This chapter describes the second literature study, a systematic mapping study into cybersecurity aspects of shadow IT. Using the definition and demarcation from the first review, we gathered and summarized existing research on cybersecurity aspects of shadow IT to create a clear description of the state of the art of cybersecurity research focusing on shadow IT. This review answers SQ2: *"What cyber threats are commonly associated with shadow IT in literature?"*.

4.1 Mapping Study Research Protocol

A mapping study, as compared to a 'normal' systematic literature review, has a broader focus with more qualitative search and classification procedures (Wohlin et al., 2012). They are a good starting point for initiating research since it creates an overview and understanding of existing research (Budgen et al., 2008) and can identify areas for more primary studies to be conducted (Kitchenham & Charters, 2007). For this mapping study, we limited ourselves to research on cybersecurity outcomes and consequences, which is the focus of this thesis. Causal factors and governance were left out of scope at this stage. For this chapter, the following questions specifying SQ2 guide the review:

- What kind of cybersecurity risks are associated with shadow IT?
- What kind of empirical work has been done in this field?

4.1.1 Identification and Selection

The electronic databases described in Section 3.1.1 were searched with the shadow IT keywords identified in Section 3.1.2, combined with the terms *"security, cybersecurity, or risk"*. Based on the results of the tertiary study, we removed '*end-user computing'* and '*EUC*'. In addition to these identification steps, this literature review identifies both works cited by, as well as works citing the identified works for additional papers, a process known as 'snowballing' (Skoglund & Runeson, 2009). Finally, the SLRs from the tertiary study were scanned for relevant cited papers since some of them mention security aspects explicitly.

For the selection, the main inclusion criterium was that a study should contribute new insights into the security consequences of shadow IT. We assessed this criterium by reading the titles and abstracts and scanning the full text. Limiting to specific dates was not deemed necessary, nor did we add other exclusion criteria since the goal of a mapping study is broad coverage, not narrow focus (Kitchenham & Charters, 2007). For the same reason, we performed no quality assessment.

4.1.2 Knowledge Extraction and Synthesis

As noted by Kitchenham & Charters (2007), data extraction for a mapping study is mainly about classification and categorization. All included papers were read to identify which ones were suitable for answering the research questions. We followed the same procedure as for the tertiary review (see Section 3.1.4), where the questions from Section 4.1 served as logical categories.

4.2 Mapping Study Results

For the second review on cybersecurity aspects of shadow IT, the initial literature search resulted in 116 works, of which 13 were duplicates. While collecting these, 44 could immediately be disregarded because they unintentionally showed up in the search results with, for instance, a sentence ending with "shadow." and the next one starting with "It". These are not counted in Table 4. Of the remaining 59 unique documents, three articles were disregarded because they were feature articles, i.e., not peer-reviewed scientific literature. For seven articles, a full-text version could not be sourced. Thirty-seven works were excluded because they did not contribute any new insights into security problems of shadow IT. Often, they only referred to existing evidence, as will be elaborated upon in Section 4.2.1. This leaves us with 12 selected articles. From (one round of) backward and forward snowballing, three more relevant papers were added. We present an overview of the results in the different databases in Table 4. A list of all works resulting from the initial search, including notes on inclusion and exclusion reasons, and notes used in synthesis, can be found in the attached data repository (see Section 8.2). The list of *selected* works is in Appendix B: Data Management Plan. The search was conducted in April 2022.

TABLE 4: NUMBER OF RESULTS PER DATABASE FOR THE MAPPING STUDY

Source	No. of results (of which unique)	No. of results after selection
Scopus	51 (51)	12
ACM Digital Library	4 (2)	0
AIS Electronic Library	1 (0)	0
Web of Science	16 (6)	0
Snowballing	3	3

4.2.1 Problemization of Shadow IT

There is little doubt in the literature about the fact that shadow IT poses a cybersecurity risk to organizations. Introductions of papers on the topic present shadow IT in terms ranging from *“one of the major IS security challenges faced by companies”* (Györy et al., 2012) to *“bringing unprecedented security risks”* (Silic & Back, 2014). There are also some empirical records of the relative size of this concern, with, for instance, Kopper (2017) finding that security issues are the most often mentioned risk of shadow IT by IT managers, with 88% of them mentioning it, the highest percentage in their comparison. In the SLR by Klotz et al. (2019), it is also the risk with the highest relative representation in the literature. **There is no apparent doubt about the fact that shadow IT can cause security problems to an organization, but it is not always as clear *why* or *how* exactly.**

Although the goal of this scoping study is to discover what cyber threats of shadow IT are mentioned in the literature, one of the most interesting findings is that very few are mentioned. While assessing each work for the inclusion criterium of ‘contributing insights to security consequences of shadow IT’ (see Section 4.1.1), it turned out that only a small number of papers worked on specifying the problem. Two patterns emerge: firstly, the reviewed literature shows a high reliance on grey literature and a few scientific sources as evidence for the fact that shadow IT causes problems for organizations. Secondly, most works cite the same one or two sources for their statement that there are serious security consequences of shadow IT instead of explaining the problem. Both patterns around defining the threats related to shadow IT will be elaborated upon in the following paragraphs.

The first finding in this regard is that papers, especially early works, often rely on grey literature to illustrate the possible problems of shadow IT. Johnson (2013) and Walters (2013) are often mentioned. They independently wrote about the security of shadow IT in two feature articles, in the same journal, with the same title, that describe potential problems. Both take a policy perspective, describing the essence of the problem as *“managing a network when you cannot control the applications and devices it connects”* and *“[CIOs not being] in charge of appraising the security of all applications hardware that are used to process and store corporate information.”* respectively. Silic & Back (2014) take a more extensive approach to the usage of grey literature and perform a review of practitioner reports. Their early work was in line with these two quotes, concluding that opening up the network to malicious users and data integrity or exposure issues were among the main problems caused by shadow IT. The reports they reviewed suspect that with the increase of ‘tech savvy’ users and mobile devices, the problem will become larger in the future. Although their approach is thorough, they do not go in-depth into technical consequences, and we expect that much has changed in any organization’s IT landscape since 2014.

A second pattern can be seen where many works rely upon the same sources. Six of the fifteen papers reviewed cite Györy et al. (2012) for their claim that security problems exist, although it is not the paper's main focus or contribution. Györy et al. take the presence of unapproved devices as the essence of the problem, which is understandable since it was a very new development at the time and had many unknown consequences. In their paper, non-compliant behavior is seen as the main problem without getting into technical details.

This idea of non-compliance as a threat in and of itself seems to be dominant and is related to the belief that existing security policies that should prevent incidents and the lack of control are evaded (Rakovic et al., 2020). Phrased differently: the usual security controls are absent (Magunduni & Chigona, 2018). Both these sources also mention the fact that the users or maintainers do not have the right knowledge about security. Besides, without central control, end-users suddenly become responsible for taking adequate security measures (Györy et al., 2012).

The papers that express possible problems explicitly, often do this on the governance level, with the 'lack of professional quality assurance or responsibility for maintenance' as the main problem (Zimmermann & Rentrop, 2014). Furstenau et al. (2017) quote an interviewee mentioning the problem where the quality assurance is performed by just *'another freelancer responsible'*. Behrens (2009) describes this as a low *'hit by a bus factor',* were the ad-hoc nature of shadow IT systems causes high reliance on a few individuals. If these individuals left the organization, nobody would be responsible for the product. This governance risk is not directly related to security, but it could lead to, for instance, unpatched systems. Rakovic et al. (2020) give the example of outdated servers and software versions.

Still, almost no literature describes specific and concrete problems. More often than not, *'IT is not being integrated or compliant with existing security protection'* or *'[shadow IT is a] violation of formal security policy'* (Chua et al., 2014; Haag, 2015) is the level of abstraction at which security problems are described. Although the above-mentioned non-compliance and governance problems can contribute to security issues, the lack of specificity or solid empirical evidence is problematic if we want to mitigate these threats.

To summarize the findings of this literature review, both scholars and practitioners see shadow IT as a cause or complicating factor of risks to IT security, with unauthorized network access and data exposure as common examples of problems. However, empirical evidence is limited and mostly anecdotal. Furthermore, factors like bad governance and non-compliance are often the end of risk analysis, while this could be the beginning of a threat assessment. Technical or detailed descriptions of cyber threats enabled by shadow IT are not present in the literature identified by this review.

4.3 Takeaways for the Next Chapter

While performing this review, it became clear that for our overarching goal of discovering the role of shadow IT in HEI's threat landscape, not many specific findings from the literature can be used. This means that instead of testing existing findings in a new context (HEIs), we will mostly be gathering new data. We will conduct interviews without expectations about results, eliciting threats from practice, and structuring these findings. This inductive approach will be central to the next chapter.

Furthermore, these results re-emphasize the added value of applying a threat modelling language to this topic. The focus on governance-level problem definitions in existing literature, (e.g., *"non-compliance is the main problem"*) leaves out details and expected consequences that cannot be missed when a structured modelling approach is taken.

To give an example of how we expect threat modelling can help specify the actionable, practical implications of shadow IT, modelling 'non-compliance as a problem' in CORAS, could lead to a specification such as: *"Non-compliance for the usage of statistics-software is a problem because the used shadow applications do not integrate with our single-sign-on solution, causing researchers to create a new account. If they then choose the same password, that could lead to a compromised institution account if leaked or badly stored."* This example could even be developed more, to then continue with the assets damaged when credentials are compromised.

5 Shadow IT in Higher Education Institutions

With the third and fourth sub-question of this thesis, we aimed to identify and evaluate shadow IT occurrences through expert interviews. It is an interpretive, bottom-up approach where threats are identified through iterative analysis and coding. We chose this approach because it is open-ended and deemed suitable for a study that has as a primary goal to gather new observations from practice. This interview research answers SQ3 and SQ4: *"What types of shadow IT are observed in Dutch higher education?"* and *"Which cyber threats related to shadow IT are perceived by experts?"*.

5.1 Interview Research Protocol

The interviews were semi-structured and designed to answer the subquestions using the lens of (CORAS) cyber threat models from Section 1.1. The interviews elicit as much input on occurrences and possible cyber threats of shadow IT as possible to enable the coding of emerging topics and patterns. Interviews were recorded and transcribed, with recordings removed immediately after transcription.

To adhere to privacy regulations and ethical standards and enable the re-use of the data for later research, the research setup was carefully thought out and reviewed. The complete research data management plan (DMP), informed consent form, and interview protocol can be found in Appendix B: Data Management Plan, Appendix C: Informed Consent Form, and Appendix D: Interview Protocol respectively. This research protocol was approved by the Science - Geo Ethics Review Board (SG ERB) at Utrecht University under reference #Bèta S-22770.

In this thesis, we refer to institutions by a number, e.g., Institution 04 or Expert 04. These numbers were assigned randomly after all interviews were conducted. Experts gave additional consent to be quoted literally.

5.1.1 Interview Subjects

The subjects of this study are security experts from Dutch HEIs. For the sample, we aimed for an approximately equal number of experts from the *HBO* sector (universities of applied sciences or higher vocational education) and the *WO* sector (research universities). For a description of the Dutch higher education system, we refer to de Haan (2014). SURF has 50 affiliated institutions from this sector, representing all publicly funded HEIs. These institutes differ greatly in the number of students, the number of staff members, research focus or approaches to IT governance. For choosing the total number of interviewees, a frequently cited analysis by Guest et al. (2006) suggests that in qualitative research, most codes are discovered within six interviews, with almost no new codes identified after twelve interviews. Based on a more recent comparison in IS research by Marshall et al. (2013), the lower bound of this estimate is expected to be too low. Justifying sample sizes in qualitative IS research is essential for scientific rigor but is often neglected (Marshall et al., 2013). To combat this issue, we conducted new interviews until data saturation occurred. This concept will be further explained in Section 5.1.4.

Since the goal is to have as complete an overview as possible, we aimed for a high sample variety. We took a mixed method approach to this extent, combining maximum variation purposive sampling and convenience sampling (Verhoeven, 2019). Convenience sampling, since the interviewees were among those active in knowledge-sharing communities at SURF[3] or known directly by them. Purposive sampling, since we aimed to represent as many variations in institution properties as possible. We selected interviewees from institutes with or without a focus on technical subjects, big and small institutes, big cities and peripheral areas, and as many permutations of these variables as possible.

We selected interviewees from both the policy side of cybersecurity practice (such as CIOs and CISOs) as well as the more practical side (such as SOC analysts and CERT members)[4]. We know these functions are often combined, especially in smaller HEIs. Again, we aimed for an equal distribution over these categories and a high variety of characteristics within these categories.

5.1.2 Protocol Development and Execution

Before establishing the interview protocol, we interviewed two experts in a less formal setting to check if the topics were clear and well-targeted to the interviewees. Additionally, two pilot interviews were conducted with researchers from Utrecht University and SURF to ensure the questions aligned with the research goals. To this end, guidelines from Castillo-Montoya (2016) were adopted as follows: Alignment between interview questions and research goals were made explicit to check their coverage, topics were introduced before asking the main questions, and specific probes were prepared to steer the participants towards the right level of abstraction, asking them to make a risk more practical and elaborate upon the attack path they hypothesize to be possible. Preparing these probes ensured that the interviewer did not have to give many examples, possibly steering the participants towards specific answers. Apart from the questions directly related to occurrences and cyber threats of shadow IT, we asked

[3] https://surf.nl/en/security-communities-working-together-on-security-and-privacy

[4] CIO = Chief Information Officer, CISO = Chief Information Security Officer, SOC = Security Operations Centre, CERT = Computer Emergency Response Team

the participants to provide background information on their working experience and the size and governance structure of their institution.

Interviews were conducted in Dutch to remove possible language barriers for the participants since this is the language most of them spoke in their daily work. We used Microsoft Teams for the interviews because of its widespread adoption in this sector and to facilitate the easy recording of the interviews with data privacy ensured, since Utrecht University has a data processing agreement with its vendor.

Interviews were transcribed verbatim, only removing filler sounds, repeated words, and interviewer encouragements. The audio was first transcribed using Office 365's built-in transcription software (Microsoft Corporation, 2022), one of the few tools of its kind supporting Dutch. Although these tools are constantly improving, manual correction was still necessary. In two of the interviews, connection issues occurred. Comments on those problems were removed from the transcript since they were irrelevant to the research.

5.1.3 Data analysis

Discovering cyber threats and their relation to the different occurrences of shadow IT required an approach closely resembling grounded theory building. Therefore, the open and axial coding paradigm was followed using ATLAS.ti (ATLAS.ti Scientific Software Development GmbH, 2022a). For occurrences of shadow IT, the coding was limited to just coding the distinct types of shadow IT mentioned. For coding the perceived threats, the threat modelling language of Section 2.3 determined some pre-set code categories used as mutually exclusive semantic domains (ATLAS.ti Scientific Software Development GmbH, 2022b). Although CORAS also provides a language for the scaling of risk, this was not done within our scope. Therefore, we only used the most central concepts of *Threat, Vulnerability, Threat scenario, Unwanted incident,* and *Asset.* An example of these categories can be seen in Figure 3. Since these categories link the codes with a framework of concepts that give meaning, that part should be viewed as axial coding. CORAS should explicitly not be viewed as a theory here but as means to operationalize the concept of a 'threat'. The chosen lens does not influence the identified threats, only how they are dissected and written down. The same goes for the

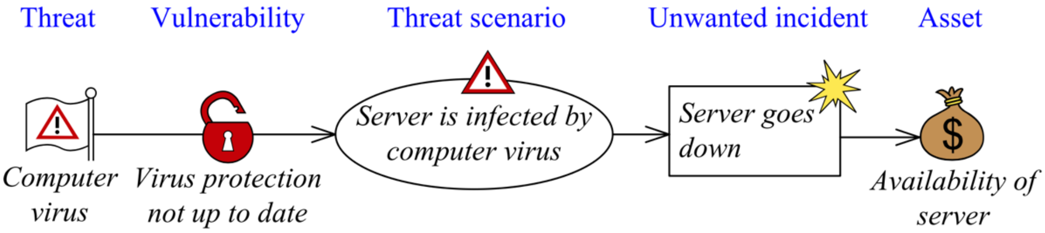


FIGURE 3: CORAS THREAT DIAGRAM EXAMPLE BY LUND ET AL. (2011)

shadow IT topology of Mallmann et al. (2019). We used this after the open coding phase to structure the found occurrences. During the creation of code categories, we made sure the different codes could be applied just as well with a different structure. The definition and demarcation of shadow IT from Section 3.3 functioned as a filter, using which we left out some observations by experts that did not fit the current research's goals. Quotes within the text were allocated to lower-level concepts, which we then assigned to the categories, or added as additional emergent codes or category.

After the transcription of interviews, the primary researcher developed an initial codebook by coding two interviews. To ensure rigor, a multiple coding approach was used. Barbour (2001) describes the essence of multiple coding as *"the furnishing of alternative interpretations and alerting the researcher to all potentially competing explanations"*. To achieve this, the first version of the codebook was shared with a second researcher, who then coded the same two documents without knowledge of the quotes (selected text) or codes the primary researcher chose. Krippendorff's alpha binary metric was used to assess the agreement in sections to be coded (Krippendorff, 2011). This metric gives a general overview of whether two different coders identify the same sections relevant to the codes. For the first two documents, this metric was 0.253, a low score. After further discussion and modifications to the codebook with a third researcher, this score increased to 0.334 for the subsequent two documents, again coded separately by the two researchers. We will discuss the consequences of these scores in the next section, but because of the low agreement, the remaining documents were also coded by two researchers and then combined after discussion.

5.1.4 EVALUATING ANALYSIS ACCURACY

The initial ICA can be interpreted as relatively low. The main reason for the low score was a misunderstanding about the difference between an actor in CORAS (the entity performing threat actions) and the actor that acquired the shadow IT occurrence. Furthermore, there were inconsistencies in applying a code multiple times when the shadow IT occurrence was mentioned more than once. These differences in interpretation could quickly be resolved during the discussions. Another explanation for the low ICA is in the chosen approach, where quotes (which pieces of text to code) were not provided to the second researcher, and the codebook was subject to much change during the first four documents. The evaluation of accuracy differed for the different semantic categories. For the concepts that relate directly to SQ3 and SQ4, the identification of shadow IT and the related threats, a more rigid approach was chosen, where the goal was to find agreement and not allow for multiple interpretations of what the interviewees mentioned.

During the interviews, it turned out that most experts had an interesting opinion on how the definition, causal factors, and governance of shadow IT are different in this sector compared to ‘normal’ companies, so it was decided to add codes for this, even though it was not part of the research questions. Because the goal here was not to come to an unambiguous interpretation but to gain input for Chapter 0, which structures the observations for interpretation and later discussion, we focused on a broad retrieval of ideas. ICA was not calculated for these categories. In the project file linked in Section 8.2, these are grouped under ‘SQ0’.

Because some segments were coded double, intended calculations for inter-coder agreement could not be performed. Overall, the interpretation of code consistency was mixed. Although the discussion after coding seldom led to disagreement and differences in coding could easily be explained, it became clear that CORAS was open to different interpretations. This adaptability to different contexts is a known property of the method (Lund et al., 2011), but it is also a downside, especially for the coding phase. For example, the exfiltration of data by a malicious actor could both be interpreted as a threat scenario or an incident. The interpretation depends on whether you view the action as a means to an end (data will be used to extort, send phishing, etc.) or as direct damage to an asset, such as the confidentiality of student data. In the discussion of results, we can easily make this nuance, but for coding, it simply leads to lower agreement. This challenge will be discussed in more detail in Section 7.5. Although the remaining documents were coded in parallel by the two researchers, all documents were checked and discussed after, together with a third researcher, to ensure consistent application of the different codes.

5.1.5 Evaluating Code Saturation

To evaluate the thematic saturation of our codes, we employed a method by Guest et al. (2020). They propose a quantitative evaluation where thematic saturation is computed as the proportion of new codes in a run of several interviews compared to the codes found in a base set of interviews. Following their recommendations, we took four interviews as the base before evaluating and then computed saturation for each run of three interviews. This base size was chosen for efficiency purposes; Guest et al. (2020) showed that taking a larger size does not effect the evaluation. The larger run size of three is slightly conservative, chosen because we expected significant differences between the contribution of new threats by the different types of interviewees. For the same reason, and since the goal was to gain a complete understanding of the threat landscape, we chose 5% as the threshold of new codes in a run before stopping.

Surprisingly, the desired saturation was reached quickly. The proportion of new codes in the first run of three after the four base interviews was already very close to the threshold for occurrences (5.9%). At the same time, for the threat modelling concepts (21.6%), a significant number of codes were still added. After the second run (a total of ten interviews), we achieved the desired threshold with a new code proportion of 0% (no new codes) for occurrences and 2.7% for threats. One more interview was conducted afterwards because it was already scheduled before the saturation was calculated. We can, therefore, confidently state that the number of interviews conducted was enough to gain a complete overview of the emerging themes within the interviews.

5.2 Interview Results

The eleven interviews all took between 40 and 70 minutes. A total of fifteen prospective participants were approached for the eleven interviews. We found an alternative expert from a similar institution if the intended experts was not available for an interview. The reasons for not participating in the research were all related to time and workload. Institution 11 initially declined because they said they did not have any shadow IT, but they were included in the research, and an interview was conducted, nevertheless. Since the final sample aligns with the initial intentions, we have no reason to suspect sample bias and are confident that we achieved the intended high sample variety necessary for a relevant overview of shadow IT in HEI.

In total, the eleven interviews resulted in a codebook of 104 codes: 18 for types of occurrences of shadow IT, 45 for the different components of threat modelling, and 41 codes on miscellaneous themes. The latter are not directly related to the research questions but contain the results of open coding on themes such as causes of shadow IT in HEI, governance approaches, technical mitigations, and indications of severity. As explained in Section 5.1.4, these codes were not restructured, combined, or discussed in detail by the coders and only served as input for an explorative section on possible treatments and governance (see 0 and 6.2.2). The complete codebook, including all descriptions of codes can be found in Appendix E: Codebook.

The 63 codes directly related to the research questions were applied 448 times and will be the focus of this chapter. We sort the occurrences by the number of times it was mentioned (a metric called 'groundedness' in ATLAS.ti), but we want to emphasize that this should not be seen as an indication of scale. Although the groundedness is helpful in providing some structure, the goal here is not to quantitatively assess the occurrences, so this number should just be viewed as an indication of how top-of-mind the type of shadow IT is within the group of security experts. Besides groundedness, we report how many different experts mentioned a specific concept. For example, 'Spreadsheets and databases (15/6)' was mentioned fifteen times by six different experts. In this chapter we introduce each code in bold once, while in the rest of the thesis quotations are used.

We support our findings with literal quotes wherever possible. These quotes were translated by the researchers from Dutch, while the original can be read in the project file, linked in Section 8.2. Direct quotes could not always be included since some of the interviewees did not consent to this, or the interviewee confirmed a statement rephrased by the researcher. In these cases, we just mention the expert. Since the transcription was partly done automatically, punctuation errors were still present in the transcripts. We corrected these in the translations to make the quotes more readable. In the pseudonymization steps taken, we redact the systems used by the institution by replacing them with [official system] or [shadow system], including their functionality if relevant. This data minimization was done because knowing the system name is not needed, and the used technology could be sensitive in terms of security, or for their supplier management. We took a similar minimization precaution for mentions of a subunit in the institution. Since specific departments as well as the general term used for a faculty, institute, school, etc. can identify an institution, we redacted these and replace them with [department].

5.2.1 Occurrence of Shadow IT

What kinds of shadow IT do you see within your institution? What kinds do you think are used the most? Is it hardware and/or software? How essential are these occurrences to the business? By whom is it used, at what level in the organisation? What would be the shadow IT hardware/software/service here?

For the coding of occurrences, we distinguish between the different types of software, hardware, and its combinations. For determining the codebook after the open coding phase, we held the role of shadow IT for the cyber threat landscape in mind. If devices or software require similar approaches in their cybersecurity management, having them in a single code would be useful. This is, for instance, why 'Locally installed applications' is a single code. Although it might matter for motivators what an application's functionality is, it does not matter as much for the security consequences since all locally installed applications *could* roughly do the same on a computer. Sometimes one code is a more specific version of the other, for instance: 'Leaked credentials' is a specific version, with specific implications, of 'Data breach'. If it was made explicit by the expert, we applied the more specific code.

As mentioned in section 5.1.3, the shadow IT topology of Mallmann et al. (2019) could be used to, retrospectively, determine a structure for the occurrences of shadow IT. This was possible without changing the topology or coding, except when the type of software was not specified. Besides, Mallmann et al. include *applications* installed on personal devices in the 'Self-acquired devices' category (see Section 3.2.1). This distinction where an application would be installed could not always be made in our data and self-installed software was discussed regardless of the device it ran on. Moreover, there were plenty of cases where the hardware might have been procured officially, but what was done software-wise was still perceived as unwanted shadow IT. To make the topology more suitable to our situation and cybersecurity perspective, we renamed 'Self-installed applications' to 'Self-acquired applications', to allow for software for which it was not known *if* and *where* it was installed. This also removed the distinction between shadow software installed on personal or work devices. Although that distinction would matter for the treatment of shadow IT (see Section 0), it could not be made in the transcripts. We will use this adapted categorization to present results in the remainder of this thesis.

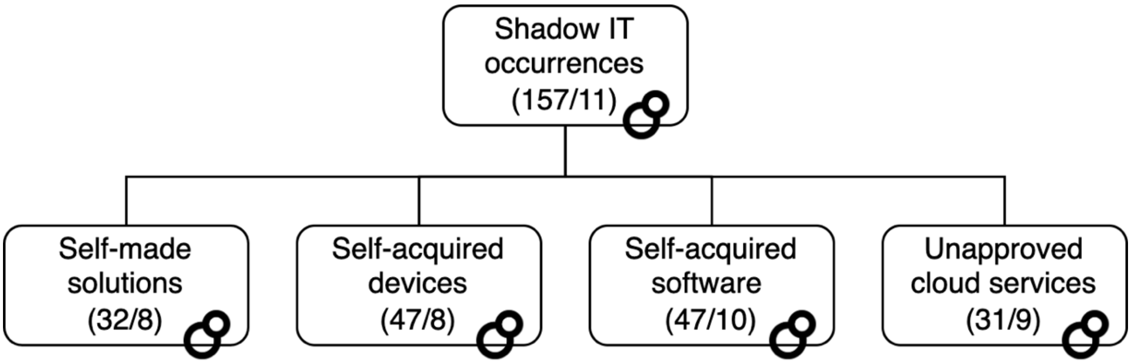

Shadow IT occurrences (157/11)
Self-made solutions (32/8)
Self-acquired devices (47/8)
Self-acquired software (47/10)
Unapproved cloud services (31/9)

FIGURE 4: THREAT COMPONENTS WITH GROUNDEDNESS AND COVER METRICS

Figure 4 shows the top-level distribution of occurrences of these categories. With each expert, a shared definition of shadow IT was agreed upon beforehand. Still, it is noticeable that not all experts observe occurrences from each category. Besides Institution 11, who only observed usage of some cloud services, all institutions observe occurrences from the different categories. The institutions *not* observing IT in a category change from category to category.

5.2.1.1 SELF-AQUIRED SOFTWARE

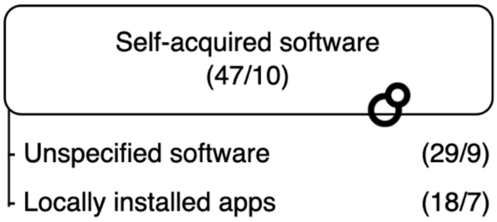

Self-acquired software (47/10)
Unspecified software (29/9)
Locally installed apps (18/7)

FIGURE 5: OCCURRENCES OF SELF-ACQUIRED SOFTWARE

Software was more prominent than hardware in the occurrences mentioned by the interviewees, accounting for 109 out of the 153 occurrences coded. **Unspecified software** and **Locally installled apps** were the most applied software codes in the data, followed by the different cloud apps which we discus later. Figure 5 reports the groundedness and cover for these codes. Examples include both freeware and open-source software, as well as professional tools like CAD drawing software and video or music editing tools. The latter were sometimes mentioned with the explanation that employees would source these tools by themselves to save costs or circumvent (lengthy) institutional procedures.

> *"[...] those teachers who then think they should have the full version of [software] and the academy thinks they shouldn't, so the money isn't there for it, and who then just go via [torrent website] or something and download a cracked version, because 'I have to have it for work'. [...] that too is not stopped." [Expert 01]*

Not a lot of different codes were used in this category for two reasons. Firstly, interviewees often did not know the exact nature of these applications. Although they gave examples ranging from torrenting software to open-source email clients, most interviewees just mentioned that software is installed, downloaded (implying a locally run executable) or gave a hypothetical example like Word Perfect. The 'unknown unknown' aspect, where we cannot know what we don't know, was mentioned by Expert 08 in this context as the reason of not knowing what was exactly present. The second reason we don't specify categories further, is that the category of software matters little for the security perspective we are taking in this thesis. All software that executes code directly on the system (as opposed to in a browser sandbox) has an equal threat model. This might also be the reason that many of the interviewees did not specify the type.

Institution 10 scanned their managed endpoints for software that was not installed by default or provided by the central department and found a staggering 18,000+ unique applications installed, while they distributed only 600 applications themselves. Although the difference includes multiple -likely outdated- software versions and distributions for different architectures, the number gives an impression of scale:

> *"No actually, look, when I talk about that [18,000+], I do talk about the whole bunch of shadow that [institution] scans. [...] Yes, and if we could have a much better view and grip on that, that would be really nice."*

5.2.1.2 Self-made solutions

The second-most-mentioned type of shadow IT overall was within the category of self-made solutions (Figure 6), namely **spreadsheets and databases**. As one of the earliest forms of shadow IT mentioned in the literature (Raden, 2005), it has apparently not diminished since. These are not only deemed 'shadow' because they duplicate data from official systems, but also because they provide functionality that should ideally be done in a more managed system.

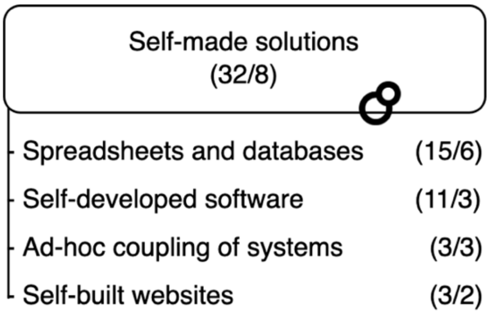

Self-made solutions
(32/8)
Spreadsheets and databases (15/6)
Self-developed software (11/3)
Ad-hoc coupling of systems (3/3)
Self-built websites (3/2)

FIGURE 6: OCCURRENCES OF SELF-MADE SOLUTIONS

> *“We have a number of applications for finance, for HR, and things like that. And then there are still people who then make their own list out of it. They do use all the approved tools for that. [...] Yes, so that a kind of unofficial copy of the data is created, which is then not as secure as the system it actually belongs in”. [Expert 07]*

In the same category of self-made solutions, there is **self-developed software**. An example that re-occurs multiply times in the interviews is that of the researcher who builds something themselves, as mentioned by Expert 04:

> *“[...], look, if you really look a shadow IT, then there are the researchers who are developing something anyway, put it into production, and we don't actually know that they are doing. Besides, they haven't applied privacy and security by design either.”*

Together with **self-build websites** and **ad-hoc coupling of systems**, these four make up the category of self-made solutions. Self-developed websites were, like self-developed software, mostly mentioned in the context of research, while ad-hoc coupling was mostly done by support staff to ‘hack’ different systems together instead of using official APIs or a service bus:

> *“Well, or that someone at some point writes a little tool, [...] maybe still from the last century, that pulls data out of the [directory service] and then throws that into a CSV that can be imported into other things.” [Expert 01]*

5.2.1.3 UNNAPROVED CLOUD SERVICES

Equal in occurrence in the data is the category of unapproved cloud services, the code overview of which we show in Figure 7. Although they are all (mostly) browser-based, we made the distinction into three categories to distinguish between the type of functionality since it could influence the impact it has on the threat landscape. They are separated into web-apps, a website for a single service; cloud productivity suites, a full-fledged workspace with roles, storage and editing, or applications that mainly provide storage. For **web-apps**, examples often included digital whiteboards and other tools for education:

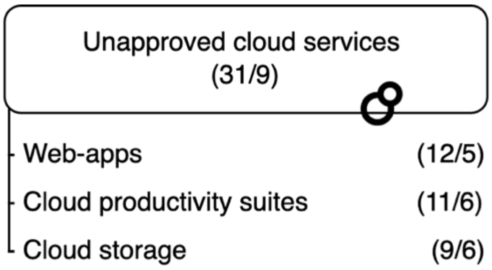


FIGURE 7: OCCURRENCES OF UNAPPROVED CLOUD SERVICES

> *"Yes, the familiar ones are [web-app used in education], [online education platform]. I think you can also speak of [two different presentation and voting tools] with whom we did happen to have contracts at one time, we still have, but those will go out the door again soon, though. But those are tools that are really used for that as well. And*

> *no matter how innocent, it is always a risk assessment that has to be made. But that has to be made by us and by the organization, and not by a teacher." [Expert 06]*

Cloud productivity suites might be among the larger scale shadow IT instances deployed. Six out of the eleven interviewees had at one time discovered smaller organizational units using their own environment, separate from that of the institution.

> *"Yes, so we have our friends here from [a specific department] who [...] have everything running at [a shadow cloud provider]. And by everything, I mean everything. They have their own mail domain on [provider] they have their own [included cloud productivity tool]. They have their own, they have some [other tool] server that's all automated through a [cloud provider] thingy where they just send a certain email to a certain email address and then everything starts. It all ties together from bits of wood and strings" [Expert 01]*

The usage of **cloud storage** was smaller in scale, often mentioned in the context of quick sharing of files. The specific longer-term collaboration use-cases stood out because it showed a need to work together with staff from other organizations, incompatible with the official system:

> *"For example, we don't have [shadow file hosting service] either, do we? We do have [official file hosting], but not [shadow file hosting service], because IT has said, yes, we are not going to manage everything. [...] Yes, and if other parties use [shadow file hosting service] then we could say well, that's not allowed, but we don't do that. So, then they will also start using [shadow file hosting service] and data will go in there." [Expert 03]*

5.2.1.4 Self-acquired devices

The highest variety of codes was in the category which Mallmann et al. (2019) call 'self-acquired devices'; shadow IT hardware. Although these could have been grouped into bigger code groups when taking a different perspective, we suspected a larger variety of vulnerabilities and threat actions to be associated with them, so we kept them separate, as shown in Figure 8. **Unmanaged PCs** (personal computers) were most named by the experts. Here, the line between regular IT and shadow IT is often blurred since unmanaged devices could be given out by departments in the institution, or the institution would allow for staff to re-install the operating system on the devices so they become unmanaged as this quote from Expert 01 illustrates:

> *"[...] we always have [laptops from a certain manufacturer], [...] if your [department] says, well, that doesn't suffice, then you get the amount that such a [laptop] costs and*

> *then the [department] adds money to that to buy something else. [...] My colleagues in [workplace management] developed a nice image for that and the first thing those guys do is put a USB stick in and reinstall [the operating system]. Yes, because 'it's all not working' and 'we're just being restricted'. Difficult."*

On the other hand, there was the more common example of staff simply using their personal hardware for the job, including their **mobile phones and tablets,** while there were no regulations guiding that type of usage. From the perspective of the central IT management personal devices are the same as **devices managed by others,** for instance when staff are employed by multiple institutes and use the managed device of only one of these institutes. Although they might have better security safeguards, from the perspective of the organization it is another unknown device.

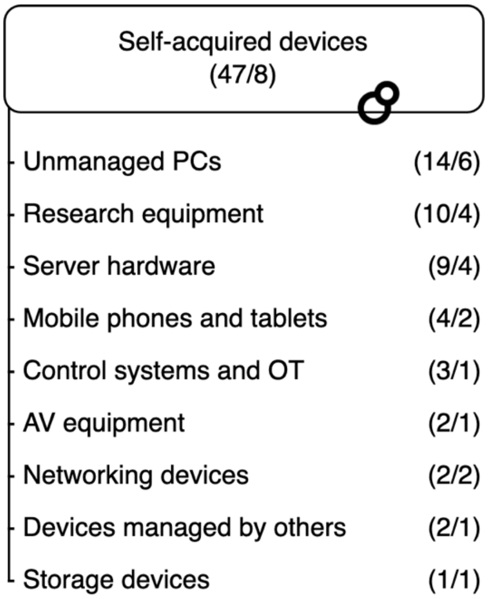


FIGURE 8: OCCURRENCES OF SELF-AQUIRED DEVICES

Research equipment included, for instance, microscopes and fridges, often running embedded software or even having networking capabilities, as Expert 07 describes:

> *"We also have a [department], and you also have all kinds of equipment there with computers built in. Yes, that is also not purchased centrally. Then, well, then an [research device] comes in and there's a [legacy] device in there. Well, is that Shadow IT? Under certain definitions it is."*

Server hardware acquired by staff ranged occurred in many different versions. Network attached storage (NAS) was often a key part of the configurations, but both education and research staff brought more elaborate server hardware into the network to fulfill their needs:

> *"For example, researchers have a certain budget for research and then it is cheaper to buy a desktop PC and start setting it up as a server than to get a real server that you put in a data center." [Expert 07]*

> *"[...] one time for the course he taught, he had configured a server, where students themselves could then build a database and a website using PHP and mySQL, etcetera, etcetera." [Expert 05]*

On the role of **control systems and OT** (operational technology), out of sight of the IT department, Expert 04 notably said that it is "*much bigger than we think*", although they were the only one to mention it. Other specific types mentioned only once or twice were **AV equipment** such as internet-connected cameras and **networking devices** like a WiFi access point or switch. Although the rogue USB stick was once infamous for its role in distributing malware (for example in Kushner (2013)), non-network attached **storage devices** seem to have a smaller role nowadays, at least in experts' perception of shadow IT occurrences.

5.2.2 Threat Components of Shadow IT

Does this shadow IT create cyber threats for your organisation? What security incidents are associated with shadow IT usage? How could shadow IT lead to… Who or what would do this? What enables it to do harm? What would then go wrong? What are the consequences? What damage is then done?

For this section, we will present the found threats, vulnerabilities, threat scenarios, unwanted incidents, and assets. We present the experts' perceived threat components, per category, as the two researchers coded them. In Chapter 0, we will discuss relations between occurrences, vulnerabilities, threats and how the combinations could be exploited.

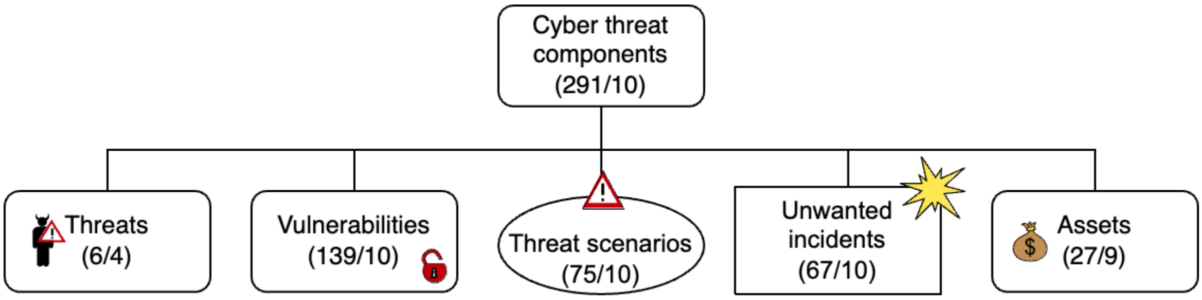


FIGURE 9: THREAT COMPONENTS WITH GROUNDEDNESS AND COVER METRICS

As shown in Figure 9 vulnerabilities, threat scenarios and unwanted incidents were mentioned most out of the five types of threat component used for coding. All experts shared perceptions on those, except for Expert 11. Coming from the smallest institution in our sample, they indicated to not see any shadow-IT related cybersecurity threats:

> *"I think the past few years have deliberately taken this turn. Because when I came in here, it was a bit freer. There were still some [departments] that were consciously doing little things independently. Well, we expelled all of those. Also while moving towards that new GDPR situation. And now everything is centralized. Yes, because of the whole cyber thing, everyone is even more alert. Even the managers. So I regularly get the question of: first go check with [Expert 11], go and see if it's allowed, if it's*

> *possible, and if they don't see anything crazy. [...] Yes, that indicates: first look and check before you start something. And of course, I have sufficient talks with larger [institutions], and you also see that things are being purchased in all sorts of nooks and crannies of the institution without them knowing it. That they also don't know what they have running. Yes, that is not the case here. And of course that is compounded by the fact that we have outsourced it. [...] So, I have some warning systems, I think, that work well, that give me signals if something is being used improperly."*

All other institutions came up with cyber threats related to shadow IT. Mostly hypothetical ones that they deemed possible given their IT, but also some incidents that actually happened.

5.2.2.1 Threats

The semantic category of threats, shown in Figure 10, was difficult to extract from the interview data. We noticed that many experts left the *who* or *what* aspect of a threat implicit or assumed it did not need explaining. *"Being hacked"* was seldom mentioned with a specific threat behind it. If we were to code for it, *"they"* would be the most mentioned threat:

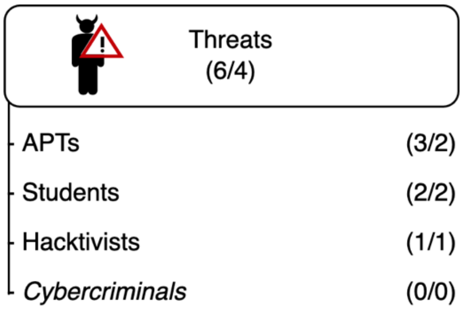


FIGURE 10: THREATS RELATED TO SHADOW IT THREAT MODELLING

> *"When that email address comes from the [institution] the first thing they are going to do is try email address at the [institution]. They are going to log in and bam! with the password, see if it works. That's what happens." [Expert 04]*

It could easily be deduced that, in most cases, a cybercriminal should be seen as the initiator of threat scenarios, but not explicitly enough to emerge during the codebook development. So, for the threats, we only coded deviations from this implied default. **Cybercriminals**, motivated by financial gain, were therefore the default actor. As explained by Expert 03:

> *"And otherwise it's all untargeted, we're just an [institution], just an institution, say, where they then happen to have a hook, where they can just go in because it's convenient. [...] Generally, let's say, our major actors, it's the criminals, that we have the most trouble with. [...] Those criminals who are not, you know, they don't want to cripple [the institution], or they're not pissed at [the institution], or they don't have any particular reason to attack [the institution] other than they think that... [...] Yes, just making easy money."*

The deviations from this default often differed in motivation. A smaller role in the threat landscape exists for **APTs** (advanced persistent threats) or state actors who would be motivated by sabotage, or more commonly mentioned, espionage:

> *"[...] there are also a lot of, and then we call it advanced persistent threats, eh, those state actors who just say, yes, you know, I don't need to take down [the institution], I don't need to do a ransomware attack, but I do really want to know exactly what they are doing day and night, I just want to know, eh, maybe a little interesting research, [...]. Well, that is of course where my big concerns are. I just don't know. I don't know if they are inside and watching. [Expert 10]*

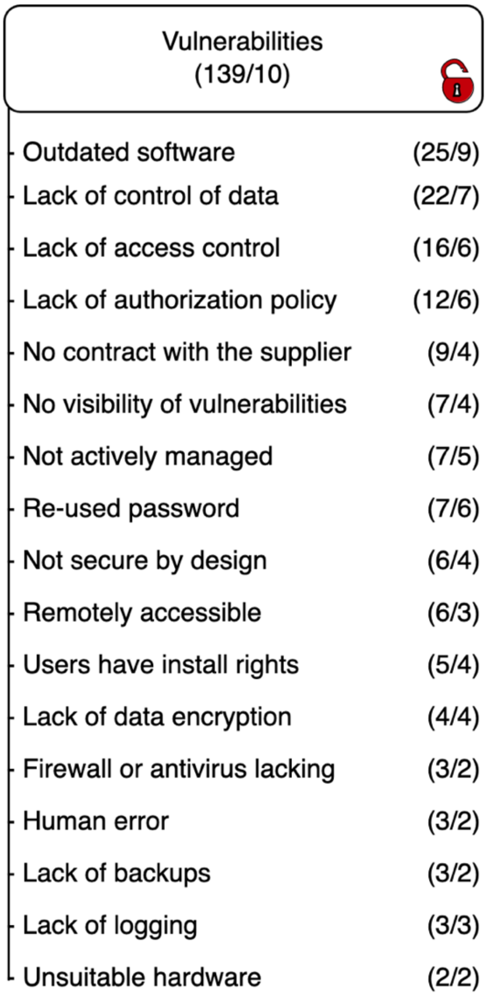
This actor was mentioned by three of the interviewees. An actor unique to this sector is the **students**, who might try to leak or change student data for financial or academic gains:

> *"Imagine if one of our students manages to find a backdoor into our grade registration system and they go and raise grades for a few tenners." [Expert 03]*

Lastly, there are **hacktivists**, groups who unite to carry out cyber-attacks in support of political causes (Fowler, 2016), mentioned laterally by Expert 03: *"Yeah, we do have hacktivists, but [...]".*

5.2.2.2 Vulnerabilities

Vulnerabilities was the semantic category within the threat modelling with the most different codes, applied most often. In Figure 11 they are already sorted by the measure of relevancy used in Section 6.1.2. It should be noted that some of these are overarching for more specific vulnerabilities. For those, we had to find a balance between coding it how the experts perceived and called the vulnerability, and what could be the underlying technical problem. In the codebook (see Section 8.2) they are described as super-vulnerability, as per the analogy of a superclass in object-oriented programming.



Vulnerabilities (139/10)	
Outdated software	(25/9)
Lack of control of data	(22/7)
Lack of access control	(16/6)
Lack of authorization policy	(12/6)
No contract with the supplier	(9/4)
No visibility of vulnerabilities	(7/4)
Not actively managed	(7/5)
Re-used password	(7/6)
Not secure by design	(6/4)
Remotely accessible	(6/3)
Users have install rights	(5/4)
Lack of data encryption	(4/4)
Firewall or antivirus lacking	(3/2)
Human error	(3/2)
Lack of backups	(3/2)
Lack of logging	(3/3)
Unsuitable hardware	(2/2)

FIGURE 11: VULNERABILITIES RELATED TO SHADOW IT THREAT MODELLING

A **lack of control of data**, for instance, is not a vulnerability that can be exploited by a threat, but it was a very common way for the experts to describe what enabled further problems with shadow IT; when the organization is not in control of the data, for instance, stored on the NAS of a researcher, it cannot ensure proper access and authorization policies, backups, and data encryption, amongst others. We chose not to add a hierarchy to these vulnerabilities since the categories would not be disjoint; lack of backup strategy, for instance, could be part of 'lack of control of data', as well as 'not actively managed'. **Outdated software**, containing known security flaws, was mentioned by almost all as a key vulnerability, making it the vulnerability that is mentioned in most interviews. Expert 05 gives an example representative for many other mentions, already giving a preview of what can go wrong with a vulnerability like this:

> *"Well, the server was also not managed very professionally, so security patches were not installed. So yes, it had been an easy target to get in for a hacker [...] from the outside. To use it as a steppingstone to attack our network and explore what's out there. So yes. Could have gone very wrong."*

This vulnerability is closely related to **lack of visibility of vulnerabilities**. Not only can vulnerabilities be present because of shadow IT, but a security professional can also be unaware of them, and not able to fix them. Three experts used the example of Log4j, the 'most critical vulnerability of the last decade'[5], to illustrate this. This vulnerability, present in a software library used by numerous other programs, had to be updated in all systems using it to prevent serious risk of remote access:

> *"I think Log4j is, once more, a very good example. When a vulnerability like this passes by, everybody knows: here is just really a leak, a vulnerability that anyone can exploit. In fact, the whole world knows that, so any hacker can take advantage of this. Still, if we don't know where Log4j is, then we also don't know if we have closed the hole. Those are the big drawbacks of shadow IT." [Expert 10]*

The code **lack of access control** is the third in terms of groundedness. We also applied this code when the access control used was not sufficient. For example: *"It had a security code of 1234"* [Expert 05] or *"Storage [...] without single sign on, without MFA, with a, well, easy to guess password."* [Expert 09]. But there were also examples where files or hardware had no access mechanism at all:

> *"Who would know that if you go to [application].domain.nl/employees.csv that you would simply get all the employees with all the dates, all the addresses and I don't know what...?" [Expert 01]*

[5] https://www.theguardian.com/technology/2021/dec/10/software-flaw-most-critical-vulnerability-log-4-shell

> *"It is [...] controlled through a PC that is at a counter. That counter is just open, and youonly have to press the keyboard, on the space bar, and it turns on. And look, I'm in the application. Yes, that's easy! [...] So in terms of security it's absolutely not wanted. Because it's a regular PC, connected to a network, and so on, and you just have to press the space bar and you're on that side." [Expert 04]*

The closely related vulnerability of **lack of authorization policy**, or the more specific version **users have install-rights** implied that some authentication was in place, but a system gave access to everybody in the same way, not checking if that person should perform a specific action, access that specific system or (sensitive) data. This would for example become clear in a related scenario *"students could look at each other's home directories, could copy all the stuff from each other"* [Expert 05].

No contract with the supplier, not secure by design and **not actively managed** occur in relation to different scenarios but are related in the sense that they imply a lack of control from the institution that causes the institution's security standards for a device or service to not be met. Devices and services are not getting *"the attention they need"* [Expert 01]. Expert 07 exemplifies how this can show:

> *"What we sometimes see is people hooking up a raspberry pi to the network, for example. And a raspberry pi, if you install it, has a default username and password, so you're in before you know it."*

When the device is also exposed to the outside network, it is **remotely accessible**. This could show as servers with unsecured APIs, or servers that are accessible via RDP. Expert 03 shared the example of this vulnerability being present in a camera:

> *"On the other hand, of course, you have devices that, yes, talk to the internet, like that crazy camera. Which, well, you wouldn't expect it, but apparently if you run a little script on that you can just snip out a password."*

Another mentioned vulnerability is a **re-used password,** which can happen when a user signs up, on their own, for a service that does not integrate with the institution's authentication solution. A note about this from Expert 11 was passwords re-used is unlikely when the usage of the shadow service is *not* initiated by the institution:

> *"No, I think that's rare anyway. Because often they already have that [account], don't they? [...] I don't get the idea that they re-use our password, which is in fact provided centrally by us. I don't see logic in that, frankly. I think those are just separate environments."*

But when it *is* initiated by an institution staff member, Expert 04 assumes this to be likely:

> *"But if a lecturer says to students, you must now use this and that application [...] they're going to use the [institution] account anyway, they say well [institution mail address format] with the [institution] password and we don't want that."*

Then there are a few vulnerabilities phrased as missing a security measure: **Lack of data encryption, firewall or antivirus lacking, lack of backups** and **lack of logging.** They are not mentioned as often as others, but are very specific. To what extent these vulnerabilities can *cause* an incident or make an existing incident worse will be discussed in Section 6.1.2. Lastly, **unsuitable hardware**, can be a very practical concern:

> *"And that little server was just a small PC that was under the desk somewhere. Then a cleaner would come by and clean it and the power cable would get disconnected and then all those students couldn't access their work, again. While they were graded on it. So, well..." [Expert 05]*

5.2.2.3 Threat scenarios

Figure 12 presents the overview of threat scenarios occurring, in order of groundedness in the data, for malicious actions related to shadow IT. These themes could be coded about half as often as vulnerabilities. Sometimes, the steps of the 'cyber kill chain' (Hutchins et al., 2011) could be recognized in the steps described. Potential attack paths in those cases included some form of initial access, installing malware, moving laterally to other systems, and then performing actions on an object to gain from the steps taken. Interestingly, Expert 09 disregarded the ransomware incidents often associated with this 'classic' hack as follows:

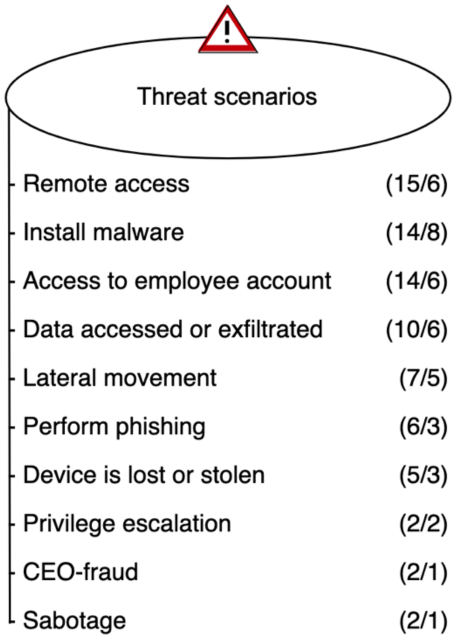


FIGURE 12: THREAT SCENARIOS RELATED TO SHADOW IT THREAT MODELLING

> *"So I think it's a bit, and I have to be careful saying this, it has had its prime days, those kind of ransomware attacks that [an institution] has experienced."*

Nevertheless, many experts mentioned steps from a multi-step scenario, with **remote access** as one of the main initial scenarios to worry about. It allows for further steps such as **install malware, data accessed or published,** or **lateral movement**:

> *"And yes, on that server there is already, well, a gigantic heap of data, already on that server alone. So that can already, well, be ransomed, or published, plus it can be used as a stepping stone to attack the rest of the network." [Expert 05]*

Of course, not all attack paths mentioned were this elaborate. A single event of data accessed or exfiltrated would often be enough to consider something a cyber incident. Another first step mentioned would be to **perform phishing**, sometimes mentioned in combination with usage of data that was gathered earlier to make such an attack more effective:

> *"[...] or you get information via other means, well, by being in a mailbox, or being inside the shadow IT, where we are using the same names as in our normal environment. It's all information points that a hacker can use to eventually get inside your organization. And that can lead to all sorts of risks. [...] To send a very good phishing mail later, yes." [Expert 04]*

The scenario above, like many others, has **access to an employee account** or other **privilege escalation** as a component. Although the latter phrasing is, mainly used in the more elaborate chain that most experts are familiar with:

> *"Yes, look, the moment then that real lateral movement occurs as well, hey, so they get more and more rights on the network, which eventually leads to a real data leak or data stealing or ransomware, you know, then of course it's already too late. And then, yes, by then the harm is done." [Expert 10]*

In terms of less intentional scenarios, a **device is lost or stolen** is a scenario that security experts must deal with. Lesser-mentioned scenarios are **sabotage** and **CEO-fraud.** The first is often a part of malware, leading to ransomware or extortion (in that case, we apply those codes), but it can also occur on its own, most notably by means of a DDoS attack. CEO-fraud is also known as 'business email compromise' (Cross & Gillett, 2020), although in the example given by Expert 01, chat was used as a medium:

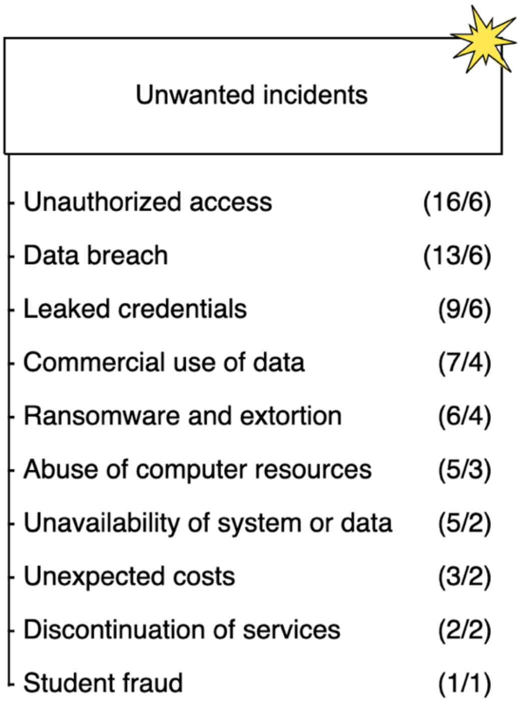
> *"there have been cases of people high up in the tree being written to by colleagues with certain requests that seemed perfectly legitimate, but then it was not the actual person [...] And that is with very direct financial gain. [...] Well, in this case it was about the nail in my coffin called [messaging app]."*

5.2.2.4 UNWANTED INCIDENTS

For the unwanted incidents, we uncovered ten repeating themes of events that can harm or reduce the value of an asset. The most prominent is **unauthorized access**, for instance to a user account or device. As opposed to the '*gaining* access' or '*exfiltrating* data' perspective that was taken when mentioned as a threat scenario, experts mentioned this as the unwanted resulting *state* of such an action:

> *"So then it's about, well, again a kind of access point on which you have no visibility. And in this case, really having access to information [...] And again, that could either be data that is in itself sensitive, personal data or the like, or lead to a very nice phishing email." [Expert 04]*

The incident of a **data breach** can be seen as an overarching concept. If the contents or even consequences of a data breach were clear, another code was used instead. Therefore, this code was mostly applied for scenarios where data was published and out in the open either by accident or to cause reputation damage. The more specific data leaks coded for were **commercial use of data** and **leaked credentials**. The first of these can also be seen as a subtype of unauthorized access, since data is accessed for a purpose not intended. It is a separate code because the intentions and consequences are different. Leaked credentials could occur as follows:



Unwanted incidents	
Unauthorized access	(16/6)
Data breach	(13/6)
Leaked credentials	(9/6)
Commercial use of data	(7/4)
Ransomware and extortion	(6/4)
Abuse of computer resources	(5/3)
Unavailability of system or data	(5/2)
Unexpected costs	(3/2)
Discontinuation of services	(2/2)
Student fraud	(1/1)

FIGURE 13: UNWANTED INCIDENTS RELATED TO SHADOW IT THREAT MODELLING

"But you see, [a cloud app provider] did get hacked once, several years ago already.... Yes, then all login details... And with [a social network] it happened too. [...] So there is that risk" [Expert 02]

Ransomware and extortion were for many experts one of the biggest incidents imaginable. With big incidents at several institutions still fresh in memory, it is also infamous as a type of incident that can be looming within the network for a long time:

> *"What you also saw at [a different incident] -I try never to refer to it again, because it's been so long ago by now, but hey- there you also saw, there they spent [several] months. [...] It speaks to the imagination indeed. There they spent [several] months before they actually flipped a switch somewhere." [Expert 03]*

With ransomware, data is made inaccessible on purpose. We also saw this **unavailability of system or data** mentioned as a side-effect or unintended event. A separate code was used for this. For cloud services, this was phrased as **discontinuation of services.** In the example by Expert 06, we also see an example of potential **unexpected costs**:

> *"If [a shadow service] were to say next week: we will make all free accounts paid. Then things do break down here and there, so to speak. [...] Then really an x number of teachers and courses are really screwed. Because yes, those students can no longer log in, can no longer access their work, can no longer run an export."* [Expert 06]

Furthermore, experts mentioned **abuse of computer resources** such as mining of cryptocurrencies [Expert 04] or hosting of phishing sites [Expert 07]. Lastly, there was **student fraud**, mentioned by Expert 05 in the context students copying each other's work from a public server.

5.2.2.5 ASSETS

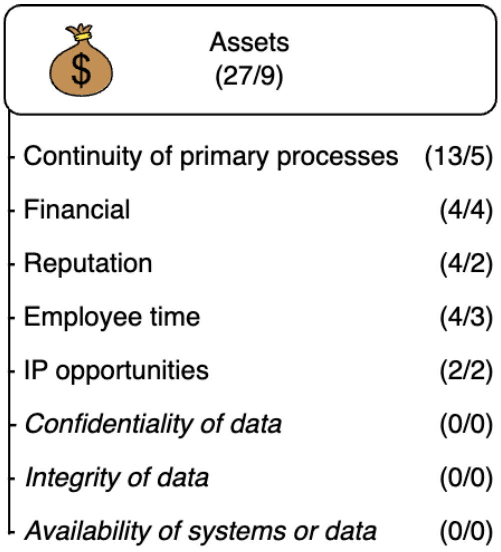


FIGURE 14: ASSETS RELATED TO SHADOW IT THREAT MODELLING

Although almost all of the incidents can be extended to financial consequences with some imagination, the examples above show that often the damage is done not directly to a tangible good but to **confidentiality, integrity** or **availability of systems or data,** also known as the CIA triad, first mentioned in a NIST publication by McKenzie & Ruthberg (1977). We added these implicit assets for later use in analysis of attack paths. Similar to the *who* or *what* aspect of threats, the experts often left the asset that was damaged implicit, just indicating there would be damage or consequences. Still, some specifically mentioned assets could be coded. **Continuity of the primary processes**, education and research, were mentioned in the context of shadow IT. In some cases, because a cloud service *"gets pulled away and you could be in the middle of your curriculum."* [Expert 05], but also in the context of having to be able to justify study results to authorities:

> *"Besides, yes, your accreditation... There are certain things you have to show when an accreditation committee comes along. You must show, well, where did you do your teaching? Yes, if that tool is then no longer there two years after you used it. Well, that that was a vital part of your teaching. It will have to be kept for some time anyway.“ [Expert 05]*

In this case, shadow IT damages the institution in the core of its existence, by damaging education results. Although damage to **financial** assets was mentioned by more experts, some experts perceived **reputation** damage as an even bigger risk, especially in the context of data breaches:

> *"And the [institution] that's the name huh. So you have a data breach, and you have the name up for grabs as well. That's actually the worst yet, reputation. So, the reputational damage, that is perceived to be big." [Expert 04].*

A specific type of asset present at a research institute consists of **intellectual property opportunities**; inventions that might be relevant to cybercriminals, but even more so to APTs, who can, according to Expert 03, *"probably gain a tactical advantage with it"* if the knowledge is innovative enough to be worth spying on. Besides losing money, trust and knowledge, an institution can also experience negative consequences from shadow IT because it takes up **employee time**:

> *"So well, there was really a gigantic, yes, management burden on a teacher. Yes, that he said: now I hardly have time to really teach. I'm constantly busy keeping that little server up and updated and addressing those students on how to do that and, well, also answering help questions."*

5.2.3 Other Threats of Shadow IT

Two other categories of threats not included in the previous section should be mentioned: threats to legal compliance and to education and research 'business' processes. Although both are not novel consequences, already mentioned in the extensive literature review into causing factors, outcomes and governance by Klotz et al. (2019), they were mentioned frequently by the experts interviewed. Some of the security experts we spoke to also had a privacy role within the organization, as the fields are often intertwined in smaller institutions. Many of them shared privacy concerns, especially in relation to software with a cloud component:

> *"Yes, with the tools on the internet, the cloud thingies, there is less of a security aspect there, but much more a privacy aspect. Yes, what happens to the data we store there, does it stay within the EU, what does such a provider do with that data?" [Expert 05]*

Under the broader concept of 'GDPR non-compliance', issues with data retention, storage outside the EU, processing of PII, and many more could arise if the software was not officially procured and assessed, or data was stored on unofficial cloud storage or hardware. Further compliance issues can arise if the usage of a certain shadow system becomes so large that a procurement process should have been started. Although not including compliancy aspects served to limit the scope of this thesis, they should not be ignored if policies would be designed for shadow IT.

The second category of problems related to shadow IT that is out of scope, but was mentioned frequently, are threats to business processes. The unmanaged nature of shadow IT can cause problems to business continuity when the person responsible for a shadow system leaves the organization or the third party providing the service stops doing so. At a smaller scale, shadow IT could cause incompatibility between files, business processes or even IT architecture in the organization:

> *"If they have some technical editor they prefer to use, or something like that that they work in, then I see more a business continuity story. Or… It occurs that people have differences between [unsupported PC] and [supported PC] or something like that. That there are just files being sent around that that are not compatible." [Expert 02]*

Furthermore, data duplicated in shadow IT could cause the incorrect information to be used for business decisions. Although a clear downside, it was difficult to relate these themes back to security threats, and they were therefore not further explored after the open coding phase.

5.3 Takeaways for the Next Chapter

This chapter presented the findings resulting from the analysis of eleven expert interviews, which resulted in an extensive list of shadow IT occurrences and recurring threat components for the cyber threats related to them. Shadow IT has a significant role in HEIs IT landscape, and it has become clear that for almost every type of hardware or software in an organization, a shadow version can exist. Despite big differences in the scale, types and users of shadow IT, the interviews gave a complete impression of the challenges organizations are facing. The completeness of these observations is shown by the high code saturation measured.

Although these observations are interesting on their own, their true value should show when connected to each other. In the next chapter, we provide this high-level overview with links between the codes to discover the patterns present in the interviewees' observations. Putting together the co-occurrences of the types of shadow IT, their vulnerabilities, and the attack paths they enable should give actionable insights into *how* shadow IT poses a risk to HEIs.

6 Patterns of Shadow IT Threats in Higher Education

From the expert interviews, many observations and perceptions of occurrences and threats were gathered in Chapter 5. The goal of this last empirical chapter before the discussion is to link the occurrences with their related threats and to link the different threat components together. By combining and exploring the insights of the different experts, we will construct a comprehensive view on the role of shadow IT in HEI's cyber threat. This will answer the fifth and last subquestion of this thesis*: "Which recurring patterns can be identified in these occurrences and cyber threats?"*

6.1 Cyber Threats of Shadow IT

For relating shadow IT occurrences to threats, as well as for the identification of attack paths, we link the elicited components in bigger diagrams. The diagrams mostly follow the meta-model of threat diagrams in CORAS (Lund et al., 2011), where a threat *initiates* a threat scenario *by exploiting* a vulnerability, *causing* an unwanted incident *damaging* an asset. Threat actions can follow multiple other threat actions, for which further vulnerabilities can be exploited. We extend this modelling language with the occurrences of shadow IT, which can *have* a vulnerability. The meta-model can be seen in Figure 15. The diagrams were made using draw.io (JGraph Ltd, 2022) and the official icons from the CORAS website. Components could be linked based on co-occurrence in the transcripts when relations were made explicit by the experts.

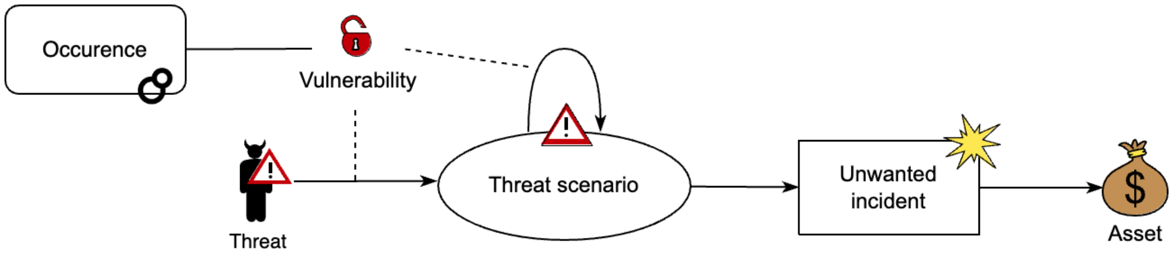
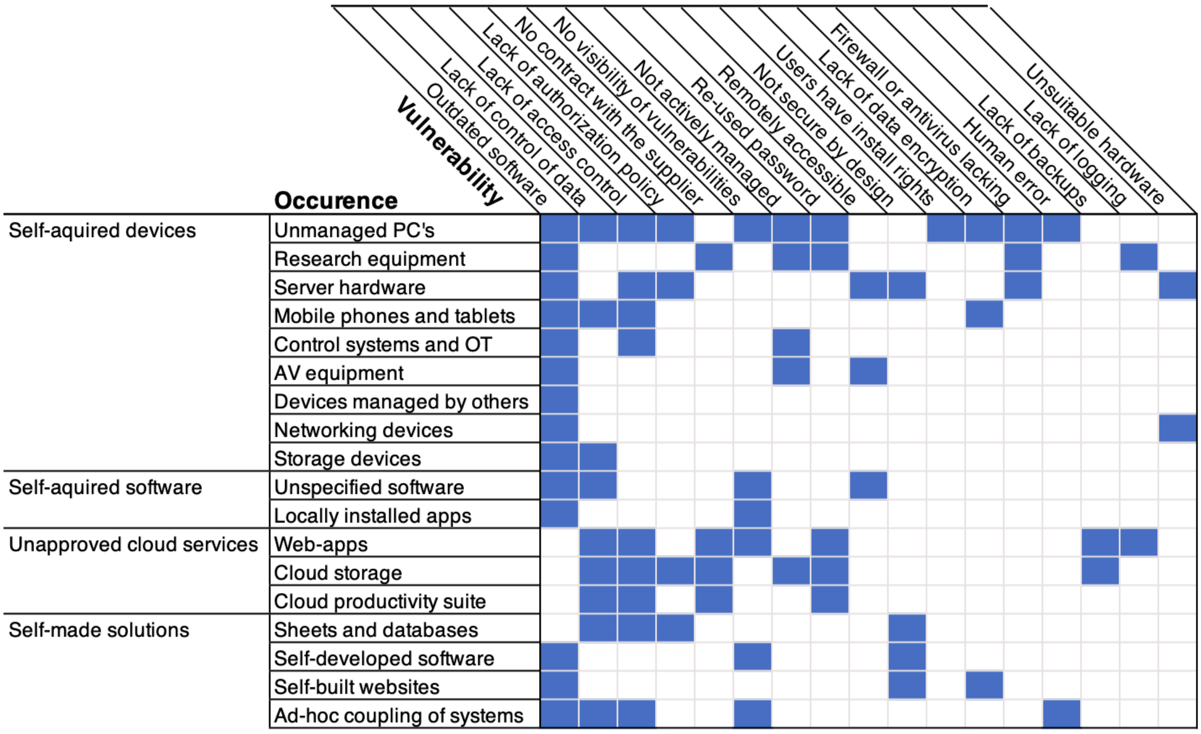


FIGURE 15: THREAT MODELLING META-MODEL

We focus on two possible views of these networks of occurrences and threats. Firstly, an "occurrence-vulnerability" view will be worked out to show with which vulnerabilities the types of shadow IT are related, according to the experts. Secondly, some of the possible attack paths will be elaborated upon. Many of the collections of codes and the links between them had to be trimmed to have understandable diagrams. However, for further exploration, the complete codebook and project are available in the data repository (see Section 8.2).

Some interviewees shared how they perceive the dangers and scale of shadow IT in their organization. Expert 04 called shadow IT "*one of my biggest risks*", while Expert 05 indicated that “*the degree of shadow IT is, I think, quite high*”. Most of them indicated it is almost impossible to get a sense of how big it is since it is, by definition, out of sight. Expert 02 shared a slightly different perspective, calling it “*part of the job*” and, although it has their attention, “*not something that keeps me up at night*”.

6.1.1 Occurrence-Vulnerability Patterns



	Vulnerability Occurence	Outdated software	Lack of control of data	Lack of access control	Lack of authorization policy	No contract with the supplier	No visibility of vulnerabilities	Not actively managed	Re-used password	Remotely accessible	Not secure by design	Users have install rights	Lack of data encryption	Firewall or antivirus lacking	Human error	Lack of backups	Lack of logging	Unsuitable hardware
Self-aquired devices	Unmanaged PC's	■	■	■	■		■	■	■			■	■	■	■			
	Research equipment	■				■		■	■					■			■	
	Server hardware	■		■	■					■	■			■				■
	Mobile phones and tablets	■	■	■									■					
	Control systems and OT	■		■				■										
	AV equipment	■						■		■								
	Devices managed by others	■																
	Networking devices	■																■
	Storage devices	■	■															
Self-aquired software	Unspecified software	■	■				■			■								
	Locally installed apps	■					■											
Unapproved cloud services	Web-apps		■	■		■	■		■							■	■	
	Cloud storage		■	■	■	■		■	■							■		
	Cloud productivity suite		■	■		■			■									
Self-made solutions	Sheets and databases		■	■	■						■							
	Self-developed software	■					■				■							
	Self-built websites	■									■		■					
	Ad-hoc coupling of systems	■	■	■			■								■			

FIGURE 16: MATRIX VISUALISATION OF OCCURRENCE-VULNERABILITY CO-OCCURRENCE

Figure 16 shows how the different vulnerabilities are related to the occurrences of shadow IT. On both axes, the items are sorted by the number of times mentioned in total. The figure provides insight into the density, the number of links from a code to other codes, of the different vulnerabilities and occurrences. It should be noted that this overview is probably incomplete. Although the *code* saturation, whether all relevant concepts are mentioned, was shown to be very high, we suspect this is not the same as *connection* saturation since the possible connections are a multiple of the codes. This means there could be missing links. Although it would be possible to fill in the gaps with knowledge from the literature, for now, we will present the data as-is.

Outdated software has the highest density of all vulnerabilities. For all types of shadow IT, except for spreadsheets and the different cloud services, outdated software is perceived as a vulnerability. It has become clear by now that this is a top-of-mind concern for experts across organizations and applicable to most of the IT landscape. We think this is partly the same as ‘no visibility of vulnerabilities’ since the main worry of outdated software would be the unpatched vulnerabilities in them:

> *“[...] a [researcher] is put in charge, who leaves. After that, there is no one left to manage it, there is no one left to do updates. So after a while, it is then very vulnerable.” [Expert 07]*

After the potential problems with vulnerabilities in outdated software, the different types of ‘lack of control of data’ are most connected. This way of phrasing is mainly used by the interviewees for places where data is stored: the cloud and devices. Often, it is combined with worries about access control:

> *“The main danger of those 100 tools is: I just don't know where that data is, what happens to it, how long it is kept, who does what with it. [...]. Yeah, I just don't think that's acceptable. If you, as a teacher, do practical research within [the institution] with your students, and you select a tool, you have to be convinced that you know exactly what happens to the data provided during the research.” [Expert 10]*

This broad category would sometimes be specified during the interviews with other data-related vulnerabilities, like lack of encryption, authorisation, or, less commonly, lack of backup strategy.

From the uncovered vulnerabilities, ‘re-used passwords’ is unique, because there is little semantic overlap with other categories. It is a very specific problem for shadow IT occurrences where a password is chosen by the end-user (cloud services and devices), and the application or device is not linked to the single sign-on solution of the institution. We suspect it was not mentioned for mobile devices since these often require a passcode instead of a password. In case of a bad passcode, it would be a ‘lack of access control’, which was defined in the codebook as a system not having *sufficient* access control (see Appendix E: Codebook), allowing for access by unauthorized users.

Besides looking at which vulnerabilities are perceived for many different occurences, the linking of vulnerabilities to occurrences can also be approached per occurrence category. There seem to be some patterns of specific vulnerabilities per category. For cloud services, re-used passwords and lack of contracts with the supplier are distinctively visible. ‘Lack of contract’ could mean for a cloud service what ‘not actively managed’ is for self-acquired devices: nobody looks after the security and data of the system in your interest. Expert 05 mentioned that for cloud tools, privacy concerns are much more prevalent than security concerns (see Section 5.2.3). This is supported by our data, where the (discarded) codes on privacy are often co-occurring here. From a security perspective, web and cloud apps have the advantage of running in the browser sandbox, being able to do less harm on the network. Although sandbox escapes exist, they are a less common scenario. On the other hand, the availability of these tools makes them take on a bigger role again.

Self-acquired devices have some specific issues that are not perceived for other types of shadow IT, such as the unsuitability of the hardware. One interviewee gave the example of a server in an office that could easily be disconnected if the cleaning staff did not recognize it as such (see Section 5.2.2.2). Besides, experts were worried about the unlimited rights that users would have in administering self-acquired or unmanaged hardware, making possible negative consequences of 'bad' actions even bigger:

> *"Software is very common with us because right now, people are still, and my mind is still blowing up when I say this: local admin. Everyone. All employees. [...] So we have employees coming in like a revolving door with infected laptops? Well, surprising..." [Expert 01]*

6.1.2 Attack Paths

In the extensive data that result from the interviews, it is not only possible to connect all different occurrences to their respective vulnerabilities, but also to how these can be exploited, what incidents they could cause, and which asset could eventually be damaged. This is what in the CORAS method is called a attack path. Although making all possible connections would be an interesting exercise, it does not provide the right context on relevance or importance. The resulting diagram would be complex and cluttered and would not do justice to the rich context that the qualitative data provides. Therefore, we chose to highlight the most relevant attack paths with their connected concepts. Relevancy, in this case, will be defined as the paths resulting from the most grounded (most mentioned by the experts) and most dense (related to most other codes) vulnerabilities, as determined in the interviews and the section above. We take the sum of these metrics as a measure of relevancy. The top vulnerabilities when sorting by groundedness were the same as when sorting by density, except for the vulnerability of 'No contract with the supplier', which was often mentioned, but since this was mostly in a privacy context, it was not connected to many other concepts in the security-focused threat scenarios.

Figure 17 shows the possible paths related to **outdated software** that emerge from the co-occurrence of codes in the interview. The unpatched vulnerabilities that outdated software contains can allow a malicious actor initial remote access, especially when the vulnerable device is reachable via the network. If remote access is not possible, it can be infected via for instance a USB stick or email attachment, for which outdated software can also increase the chances of success. After this initial access, a malicious actor can either directly access and exfiltrate potentially sensitive data or move laterally through the network to more sensitive data and more important servers. Although steps can also be skipped depending on which systems are vulnerable. Besides the accessing and exfiltration of data, results could be unavailability caused by outages, or caused deliberately by the malicious actor if they deploy ransomware. Ransomware and extortion points to all these consequences since they were associated with all of them by the experts.

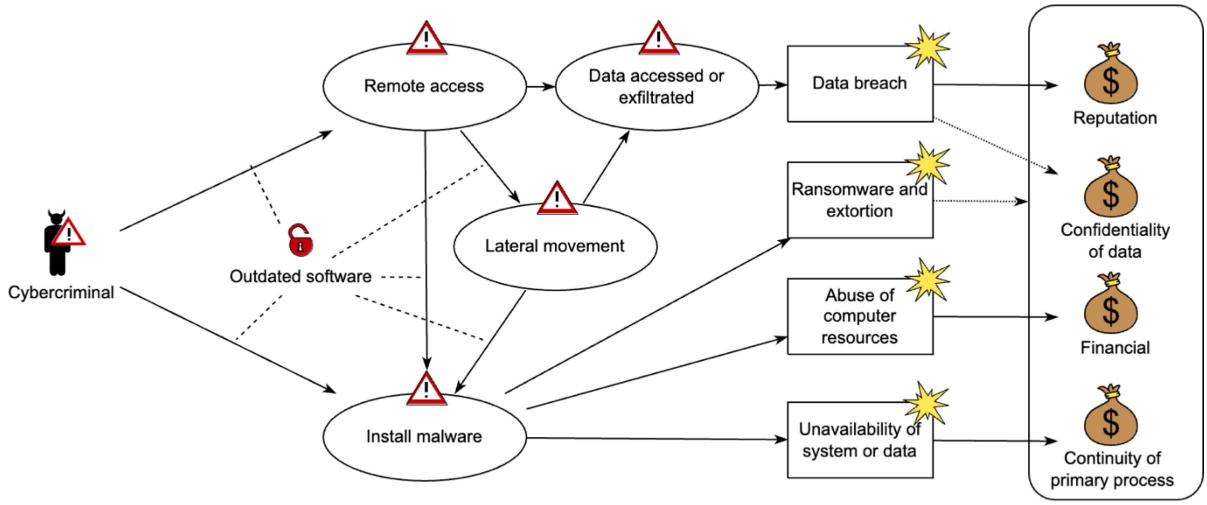


FIGURE 17: CASE 1: ATTACK PATHS RELATED TO OUTDATED SOFTWARE

We chose to create a diagram with the related threat, scenarios, incidents, and assets, while focusing on a single vulnerability. Although these components and the relations between them often have more associated vulnerabilities, we chose not to add theme here to reduce clutter. Still, the examples mentioned in the data are numerous:

- Lack of a firewall/antivirus contributes to the likelihood of malware being installed successfully on a shadow system.
- Lack of suitable access control, for instance when accessing a (shadow IT) server from a workstation, can speed up the lateral movement through a network.
- Lack of authorization could make it easy for a malicious actor who compromised a ‘normal’ user's device to quickly get to more sensitive data.
- Lack of logging on a shadow system makes it harder to detect or revert the actions of a malicious actor.

This list of vulnerabilities that cause scenarios to *not be stopped,* as opposed to vulnerabilities that can be exploited to *start* a scenario, relates our results back to (Rakovic et al., 2020), who saw ‘existing security policies being evaded’ as the main problem of shadow IT. However, vulnerabilities associated with shadow IT can not only enable threat scenarios, but they also increase the negative effects of unwanted incidents *if* they happen. A lack of encryption on shadow devices makes the effect of compromised data storage worse, while a lack of backups could severely increase the effects of sabotage, ransomware, or unavailability. Overall, this attack path of deliberate criminal activity associated with the vulnerability of outdated software is the most comprehensive, with potentially the widest impact. For easy later reference, we call this set of attack paths 'Case 1'.

The vulnerability of **lack of control of data** is phrased broadly. In this second set of attack paths, visualized in Figure 18, the focus is on the incidents that would be possible with cloud services or mobile devices outside of the network. This is the context in which the vulnerability is most mentioned. When the data is 'not controlled', a possible threat scenario is access by the third party storing the data. Such access or even commercial use could damage confidentiality or integrity. Integrity is used because the lack of contractual agreements about the data, combined with a commercial provider could lead to functionality and access being limited:

> *"Well, let's suppose you get three [environments] for free. Then they suddenly make that only one, while you have three classes. Then, as a teacher, you have a problem. And it's a very… If you think 'I want to make money with my tool' then these are very logical steps. To first let people have a good look around, and then you remove half of it and then…" [Expert 06]*

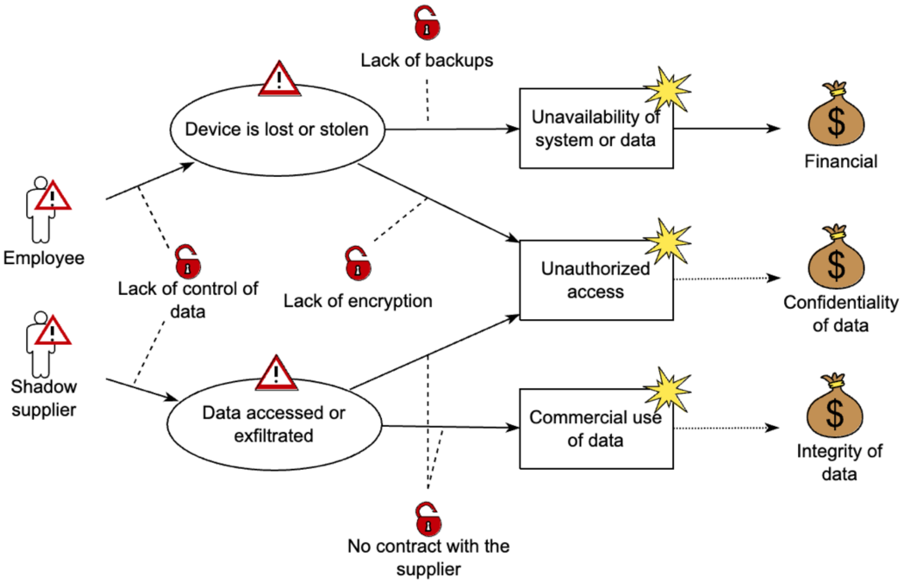


FIGURE 18: CASE 2: ATTACK PATHS RELATED TO A LACK OF CONTROL OF DATA

When the 'uncontrolled' data is on a device, the loss of that device is the primary threat. Consequences of loss are increased when the device is unmanaged, resulting in unavailability or unauthorized access to the data. When this data contains credentials, it will enable the more deliberate attack of Case 1 in Figure 17 as well. This shows once more how interconnected the different scenarios are, and how the most serious incidents are always a chain of scenarios and incidents. Again, the potential impact of these effects could be worsened by other vulnerabilities associated with shadow IT, such as lack of encryption and a lack of backups. We will refer to this set of attack pats as 'Case 2'.

The last attack path presented as a diagram in , Case 3, focuses on possible paths of a threat from *within* the organization. Although the privilege escalation that is made possible by a **lack of authorization** in a system can be related to the Case 1: actions performed by a malicious actor with financial motivation, the incidents presented in this case are especially relevant between different individuals in the organization. For example, a student not needing authorization to see staff data or students being able to access each other's environments (as exemplified in Section 5.2.2.2). Possible incidents are abuse of resources by users who are not authorized to do so, access to data that should not be possible or even fraud with (student) data. An example of abuse of institution resources, given by Expert 04, was the mining of cryptocurrency on institution resources.

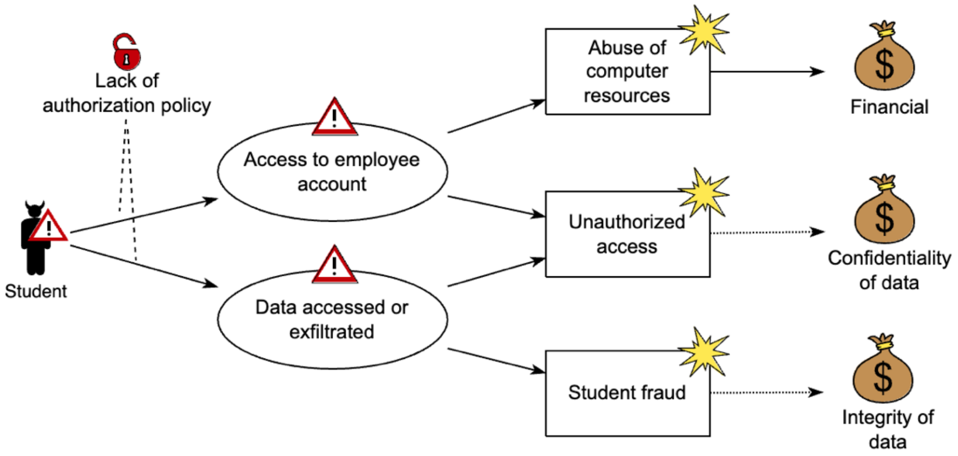

Lack of authorization policy
Student
Access to employee account
Data accessed or exfiltrated
Abuse of computer resources
Unauthorized access
Student fraud
Financial
Confidentiality of data
Integrity of data

FIGURE 19 CASE 3: ATTACK PATHS RELATED TO A LACK OF AUTHORIZATION POLICY

Expert 10 introduced the methaphor of guarding a building with doors and windows. In physical security, you want to guard and lock the doors and close the windows when there are valuable assets in the building. With shadow IT in the organization, a security officer doesn't even know where the doors and windows are located. To take this methaphor even futher: with outdated software, the windows could be broken or opened easily, while a lack of access control simply means some are open. With a lack of authorization, the same key works on every door in the house, while you probably want to keep different people in different areas. Meanwhile, the lack of control of data could mean that the security officer isn't even sure in which building the assets that should be protected are stored.

No visibility of vulnerabilities is the fourth vulnerability when ranked by the specified relevancy metric. However, when reading the coded interview segments in context, it turns out that these invisible vulnerabilities are often there because of outdated software with known security issues, and the consequences are the same as in Case 1, just viewed from a slightly different perspective:

> *"And then if you have such a Log4j thing, and the message of 'you really should check everything for Log4j' doesn't reach those people, then you could have a thingy there. That could then be another steppingstone to threaten your infrastructure from there. [Expert 02]*

A tangent to the 'no visibility' problems the interviewees raise is the **lack of logging**. Lack of logging is not a vulnerability that can be exploited, but it prevents the organization from noticing malicious activity or unwanted conduct by their own staff or students. This adds to the sense of 'lack of visibility' for IT staff on shadow IT:

> *And that's a shame, because hey, we do pay quite a bit for [a SIEM service], for example. And yeah, if we then have devices that just don't report to it, that's a waste. [Expert 03]*

When working out the different attack paths, the overarching nature of some of the vulnerabilities becomes clear. A good example is the **not actively managed** vulnerability. Expert 07 noted: "*Actually, it is not so much about shadow IT, but poorly managed shadow IT that is the problem."* Badly managed, in their view, meant that software was not updated regularly, and bad passwords were used, all while gathering and storing too much data. These three components of bad management can all be related back to previously discussed issues: outdated software, lack of access control, lack of control of data and out-of-scope privacy concerns. In the context of unmanaged devices, the lack of storage encryption, logging and updates were additionally mentioned as specifications of badly managed. This would again allow us to relate the vulnerabilities to known attack paths; Case 2 in:

> *"We should just be able to say: yes, that phone is properly secured, it has an adequate PIN, the storage is encrypted, etcetera, etcetera. We can wipe it remotely, things like that. We can't guarantee things like that with a personal phone. So, if that personal phone is nicked, then, yes, we immediately have a data breach. We can't rule out that people didn't get to that [institution] information." [Expert 07]*

The last path we want to elaborate upon is the one that exploits **re-used passwords**. This is a vulnerability unique to shadow IT since all interviewed institutions had a SSO implemented for the services managed by the organizational IT department. For these managed services, students and staff could use the same password and the institution was in control of its security. The potential problem with a re-used password would be a data breach at a shadow IT supplier, causing the re-used password to become public and compromising institution account security. The possible consequences would then again be like Case 1 in Figure 17, since it leads to a circumvented access control. Some experts made the nuance that a leaked password does not always have to cause incidents if, for instance, multifactor authentication is in place. In such a case, a leaked password alone would not be enough to access a user account. This treatment scenario will be concisely covered in Section 0

When exploring the possible attack paths from the last four vulnerabilities, we showed that these could all be related back to one of the first three cases that emerged. Case 1 is the very elaborate view of the different actions that could be performed by a deliberate malicious actor. It shows how shadow IT can create vulnerabilities allowing for the steps within the cyber kill chain. Ransomware could be the ultimate incident, but different threats to data confidentiality and availability are less complex possible paths. Case 2 focuses on the loss of control of data and what could happen when the data storage is lost, or data stored in a cloud environment is accessed by third parties. Case 3 focuses on threats within the organization and unauthorized use of resources, unauthorized access, and manipulation of data. As noted, before, all of them can be connected back to each other and extended, but these are the three main scenarios that most shadow IT vulnerabilities relate back to.

6.2 Prevention, Detection and Mitigation of Cyber Threats

Although this thesis' research questions do not focus on combatting shadow IT attack paths, the interview data does allow for some discussion of treatment scenarios and governance approaches. A treatment scenario is the term used in CORAS for taking measures that can reduce the risk of a vulnerability or incidents. During the interviews, experts were keen to point out how shadow IT could be prevented, how vulnerabilities could be fixed, or at which stages of an attack path certain measures could prevent damage. A complete mapping of the uncovered threats to possible mitigations would not be possible with the number of observations made in the interviews and dilute the focus of the thesis (it was not the focus of interview questions nor coding), but some observations by the experts will be shared. Although the selection is made based on the judgement of the researchers, they are supported by expert observations and assessments. The first section will present some specific mitigations of vulnerabilities and threat scenario risks, while the second section will then discuss the different possible approaches that emerged from the interviews and what we might be able to learn from the current ways in which shadow IT is governed in HEIs.

6.2.1 Treatment Scenarios

In terms of technical measures, network segmentation and the implementation of multi-factor authentication (MFA) for user accounts are clear treatment scenarios mentioned by all but two of the experts. Network (micro)segmentation or even air-gapping can prevent remote access or lateral movement to and from vulnerable devices, significantly reducing the impact of the 'deliberate malicious actor' attack path. It also reduces the possibility for malware on self-acquired hardware to spread or perform malicious actions like scanning the network or moving laterally. Expert 02 discussed how mitigating shadow IT risks is similar to mitigations against other scenarios of hows a cybercriminal can gain access to the institution's network:

> *"Another new [phishing] technique will come along, and someone will fall for it because, well, whatever. So the strategic thing to do is -you know it's going to happen- make sure the impact if it happens is as small as possible. So you have to make sure that that end user can do little harm when they click on that phishing email. Well, so that means for username password, you must attach multifactor to that to mitigate it a bit, you have to get rid of the local admins, you have to do microsegmentation, so that from a single workstation they cannot explore the whole network. Well, those measures are almost the same for that Shadow IT, as for clicking on a phishing email. It's going to happen. And then if you look at the measures, I think they are, well, I hardly see any nuance difference in the measures."*

Many of the institutions already put student or BYOD devices in a separate (virtual) network, something that can be extended to other unmanaged devices. Enforcing MFA limits the effect of leaked credentials so a leaked password alone cannot compromise an account's security. Furthermore, MFA can raise barriers for actors trying to access official systems from shadow IT if they already have a drop-out point on a compromised device. Expert 09 summarizes the technical measures needed to get back 'in control' as follows:

> *"[You have] to think about how you move towards a zero-trust environment, where you assume everything is suspicious, and also properly verify digital identity with biometrics, go passwordless. So you're thinking of a completely different way of working where you're just sure of that identity, of who you're dealing with."*.

The risks of self-installed or self-made software on managed devices can be reduced with appropriate endpoint detection and response tools. Even when the devices themselves are unmanaged, malicious activity can still be detected and responded to at the network level or when unmanaged devices perform actions on a managed server. Besides the traffic, a network can be preventively scanned both from the inside and from the outside to detect vulnerable or misbehaving shadow IT. Institution 07 was successful with an approach to governance of shadow IT that had network monitoring as a key component:

> *"Yes, and we scan the network regularly, so if we come across things like [compromised devices], then.... Look, sometimes a new device like that comes in, is quickly connected, without them requesting a separate connection for it, for example. You come across things like that. [...] We use [brand name] as a security system, intrusion detection, and protection system on our network. And it has been able to stop quite a few attacks over the years we have been using it. So in that respect, I am less afraid than at an average other institution."*

To prevent shadow copies of data on unmanaged hardware, Expert 08 suggests that software used in the institution should be designed to not store unencrypted files locally, to not leave traces on the device it runs on, and to not give possibilities to export the data to unknown locations. Expert 09 adds that with mobile device management (MDM), an institution can limit the extent to which sensitive data can be stored and spread on mobile devices, even if the devices are not managed by the institution. Although they also mention the challenge of resistance from decentral departments:

> *If you lose your managed laptop, then you can sleep easy. If you lose your [unmanaged laptop], then sorry, you have to solve it yourself. We are trying to change that with different programs, so mobile device management for example. But you notice, that because those devices are actually owned by the [department], you do get resistance to those kinds of efforts.*

To prevent the usage of unapproved cloud services, multiple strategies were suggested and already adopted. When cloud services are used within an organizational unit, repeated talks with directors was mentioned as an effective solution by Expert 03:

> *"We did take a very clever approach to ensure that those [department directors], you know, within those [departments], who really deal with it, that they were just made responsible. And then you saw, really after a couple of times making it very clear... [...] We always do a bit of ranking the stars as well, which is also very nice. say, look they are doing so well, you are over there somewhere. [...] I see all kinds of risks here. It does seem like a problem to me, but yes, just see what you are going to do with it. And that's been a really good move, to make them responsible and make them, let's say, also start to feel like: well *** I have to do something about this."*

Decentral departments never deployed shadow IT with bad intentions, but instead, it was deemed necessary because the official system was not familiar to them or supported well.

Several institutions shared positive outcomes of asking questions about where to store data and which services to use in existing procedures such as GDPR procedures, data management plans for researchers, data risk assessments for supporting services and education, or financial processes. If these processes were well embedded in the institution, adding a security aspect to it could make users aware that they have to take responsibility in choosing a secure service:

> *"Well, fortunately, we have data stewards. So you always have to submit a data management plan, research management plan, they discuss things with you. [...] There are all kinds of conditions you have to meet in order to do research, then you can easily add some conditions." [Expert 06]*

> *"Well, we also want for a process that already exist, for example the procurement process, that a security component is deployed there, right? That that is safeguarded in there, instead of us coming up with separate processes everywhere" [Expert 03]*

But besides technical, organizational, and awareness solutions, one overarching measure remains: institutions can ensure that users' needs are met by an official system and that they are pleasant to use. Many of the experts mentioned the difficulty of IT management in a HEI, and how personal freedom, openness and decentral management made their end-users difficult to please. Supplying managed devices and keeping people's work within the official applications was generally considered key to securing the institution and making them pleasant to use might be the only way to get people to not switch to a personal device or self-acquired service. HEIs should ensure there is a way for researchers and innovative educators to fulfil their complex IT needs and ensure researchers can work together across different organizations, so they don't have to resort to that cloud tool with their private email address. Expert 05 also emphasized the importance of guidance in the organization that help end-users find their way towards official systems and signal changing needs for services:

> *"And we are also pushing [departments] to play a role in this with [IT] coaches and little teams that will also think within such a [department] about: yes, what do we need? What will we face in the future? What is our teaching staff working on? What IT support do they have? That starts slowly, starts growing and, yeah, starts to work. I think that will become the best solution or the biggest solution for shadow IT: to just have that conversation [...] so that people can voice their desires and needs somewhere."*

In the interviews, a pattern became visible where the security measures in place seemed to determine what experts saw as the main problem of shadow. Institutions with rigid network segmentation, like 09 and 11, were not too worried about self-acquired devices, while institutions that had most employees on managed devices, like 03, 08 and 10, had little to worry about self-installed applications. If procedures for data classification were in place and complied with, cloud storage was barely a cause for concern, as examples from 03 and 08 show. Although it is difficult to judge when the experts were rightfully confident and when they might be naïve without thoroughly assessing their IT infrastructure, the difference across institutions was a clear signal. Although specific vulnerabilities and threats of shadow IT might be easy to mitigate, the broadness of the issue causes its complexity.

6.2.2 Governance Approaches

Besides the specific mitigations mentioned by the interview participants, overarching approaches emerge. It seems like there are two ways shadow IT in an organization can be approached. The first is the 'strict' approach: keeping all staff on managed devices with only a set of packaged applications, all student devices in a strictly separated network, all unknown devices out of the network, with institution data not being able to be accessed from private devices. Although not a single institution follows this approach completely, its principles serve as an example to oppose the other approach. It does resemble the situation of Institution 11, who experienced no issues with shadow IT, as exemplified in Section 5.2.2.

The second approach assumes shadow IT to always be present to a certain extent. Institutions taking this stance to monitor their systems, networks and (if possible) end-user devices for malicious activity but do not restrict as much. Self-installed applications are allowed and accounted for in other policy choices, for instance, by making local admin rights available, albeit temporarily. The goal here is to provide a path for the wide variation in needs while keeping users on board with official IT for as long as possible, following the idea that it is better to have everybody on board with a lightly managed device than to heavily manage devices and have users resort to shadow IT as an effect. This 'keeping users on board' seems to be especially important for hardware, since it allows preventive software measures to be taken:

> *"So yes, I realize that that it is used. We do have [antivirus] running on all the managed and also on a large part of the unmanaged workstations. So if there is a problem with an application or something like that, it does get detected."* [Expert 07]

This approach views staff and students more like potential customers than as a given consumers of services. The user-centric approach could go as far as taking over the management of self-acquired hardware *and* building a better alternative, as demonstrated again by Institution 07:

> *"I do notice that in the past, we had a lot more shadow IT. See, it used to be very easy for a user to walk to [an electronics store] and buy [a network attached storage] system*

> *and put it under their desk. [...] At one point we adopted the policy of: yes, we can also manage that kind of thing. We are not going to get fussy about that. [...] If you want, we also manage your [storage] [...]. But next time, preferably don't do it."*

> *"That is also why we have deployed large-scale central storage in various flavors. Very good ones with redundancy, backup, ransomware protection, and things like that. A very cheap one, which is cheaper than any commercial provider."*

Besides general attitudes towards shadow IT, some specific governance strategies emerged from the interviews. Institutions 07, 08, and 11 govern or prevent shadow IT via finances. The idea being that all IT purchases require budget approval and that shadow IT would simply not be approved. They do recognize that this is only effective for a specific subset of IT since it does not prevent cloud applications, which often offer (some of) their services for free. Other institutions specifically mentioned the lack of control over researchers' budgets (and emphasized the importance of not having this control) or that users would simply pay for applications or hosting with their own money, as happened in institution 09, or this example from Expert 03:

> *"So sometimes I hear, say, that then, I don't know, a credit card is pulled out somewhere to just purchase an application, or something. But well, hey, there could be a risk in that. But generally speaking, that application then just runs nicely on [institution] laptops that have been managed. Yes, we have quite... I worry less about that in that case."*

A different approach, focusing on finances that did seem to be more broadly applicable, was developing a competitive IT solution at the central level or pre-financing these. This was for example mentioned Expert 07, as cited in the beginning of this section. When the official solution is good, cheap, or already paid for, the choices of researchers with control over their own budget can be steered towards a safe solution.

Governance of decentral units with autonomy and control over their own budget and IT environment was a repeating theme. When risks of IT were still primarily financial, they could be assessed and mitigated or accepted within a department, faculty, or institute. But recently, it seems like IT security risks are too much shared within the institution to be left to the departments, raising questions about what level of control the central IT department should have over these systems. Differences in approach between institutions in this regard are very big, and the data from our interviews is not sufficient to cast a judgment on these approaches to dealing with organizational sub-units. Asked whether it would be the goal to centralize all IT, Expert 10 answered the following:

> *"Well, maybe for me [...] as somebody who is crazy about their profession, so to speak, yes, you would prefer to do that. Only well, is that realistic in a sector that is obviously hugely dynamic and wants to grow with the market quickly? So my philosophy is: look, I'm fine with leaving that freedom with institutes. But I can only give them that freedom if they also comply with the regulations. So you have to train a lot, facilitate, so they actually do so."*

7 Discussion and Interpretation

After we conducted the interviews and structured the results into the different attack paths, the thesis as a whole is discussed, and findings are put into perspective. This chapter presents the results from all previous chapters together, summarizes key findings, translate them to contributions for science and practice, and reflects on the research' limitations.

7.1 Key Findings

We will present the key findings of the different chapters by answering the sub-questions as posed in Chapter 2.

SQ1: *"How is shadow IT defined and differentiated from similar concepts in literature?"*

This question was answered with a tertiary study into existing literature reviews about the topic. It resulted in a clear understanding of the concept and a definition by Haag & Eckhardt (2017) that could be used throughout the thesis:

> *"[Shadow IT is] hardware, software, or services built, introduced, and/or used for the job without explicit approval or even knowledge of the organization."*

In research, shadow IT is often related to, or confused with, concepts like IT consumerization, lightweight IT, BYOD, and many more. The concept of shadow IT and the related research can and should be delaminated because it is about the occurrences of unofficial IT as an *outcome*, not the culture or policy by which unofficial IT gets invited into the organisation. The IT hardware, software and services are unofficial in the sense that the organization does not approve of, or knows about them. The literature review further revealed that current research is often about causal factors and consequences of shadow IT and that governance has seen a big increase in interest recently. Security concerns are the most discussed risk according to the literature reviews.

SQ2: *"What cyber threats are commonly associated with shadow IT in literature?"*

A second literature review gathered the existing knowledge about cyber threats of shadow IT in a structured mapping study. The surprising outcome was that research on this topic seldom goes into practical or technical details. Although a lot is written about how shadow IT causes management problems like a loss of control or lack of compliance, specific vulnerabilities, or scenarios of what can go wrong are seldom mentioned let alone empirically supported. This lack of a problem statement makes it difficult to assess what an organization could do about shadow IT, especially if preventing it with strict limitations is not a realistic approach, as is the case in HEI given the values of academic freedom and openness. This literature review prompted the focus on the 'how?' aspect of shadow IT cyber threats for the rest of the thesis.

SQ3: *"What types of shadow IT are observed in Dutch higher education?"*

This question was the first of three questions answered by the qualitative analysis of interviews. To answer this question, we extracted occurrences and their relative importance from the interviews with eleven information security professionals from HEIs. After gathering the observations, we divided these into a topology of shadow IT, adapted from Mallmann et al. (2019). From this, we concluded that for almost every type of IT a shadow version can exist. Software, locally installed or in the cloud, is most top-of-mind for the interviewed experts, accounting for more than two-thirds of the mentions. Especially the mentioned cloud software had a big range in instance size, from a small web-tool used in education to complete productivity suites for departments or faculties. Besides these types of software and services, institutions frequently mentioned spreadsheets, duplicating data and functionality of official systems. A type very specific to research were the self-developed applications, not seen as shadow IT because they should have been procured centrally but because there was no oversight on risks. Self-acquired devices added variety to the IT landscape. Where most software would be piled into one category by the experts, many different types of hardware would be mentioned. However, unmanaged PCs, research equipment, server hardware and mobile devices were most frequently mentioned.

SQ4: *"Which cyber treats related to shadow IT are perceived by experts?"*

The experts we interviewed listed a multitude of different ways in which shadow IT could enable or worsen cyber threats to the institutions. We captured the emerging threat components using the lens of CORAS (Lund et al., 2011) to operationalize the concept of a 'cyber threat' as threat (actors), vulnerabilities, threat scenarios, unwanted incident, and asset (damaged). Most experts phrased perceived threat components in terms of vulnerabilities, threat scenarios, and unwanted incidents, while the threat and asset were often left implicit. The main concerning vulnerabilities were outdated software, lack of control of data, lack of access control and a lack of authorization policy in shadow IT. The most mentioned threat scenarios were remote access into the network, access into employee accounts, installing malware and accessing data. This could cause incidents like unauthorized access, data breaches, leaked credentials, commercial use of data, and ransomware and extortion. Again, we presented the different components as-is.

SQ5: *“Which recurring patterns can be identified in these occurrences and cyber threats?”*

To gain further insights into the interview data, we thereafter linked the different shadow IT occurrences and threat components to each other. This was done based on co-occurrence in the interview transcripts and relations explicitly mentioned by the experts. The number of mentions by experts combined with the number of related occurrences gave us a metric of relevancy according to which we further explored possible attack paths enabled by these vulnerabilities. We identified three main ‘cases’ of attack paths, to which many other scenarios and incidents could be related as well. The first case is that of network infiltration by a cybercriminal for financial gain. It is associated with the classic cyber kill chain and the most elaborate, allowing for severe outcomes like ransomware and large data breaches with damage to reputation and the institutions finances, among others. A second case focused on the consequences of a lack of control of data, when a (shadow) device was lost or cloud storage was accessed by a (shadow) supplier. The resulting unavailability, unauthorized access, or commercial use of data can cause financial damage or compromises integrity or confidentiality of data. The last identified case was that of an insider. It specified how a lack of authorization policy, the fourth-most mentioned vulnerability by experts, allowed different parties within the institution to access each other’s data. Besides being vulnerabilities that can be exploited by a threat, the potential risks of shadow IT were often related to how it would worsen the effect of threat scenarios or prevent their detection.

The structuring of emerging codes also allowed us to identify relevant treatments. MFA, network (micro)segmentation, endpoint detection and response tools, security incident monitoring and both internal and external network scanning are among the technical measures reported as effective to limit the risks of shadow IT by the institutions. Configuring the applications to not leave data on unmanaged devices and providing guidance on the use and wishes of innovative applications are among possible treatments as well. Institutions report satisfactory results when making security checks part of existing processes, especially in the context of preventing data from being stored in the cloud or on unmanaged devices.

Both to prevent shadow IT, and to prevent the above-mentioned lack of detection and containment possibilities, it is still paramount to keep users on managed devices and solutions whenever possible. The simply said but hard to execute treatment here would be to make the official solutions convenient to use, and to always account for -the sometimes exotic- user needs of students and staff in higher education. By ensuring a solid basis of (lightly) managed devices and a well-secured network, other shadow IT problems can be prevented, detected or mitigated.

7.2 Evaluation of Results

The initial starting point of this thesis was a curiosity if shadow IT in HEIs was a phenomenon that required special attention and governance approaches. After conducting literature research in the first phase, a lack of practical understanding of shadow IT's cybersecurity consequences in the literature prompted the threat modelling focus of this rest of the thesis. During the interviews and by responses from stakeholders the relevancy was quickly confirmed since many were interested in participating and hearing back about the results. Higher education turned out to be an interesting sector to study this phenomenon. Because of (academic) freedom, a drive to innovate, openness and diverse user needs, a standardized IT environment is simply not sufficient in this sector. Almost every institution interviewed expressed the uniqueness of their situation and management challenges. From the modeling of threats, however, it does not immediately become clear if shadow IT is a justified focus for research on cyber threats. This might explain why so little shadow IT research takes this specific perspective.

Relating the modeled threats to the prevention, detection and mitigation measures suggested by the interviewed experts, it seems that many of the problems of shadow IT can be solved by already known, broadly applicable cybersecurity measures and best-practices. The fact that the perceived problems of shadow IT were different in each institution and highly reliant on the implemented security measures contributes to this idea. Shadow IT has the possibility to cause big security problems, mainly when already known best practices for cybersecurity are not implemented. Other problems expressed by experts could often be related to abuse of official IT or simply a lacking IT environment, for instance because of staff shortages. It is already known how to deal with cybersecurity problems of shadow IT.

What remains are the challenges of preventing shadow IT from becoming too big, and (security) governance in an organization with partly independent decentral units like faculties Regarding the second issue, the interviewed experts had little solutions to mention. This is a clear need for future research, although it can probably be related to the concept of 'business-managed IT', identified in the literature reviews of this thesis (Klotz et al., 2020). On the topic of prevention, many suggested policies and best practices emerged, which we will cover in Section 7.4.

7.3 Scientifical Implications

This thesis project gathered a large set of empirical data on the phenomenon of shadow IT in a previously unexplored context, namely HEIs. The interview transcripts and the associated codebook are both published, not only as supporting evidence for this research, but also for further research into this topic. Because the main focus in this project was to extract observations on occurrences and threats of shadow IT, these themes were reported in a structured manner. The interviews also resulted in an unexpected wealth of data on practical solutions to shadow IT problems.

To the best of our knowledge, this paper was the first to not only elicit perceptions on shadow IT occurrence, but also work out the related cyber threats and their consequences in a practical and actionable way. We think this perspective is very beneficial for further work on governance of shadow IT as it provided previously missing observations to link occurrences to threats and their possible treatments. This approach showed how a 'deep dive' into practical details can allow scientific research to emerge with practical solutions and generalizable insights.

The evaluation of results, provided a critical reflection on shadow IT as area of research, concluding the perspective mainly has a lot of value when looking at its prevention, which can be an important part of governance. Besides, for scholars working on risk quantification or detection via technical means, it is very relevant to consider a focus on shadow IT. However, from a technical perspective, no specific mitigations are required for cyber threats related to shadow IT and most professionals knew what was lacking in terms of security measures. All actions an institution can take to prevent damage related to shadow IT are, or should be, part of normal security operations.

7.4 Recommendations for HEIs

The main broad takeaway for practitioners is that, unless you are a small and orderly institution, nobody should have the illusion that a strictly managed IT workspace and environment is a good match with higher education. Shadow IT is inherent to higher education and trying to ban it completely would be naïve. Institutions that report little problems with shadow IT are the ones who've accepted its existence and take a pragmatical stance. Further recommendations are:

- **Adopt existing cybersecurity best practices** to mitigate the main threats of shadow IT. These include MFA for user accounts, proper password policies, privileged access management, network segmentation, adopting zero-trust principles, having well-configured antivirus and firewalls, endpoint detection and response, and security incident monitoring.
- **Scan and monitor the network**, both from the inside as well as from the outside. This can help identify vulnerable or problematic shadow IT instances, without the need to completely ban every strange device.

- **Provide a managed workplace that people will actually use.** It can be beneficial to have a workplace solution in between a completely locked down, managed workplace and unmanaged (shadow) IT. HEI user needs will always be exotic, and by having a defined path for unconventional needs, you can prevent users from having to resort to shadow IT while keeping control of essential measures such as endpoint detection and response.
- **Limit the extent to which applications store data on personal devices** when accessed via them. Disable unnecessary export or download options and use mobile device management tools when necessary to prevent your applications from insecurely storing data on private devices.
- **Have (repeated) conversations with directors** about their (shadow) IT in case the responsibility for IT security is delegated to a lower level. Moving responsibility for cybersecurity to the 'first line' in no way guarantees security of their self-procured IT or usage of official systems.
- **Ensure guidance is available for educators and researchers** who want to use new, innovative (and unprocured) services. Telling them it is not allowed is unlikely to secure the institution. Coaching them in responsible usage, while at the same time taking the opportunity of noticing changing needs from the most innovative users has shown promising results in other institutions.
- **Think of procedures in which security components can be added**. Data classification for GDPR-compliance of central services, Data Management Plans for research data, *etc.*; can include checks on storage security, while research and education support centers can help disseminate expertise on responsible choosing of systems.
- **And maybe most importantly: think about the considerations end-users have when choosing between going the official route or using something different**. Create financial incentives to use official services, for instance by having them be pre-financed or at least cheaper than commercial alternatives. And most importantly, offer services that align with the needs of users, and have a pleasant user experience. This might not usually be within the scope of IT security teams, but it is nevertheless **essential** if we want to keep a grip on the IT environment.

7.5 Limitations of the Study

Several limitations of this study can be named, with a first point of discussion being whether the study gives a complete overview of all threats related to shadow IT in HEIs. The achieved code saturation, as discussed in Section 5.1.5 provides evidence that all the relevant recurring themes were retrieved from the experts. However, as mentioned in Section 6.1.1, we cannot ensure that all relevant *links* between the codes also surfaced in these interviews. The only thing we can be sure of is that the most important attack paths were mentioned because we can expect these to be more top-of-mind for security experts. This concept of 'importance' is another limitation. The number of times, and by how many experts an occurrence or threat is mentioned gives a good impression of how important it is for the cyber threat landscape, but is not the same as 'how often is this type of shadow IT used' or 'how likely is this cyber threat'. It remains a proxy. Quantifications of occurrence and risk were not part of the research setup and should not be inferred from our results.

Another limitation might be in the chosen modeling approach. Using CORAS enabled the detailed modeling of cyber threats and allowed us to relate the occurrences to threats and further consequences, but it also made the coding of data an exercise in semantics and ontology. Especially the properties of shadow IT that *worsen* the effect of threats was difficult to grasp in CORAS and had to be modeled as a vulnerability. It is a property of the CORAS method that there is not just one way to model a situation, but that the phrasing of the different components depends on the perspective of participants (Lund et al., 2011). The modeling choices made here were agreed upon by multiple coders and contributed to answering our research questions, but it should not be interpreted as an absolute model of reality. The aim was not to test the validity of a model and inter-coder agreement was never a goal in and of itself, only a means to ensure consistent interpretation.

For the sections on treatments and governance, we used only the results of open coding by both researchers. We did not anticipate the step from defining the problem (our main research goal) to providing solutions to be this small, so both the interview and coding setup were not aimed at retrieving data on treatments and governance, and the fact that findings about solutions would emerge was unforeseen. Therefore, we cannot ensure the same interpretation validity and completeness as for the occurrences and threats, but the input was nevertheless suitable to the more explorative nature of this section. Conclusions like 'there seems to be a treatment for every solution' or 'most measures are known best practices' are appropriate with this type of interpretation, but the results shouldn't be interpreted as a technical manual towards mitigations of threats.

8 CONCLUSIONS

Let us re-iterate the main research question:

MRQ: *"What is the role of shadow IT in the cyber threat landscape of Dutch higher education institutions?"*

To answer the main research question, this thesis investigated the definition of shadow IT, the cyber threats commonly associated with shadow IT, in what form shadow IT occurs in HEIs, what vulnerabilities and threats are associated with its occurrence, and what kind of attack paths it enables or worsens. Our results show that shadow IT is an inherent part of HEIs IT environments. It occurs in many different forms: locally installed and other types of software are the most prominent occurrences, but the variety is big, and after talking to eleven experts, there is almost no type of IT left that does not have a shadow version somewhere. It ranges in size from a single user, research groups, courses, to whole institutes, faculties, and the like. Most cyber threats associated with shadow IT revolve around the vulnerabilities in outdated (locally installed or embedded) software, and the lack of visibility thereof. Another category of vulnerabilities that raise concerns are a lack of control over data: where and how it is stored, if backups are made, and if authentication and authorization are performed was well as in official systems. We combined these vulnerabilities with the threat scenarios, incidents and assets mentioned by the experts to create comprehensive and practical overviews of what can potentially go wrong with shadow IT. This led to the identification of three 'cases' of attack paths, to which most other vulnerabilities, threat scenarios and incidents could be linked. This specific modeling of threats also allowed us to relate these threats to countermeasures. It led to the insights that the cyber problems related to shadow IT are highly dependent on the different defense measures already taken by institutions. The role of shadow IT in the threat landscape can be very manageable if the institution accounts for its existence. What remains are the challenges of allowing shadow IT in a responsible manner to some extent and preventing shadow IT when possible. To achieve this, we provided practical recommendations, while at the same time, we identify needs for future work.

8.1 FUTURE WORK

Directions for future work, directly continuing our line of research, could include the quantification of both the size of shadow IT in HEIs, and quantification of its risks. These are details that could not be provided through our qualitative research setup. Insights from such research could be used to prioritize and target efforts within the organization or sector. Based on this paper's conclusion that technical measures can provide prevention, detection, and mitigation for many of the cyber treats related to shadow IT, specific measures and the effectiveness of their implementation could be a focus for future research as well. For example: how to effectively scan for vulnerable or compromised shadow IT, what mobile device management (MDM) tools work best and how these types of controls are perceived, or where to start with implementation of zero-trust concepts in a complex environment. Given the identified importance of usability and alignment to needs of official solutions as the main preventive measure, this research additionally stresses the importance of usable security research (Garfinkel & Lipford, 2014).

Technical measures and their importance are often well understood by security experts, but the current research also emphasized the importance of IT environment usability as a factor contributing to an institution's cybersecurity. Not only does this require different expertise, but it might also require coordination at a higher level. Responsibility for security and IT infrastructure design are likely allocated with different people with different priorities. This adds an element of complexity that might require new ways of thinking and approaches to IT security governance.

A final topic in the IT governance domain, that needs more attention, is the dealing with decentral units when doing security. Our research found that shadow IT does not only emerge from individual users, as is often assumed in previous research, but also from departments or smaller groups within a faculty or institute. This raises questions about what aspects of IT can safely be left to smaller organizational units, what safeguards might be required, and how a CISO can keep the overview of their environment. This becomes especially relevant when we want to allow educators and researchers to quickly adopt new, innovative tools while ensuring security (and privacy and compliance).

8.2 Data Repository

All the thesis' data are published and can be accessed to verify validity of results and for further exploration and research. [6] This primarily includes the interview transcripts. They are anonymized using the guidelines in Appendix F. The complete project file including the interview transcript and coding can be downloaded in both .atlproj22 and .QDPL format, ensuring maximum compatibility with other qualitative analysis software packages. The metadata documentation can be found in the .README file, this should ensure the data to adhere to the FAIR -findability, accessibility, interoperability and reusability- guiding principles (Wilkinson et al., 2016).

[6] At the time of submission of this thesis, the publishing of the data repository is still awaiting the required approvals. The link to the data will be available in a later version of the thesis, published via https://edu.nl/683h9.

ACKNOWLEDGEMENTS

I'd like to thank, first and foremost, the security and privacy professionals whom I could talk to, not only for the 'on the record' interviews, but also in the scoping phase and for the pilot interviews. Besides the research results, I learned a great deal from them, and the passion with which they talk about their sector and area of expertise is contagious. I was pleasantly surprised by the interest in my research and their desire to contribute and know more about what I was doing. You made me want to work in this security world of yours and helped me get there.

Of course, this research would have gone nowhere without my supervisors and other researchers at Utrecht University and SURF. Your guidance, help and feedback has been tremendous and allowed me to improve a great deal, not only in terms of this thesis' content and process, but also as a person. Wrapping up this thesis was a big part of the work we are doing together, but I look forward immensely to the next steps we will be taking with the results, both in academia as well as in practice, working towards more secure IT ecosystems and organizations.

REFERENCES

ATLAS.ti Scientific Software Development GmbH. (2022a). *ATLAS.ti* (22.2.0) [MacOS]. https://atlasti.com

ATLAS.ti Scientific Software Development GmbH. (2022b). *Requirements For Coding*. Requirements For Coding - ATLAS.Ti 22 Mac - User Manual. https://doc.atlasti.com/ManualMac.v22/ICA/ICARequirementsForCoding.html

Bandara, W., Miskon, S., & Fielt, E. (2011). A systematic, tool-supported method for conducting literature reviews in information systems. *ECIS 2011 Proceedings*, 15.

Barbour, R. S. (2001). Checklists for improving rigour in qualitative research: A case of the tail wagging the dog? *BMJ*, *322*(7294), 1115–1117. https://doi.org/10.1136/bmj.322.7294.1115

Behrens, S. (2009). Shadow systems: The good, the bad and the ugly. *Communications of the ACM*, *52*(2), 124–129. https://doi.org/10.1145/1461928.1461960

Bongiovanni, I. (2019). The least secure places in the universe? A systematic literature review on information security management in higher education. *Computers & Security*, *86*, 350–357. https://doi.org/10.1016/j.cose.2019.07.003

Budgen, D., Turner, M., Brereton, P., & Kitchenham, B. (2008). Using Mapping Studies in Software Engineering. *Proceedings of PPIG 2008*, 10.

Bygstad, B. (2015). The Coming of Lightweight IT. *ECIS 2015 Completed Research Papers.*, 17.

Candia, T. (2021). How to Proactively Manage Shadow IT. *EdTech Magazine*. https://edtechmagazine.com/higher/article/2021/03/how-proactively-manage-shadow-it

Castillo-Montoya, M. (2016). Preparing for Interview Research: The Interview Protocol Refinement Framework. *The Qualitative Report*. https://doi.org/10.46743/2160-3715/2016.2337

Centre for Reviews and Dissemination (CRD). (1995). *Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews*. Centre for Reviews and Dissemination (UK).

Chua, C. E. H., Storey, V. C., & Chen, L. (2014). Central IT or Shadow IT? Factors shaping users' decision to go rogue with IT. *ECIS 2012 Proceedings*. 35th International Conference on Information Systems 'Building a Better World Through Information Systems', ICIS 2014.

Coen, M., & Kelly, U. (2007). Information management and governance in UK higher education institutions: Bringing IT in from the cold. *Perspectives: Policy and Practice in Higher Education*, *11*(1), 7–11. https://doi.org/10.1080/13603100601127915

Computer Security Resource Center. (2022). *Cyber Threat—Glossary | CSRC*. National Institute of Standards and Technology COMPUTER SECURITY RESOURCE CENTER. https://csrc.nist.gov/glossary/term/Cyber_Threat

Cross, C., & Gillett, R. (2020). Exploiting trust for financial gain: An overview of business email compromise (BEC) fraud. *Journal of Financial Crime*, *27*(3), 871–884. https://doi.org/10.1108/JFC-02-2020-0026

D'Arcy, P. (2011). *CIO Strategies for Consumerization: The Future of Enterprise Mobile Computing* (Dell CIO Insight Series).

de Haan, H. (Helen). (2014). Internationalization: Interpretations Among Dutch Practitioners. *Journal of Studies in International Education*, *18*(3), 241–260. https://doi.org/10.1177/1028315313496571

Digital Curation Centre (DCC). (2022). *DMPonline*. DMPonline. https://dmponline.dcc.ac.uk

FireEye, Inc. (2015). *Storming the Ivory Tower: Why Cyber Attackers Are Targeting Higher Education, and What Universities Can Do About It* (p. 12). https://www.fireeye.com/content/dam/fireeye-www/solutions/pdfs/storming-the-ivory-tower-whitepaper.pdf

Fowler, K. (2016). Chapter 1—An Overview of Data Breaches. In K. Fowler (Ed.), *Data Breach Preparation and Response* (pp. 1–26). Syngress. https://doi.org/10.1016/B978-0-12-803451-4.00001-0

French, A., Guo, C., & Shim, J. (2014). Current Status, Issues, and Future of Bring Your Own Device (BYOD). *Communications of the Association for Information Systems*, *35*. https://doi.org/10.17705/1CAIS.03510

Furstenau, D., Rothe, H., & Sandner, M. (2017). Shadow systems, risk, and shifting power relations in organizations. *Communications of the Association for Information Systems*, *41*, 43–61. https://doi.org/10.17705/1cais.04103

Garcia, L., & Quek, F. (1997). Qualitative Research in Information Systems: Time to be Subjective? In A. S. Lee, J. Liebenau, & J. I. DeGross (Eds.), *Information Systems and Qualitative Research* (pp. 444–465). Springer US. https://doi.org/10.1007/978-0-387-35309-8_22

Garfinkel, S., & Lipford, H. R. (2014). Usable Security: History, Themes, and Challenges. *Synthesis Lectures on Information Security, Privacy, and Trust*, *5*(2), 1–124. https://doi.org/10.2200/S00594ED1V01Y201408SPT011

Gartner. (2016). *Top 10 Security Predictions 2016*. Gartner, Inc. https://www.gartner.com/smarterwithgartner/top-10-security-predictions-2016

Godefroid, M.-E., Plattfaut, R., & Niehaves, B. (2021). IT Outside of the IT Department: Reviewing Lightweight IT in Times of Shadow IT and IT Consumerization. *Lecture*

Notes in Information Systems and Organisation, *48 LNISO*, 554–571. https://doi.org/10.1007/978-3-030-86800-0_39

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

Guest, G., Namey, E., & Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. *PLOS ONE*, *15*(5), e0232076. https://doi.org/10.1371/journal.pone.0232076

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, *11*(2), 181–217. https://doi.org/10.1002/jrsm.1378

Györy, A., Cleven, A., Uebernickel, F., & Brenner, W. (2012). Exploring the Shadows: IT Governance Approaches to User-Driven Innovation. *ECIS 2012 Proceedings*, 13.

Haag, S. (2015). Appearance of Dark Clouds? - An Empirical Analysis of Users' Shadow Sourcing of Cloud Services. *Wirtschaftsinformatik Proceedings 2015*, 16.

Haag, S., & Eckhardt, A. (2017). Shadow IT. *Business & Information Systems Engineering*, *59*(6), 469–473. https://doi.org/10.1007/s12599-017-0497-x

Haag, S., Eckhardt, A., & Bozoyan, C. (2015). Are Shadow System Users the Better IS Users? – Insights of a Lab Experiment. *ICIS 2015 Proceedings*. https://aisel.aisnet.org/icis2015/proceedings/ITimplementation/16

Huber, M., Zimmermann, S., Rentrop, C., & Felden, C. (2016). The Relation of Shadow Systems and ERP Systems—Insights from a Multiple-Case Study. *Systems*, *4*(1), Article 1. https://doi.org/10.3390/systems4010011

Huber, M., Zimmermann, S., Rentrop, C., & Felden, C. (2017). Integration of shadow it systems with enterprise systems–a literature review. *Proceedings Ot the 21st Pacific Asia Conference on Information Systems: "Societal Transformation Through IS/IT", PACIS 2017*.

Hutchins, E. M., Cloppert, M. J., Amin, R. M., & others. (2011). Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Leading Issues in Information Warfare & Security Research*, *1*(1), 80.

JGraph Ltd. (2022). *Draw.io* (20.3.0). https://www.diagrams.net/

Johnson, S. (2013). Bringing IT out of the shadows. *Network Security*, *2013*(12), 5–6. https://doi.org/10.1016/S1353-4858(13)70134-X

Käss, S., Godefroid, M., Borghoff, V., Strahringer, S., Westner, M., & Plattfaut, R. (2021). *Towards a Taxonomy of Concepts Describing IT Outside the IT Department*. 12.

Kitchenham, B., Brereton, P., Budgen, D., Tuner, M., Bailey, J., & Linkman, S. (2007). *Protocol for a Tertiary study of Systematic Literature Reviews and Evidence-based Guidelines in IT and Software Engineering*. https://community.dur.ac.uk/ebse/resources/studies/protocol/tertiary_slr_v2.pdf

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. *2*.

Klotz, S., Kopper, A., Westner, M., & Strahringer, S. (2019). Causing factors, outcomes, and governance of Shadow IT and business-managed IT: a systematic literature review. *International Journal of Information Systems and Project Management*, *7*(1), 29.

Klotz, S., Westner, M., & Strahringer, S. (2020, June 20). *From Shadow IT to Business-managed IT and Back Again: How Responsibility for IT Instances Evolves Over Time*.

Kopper, A. (2017). *Perceptions of IT managers on shadow IT*. *2017-August*.

Kopper, A. (2019). *From shadow IT to business-managed IT* [Hochschulschrift]. Technische Universität Dresden.

Kopper, A., Fuerstenau, D., Zimmermann, S., Klotz, S., Rentrop, C., Rothe, H., Strahringer, S., & Westner, M. (2018). *Shadow IT and Business-Managed IT: A Conceptual Framework and Empirical Illustration*. *9*, 53–71. https://doi.org/10.4018/IJITBAG.2018070104

Kopper, A., & Westner, M. (2016a). Deriving a framework for causes, consequences, and governance of shadow IT from literature. In S. D. Nissen V. Strassburger S. ,. Fischer D. (Ed.), *Multikonferenz Wirtschaftsinformatik, MKWI 2016* (Vol. 3, pp. 1687–1698). Universitatsverlag Ilmenau.

Kopper, A., & Westner, M. (2016b, August 11). Towards a Taxonomy for Shadow IT. *AMCIS 2016 Proceedings*. Americas Conference on Information Systems.

Kretzer, M., & Maedche, A. (2014). Generativity of business intelligence platforms: A research agenda guided by lessons from shadow IT. In S. L. Kundisch D. Beckmann L. (Ed.), *Tagungsband Multikonferenz Wirtschaftsinformatik 2014, MKWI 2014* (pp. 207–220). University of Paderborn.

Krippendorff, K. (2011). *Computing Krippendorff's Alpha-Reliability*. 12.

Kushner, D. (2013). The real story of stuxnet. *IEEE Spectrum*, *50*(3), 48–53. https://doi.org/10.1109/MSPEC.2013.6471059

Lund, M. S., Solhaug, B., & Stølen, K. (2011). *Model-Driven Risk Analysis*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-12323-8

Lund-Jensen, R., Azaria, C., Permien, F. H., Sawari, J., & Bækgaard, L. (2016). Feral Information Systems, Shadow Systems, and Workarounds – A Drift in IS Terminology.

Procedia Computer Science, *100*, 1056–1063. https://doi.org/10.1016/j.procs.2016.09.281

Magunduni, J., & Chigona, W. (2018). Revisiting shadow IT research: What we already know, what we still need to know, and how do we get there? *2018 Conference on Information Communications Technology and Society (ICTAS)*, 1–6. https://doi.org/10.1109/ICTAS.2018.8368735

Mallmann, G. L., de Vargas Pinto, A., & Maçada, A. C. G. (2019). Shedding Light on Shadow IT: Definition, Related Concepts, and Consequences. *Lecture Notes in Information Systems and Organisation*, *31*, 63–79. Scopus. https://doi.org/10.1007/978-3-030-14850-8_5

Managing and Sharing Qualitative Data. (2019). [Data set]. figshare. https://doi.org/10.6084/m9.figshare.7637288.v1

Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does Sample Size Matter in Qualitative Research?: A Review of Qualitative Interviews in is Research. *Journal of Computer Information Systems*, *54*, 11–22. https://doi.org/10.1080/08874417.2013.11645667

McKenzie, R. G., & Ruthberg, Z. G. (1977). *Audit and evaluation of computer security* (NBS Special Publication, p. 268). U.S. Department of Commerce, National Bureau of Standards.

Microsoft Corporation. (2022). *Microsoft Word* (16.0.15728.41006) [Online]. https://word.office.com

Orduna-Malea, E., Ayllón, J. M., Martín-Martín, A., & Delgado López-Cózar, E. (2015). Methods for estimating the size of Google Scholar. *Scientometrics*, *104*(3), 931–949. https://doi.org/10.1007/s11192-015-1614-6

Peter, S., & Deimann, M. (2013). On the role of openness in education: A historical reconstruction. *Open Praxis*, *5*(1), 8.

Petticrew, M., & Roberts, H. (2008). *Systematic Reviews in the Social Sciences: A Practical Guide*. John Wiley & Sons.

Raden, N. (2005). *Shedding Light on Shadow IT: Is Excel Running Your Business?* 11.

Rakovic, L., Duc, T. A., & Vukovic, V. (2020). Shadow it and ERP: Multiple case study in German and Serbian companies. *Journal of East European Management Studies*, *25*(4), 730–752. https://doi.org/10.5771/0949-6181-2020-4-730

Rakovic, L., Sakal, M., Matkovic, P., & Maric, M. (2020). Shadow IT - A Systematic Literature Review. In *Information Technology and Control* (Vol. 49, Issue 1, pp. 144–160). KAUNAS UNIV TECHNOLOGY. https://doi.org/10.5755/j01.itc.49.1.23801

Schlagwein, D., Conboy, K., Feller, J., Leimeister, J. M., & Morgan, L. (2017). “Openness” with and without Information Technology: A Framework and a Brief History. *Journal of*

Information Technology, *32*(4), 297–305. https://doi.org/10.1057/s41265-017-0049-3

Silic, M., & Back, A. (2014). Shadow IT – A view from behind the curtain. *Computers & Security*, *45*, 274–283. https://doi.org/10.1016/j.cose.2014.06.007

Silic, M., Barlow, J. B., & Back, A. (2017). A new perspective on neutralization and deterrence: Predicting shadow IT usage. *Information and Management*, *54*(8), 1023–1037. https://doi.org/10.1016/j.im.2017.02.007

Skoglund, M., & Runeson, P. (2009). *Reference-based search strategies in systematic reviews*. 13th International Conference on Evaluation and Assessment in Software Engineering (EASE).

SURF. (2022). *Cyberdreigingsbeeld 2021-2022—Onderwijs en Onderzoek*. 50.

Tuma, K., Calikli, G., & Scandariato, R. (2018). Threat analysis of software systems: A systematic literature review. *Journal of Systems and Software*, *144*, 275–294. https://doi.org/10.1016/j.jss.2018.06.073

Ulven, J. B., & Wangen, G. (2021). A Systematic Review of Cybersecurity Risks in Higher Education. *Future Internet*, *13*(2), Article 2. https://doi.org/10.3390/fi13020039

Verhoeven, N. (2019). *Doing research: The hows and whys of applied research*.

Walterbusch, M., Fietz, A., & Teuteberg, F. (2017). Missing cloud security awareness: Investigating risk exposure in shadow IT. *Journal of Enterprise Information Management*, *30*(4), 644–665. https://doi.org/10.1108/JEIM-07-2015-0066

Walters, R. (2013). Bringing IT out of the shadows. *Network Security*, *2013*(4), 5–11. https://doi.org/10.1016/S1353-4858(13)70049-7

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), Article 1. https://doi.org/10.1038/sdata.2016.18

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in software engineering*. Springer Science & Business Media.

Yan, J., Zhang, S., Milic, N., Koch, H., & Curry, P. (2016). IT Consumerization and New IT Practices: Discriminating, Firefighting and Innovating. *Twenty-Second Americas Conference on Information Systems*, 10.

Zimmermann, S., & Rentrop, C. (2012). Schatten-IT. *HMD Praxis der Wirtschaftsinformatik*, *49*(6), 60–68. https://doi.org/10.1007/BF03340758

Zimmermann, S., & Rentrop, C. (2014). On the emergence of shadow it—A transaction cost-based approach. *ECIS 2014 Proceedings*. 22nd European Conference on Information Systems.

Zimmermann, S., Rentrop, C., & Felden, C. (2014). *Managing Shadow IT Instances – A Method to Control Autonomous IT Solutions in the Business Departments*. 13.

Zimmermann, S., Rentrop, C., & Felden, C. (2016). Governing identified shadow IT by allocating IT task responsibilities. *AMCIS 2016: Surfing the IT Innovation Wave*. 22nd Americas Conference on Information Systems.

Appendix A: Sources Used in the Literature Reviews

List of sources left after selection in the tertiary review (Table 5) and mapping study (Table 6).

TABLE 5: LIST OF SOURCES INCLUDED IN THE TERTIARY REVIEW

(Kretzer & Maedche, 2014)	Generativity of business intelligence platforms: A research agenda guided by lessons from shadow IT
(Kopper & Westner, 2016a)	Deriving a framework for causes, consequences, and governance of shadow IT from literature
(Kopper & Westner, 2016b)	Towards a Taxonomy for Shadow IT
(Klotz et al., 2019)	Causing factors, outcomes, and governance of Shadow IT and business-managed IT: a systematic literature review
(Mallmann et al., 2019)	Shedding Light on Shadow IT: Definition, Related Concepts, and Consequences
(Rakovic, Sakal, et al., 2020)	Shadow IT - A Systematic Literature Review
(Godefroid et al., 2021)	IT Outside of the IT Department: Reviewing Lightweight IT in Times of Shadow IT and IT Consumerization
(Käss et al., 2021)	Towards a Taxonomy of Concepts Describing IT Outside the IT Department

TABLE 6: LIST OF SOURCES INCLUDED IN THE MAPPING STUDY

(Györy et al., 2012)	Exploring the Shadows: IT Governance Approaches to User-Driven Innovation
(Chua et al., 2014)	Central IT or Shadow IT? Factors shaping users' decision to go rogue with IT
(Furstenau et al., 2017)	Shadow It Systems: Discerning the Good and the Evil
(Zimmermann & Rentrop, 2014)	On the emergence of shadow it - A transaction cost-based approach
(Zimmermann et al., 2014)	Managing Shadow IT Instances – A Method to Control Autonomous IT Solutions in the Business Departments
(Silic & Back, 2014)	Shadow IT – A view from behind the curtain
(Haag, 2015)	Appearance of Dark Clouds? - An Empirical Analysis of Users' Shadow Sourcing of Cloud Services
(Zimmermann et al., 2016)	Governing identified shadow IT by allocating IT task responsibilities

(Furstenau et al., 2017)	Shadow Systems, Risk, and Shifting Power Relations in Organizations
(Kopper, 2017)	Perceptions of IT managers on shadow IT
(Silic et al., 2017)	A new perspective on neutralization and deterrence: Predicting shadow IT usage
(Walterbusch et al., 2017)	Missing cloud security awareness: investigating risk exposure in shadow IT
(Magunduni & Chigona, 2018)	Revisiting shadow IT research: What we already know, what we still need to know, and how do we get there?
(Klotz et al., 2019)	Causing factors, outcomes, and governance of Shadow IT and business-managed IT: a systematic literature review
(Rakovic, Duc, et al., 2020)	Shadow IT and ERP: Multiple Case Study in German and Serbian Companies

Appendix B: Data Management Plan

Created using DMPonline (Digital Curation Centre (DCC), 2022) for the project 'Shadow IT in Higher Education and Research'

Data Collection

1.1 Will you re-use existing data? If yes: explain which existing data you will re-use and under which terms of use.

No, I will be collecting/generating new data

1.2 Describe your data. Fill the table below with a brief description of the data, including the type, format and volume.

Data Description	*Data Type*	*Format*	*Total Volume*
Interview recordings	Audio	.wav	< 10 GB
Interview transcripts	Text	.docx	< 100 MB
Organized and analyzed transcripts	ATLAS.ti project	.altproj22	< 200 MB
Informed consent forms	Text	.pdf	< 10 MB

Data Documentation

2.1 Describe the documentation and metadata that you will use to make your data reproducible and interoperable. Describe which files you will provide, along with a brief description of the information they will contain, to make your data reproducible and interoperable. Describe the information that you will provide to make the data items in questions 2.1 reusable and interoperable. If using a specific metadata standard, please mention this below.

For the project as a whole, a README.md file will be written with references to published documentation of methods; the completed master's thesis.

For both the audio files as well as the transcripts, the metadata will consist of only the file metadata (size, time of creation, duration, quality) while the file name (see 2.2) should provide the context necessary for use.

Pseudonym identifiers will be stored with the recordings, to make sure these are not accidentally shared together with the transcripts.

Although not an open standard, the ATLAS.ti project is interoperable with other software from QSR, MaxQDA and Framework.

2.2 Describe the folder structure you will provide to make your data reproducible and interoperable. Describe the folder structure, naming conventions and/or version control you will use for this project.

The folder structure will contain a folder for the recordings, (pseudonymized) transcripts and the project as follows:

```
Project_Folder
  ATLAS.ti_project.altproj22
  Data_gathering
    Pseudonym_identifiers.txt
    Informed_consent_forms.csv
    Audio
      Organisation0.wav
      Organisation1.wav
  Data_analysis
    Organisation_pseudonym0.docx
    Organisation_pseudonym1.docx
```

No version control methods will be used, except for the build-in version control that OneDrive provides in case of accidental removal of files.

Data Storage

3.1 Select the storage solution where you will store and back-up your data. Select the locations where your data will be stored. You may select more than one. Please describe the storage solution and the backup strategy of your storage solution if it does not appear in the list below.

Everything will be stored in OneDrive. We use its built-in backup solution.

Before downloading, the filled-in consent forms are stored in the UU Qualtrics instance.

After the project's completion, data will be archived in Yoda.

Data Privacy and Security

4.1 Will you be collecting or using personal data? Personal data is any data which, alone or in combination with other information, can identify a living person. Such data must abide by the GDPR and requires additional safeguards and documentation to be processed lawfully.

Yes, I will collect and/or use personal data

4.2 What is the legal basis by which you are collecting and/or processing this data? If you are uncertain as to which legal basis applies to your type of research; please do not hesitate to contact us at info.rdm@uu.nl or by using the "Request feedback" button and leaving a comment alongside this question.

Informed consent

4.3 Select the privacy and security measures you will employ to protect the privacy of your data subjects. Check all that apply.

Aggregation/Abstraction, Encryption, Access control, Minimization, Pseudonymization, Secure storage

Access to the recordings is restricted to just the interviewer, who transcribes the recording and lets the interviewee review the transcript before it is further processed. Access to the transcripts is restricted to the researchers involved in analysis. After analysis, the transcripts and results will be pseudonymized. The pseudonymized transcripts will be double-checked by the project PI and one person who is not involved in the research project but works with the researchers in UU or SURF. The anonymization of the transcripts will be done in line with the recommendations provided in Slide 31 in: Managing and Sharing Qualitative Data (2019) https://doi.org/10.6084/m9.figshare.7637288.v1

Data on OneDrive is encrypted in transfer and storage by default. All researchers use devices with disk-level encryption.

4.4 Who is the controller of the personal data? The controller of the personal data is the entity which determines what is done with the data. In most cases the controller is Utrecht University.

Utrecht University is the controller of the collected personal data. Nevertheless, the master student collecting and analyzing the data will ensure that the data is handled and processed in accordance with the GDPR, in consultation with his supervisor.

4.5 How will ownership and intellectual property rights of the data be managed? Describe who controls access to the data and who determines what is done to the data.

The PI supervising this project will determine who has access to the data within the research group. All intellectual property rights belong to Utrecht University. During the project, only the master student and his supervisors have access to the data.

Data Selection, Preservation & Sharing

5.1 Describe the data you will be preserving and the storage solution where it will be preserved? Describe which data will be preserved under long-term storage. You may refer back to the data described in question 1.2 to specify which data will be preserved. Explain where you will preserve your data, and how procedures are applied to ensure the survival of the data for the long term.

All collected data except audio recordings will be preserved. The audio recordings will be transcribed and deleted afterwards.

The pseudo-anonymized data will be kept for at least ten years. After completing the project, the data will be stored in Yoda. Yoda is an infrastructure developed at the Utrecht University and provides an integrated collaboration, secure and (long-term) storage environment. The data will be preserved in a vault where the data are kept safe and cannot be tampered with.

5.2 Describe the data you will be sharing and the repository where it will be shared? Describe which data you will be sharing. Select where you will make your data findable and available to others. If selecting "Other" please specify below which repository and provide a URL. Please also write below if you will apply any conditions to the re-use of your data. (i.e., Creative commons license or Data Transfer Agreement).

Other. The pseudonymized transcripts and analysis in ATLAS.ti will be made publicly available through Yoda repository if the interviewee has separately opted-in for this.

The results will be re-usable under a CC-BY 2.0 creative commons license.

5.3 Are specialized, uncommon or expensive software, tools or facilities required to use the data? Please list any specialized, uncommon or expensive software, tools or facilities that are absolutely required to obtain, use or handle your data, if any.

The transcripts can be accessed by free, open-source or non-proprietary software.

The ATLAS.ti project file can be opened with software from QSR, MaxQDA and Framework, all of which proprietary packages. Some offer a trail or free version.

Data Management Costs and Resources

6.1 What are the foreseeable research data management costs and how do you expect to cover them? Please specify the known and expected costs involved in managing, storing and sharing your data. Also explain how you plan to cover these costs.

The data storage needs will not exceed the amount of storage allocated to each individual by Utrecht University, so no additional costs are expected. The long-term storage will be done in Yoda for 4€ per TB/month, covered by the research group's budget.

6.2 Who will be responsible for data management? Please specify who is responsible for updating the DMP and ensuring it is being followed accordingly.

The project's PI Dr. Kate Labunets will be responsible for maintaining the DMP up to date, granting permissions, and ensuring the data is deposited in the repository.

6.3 State if you contacted an RDM consultant from Utrecht University to help you fill out your DMP. Please list their name and date of contact. This is mandatory for NWO grants.

Jaques Flores from Utrecht University RDM was contacted for information about the processing of company data and informed consent on 12 May 2022.

Appendix C: Informed Consent Form

Information about the research

The interview you are asked to participate in is part of scientific research aiming to gain insights into the presence and cybersecurity problems of shadow IT in Dutch higher education and research institutions. Shadow IT is defined as “hardware, software, or services built, introduced, and/or used for the job without explicit approval or even knowledge of the organization”.[7] The research is conducted in collaboration with SURF. Although they are actively involved in the goals and dissemination of research, they are not involved in the carrying out of the study.

How will the study be carried out?

The interview will take at maximum one hour, during which the researcher will ask questions in a semi-structured format. The interview will be recorded. After the recordings are transcribed, you will get the opportunity to remove any information from the text that should not be included in further analysis. Following the researchers’ analysis of these transcripts, you will be asked to evaluate and add to a summary of the results that are based on the interviews. You will not be reimbursed for your participation in this study.

What will we do with your data?

During this interview, data about your (and your institution's) experiences with shadow IT will be collected. Although the objectives and design of this study do not require specific personally identifiable information, the data collected should be considered as such. The interview will be recorded before it is transcribed. Interview recordings will be retained for up to six months until transcribed. The non-pseudonymized transcripts will only be processed by UU researchers who are collaborating in the study, or who are responsible for assessing its implementation. After analysis, the transcripts will be further pseudonymized as described in the next section. There are no specific increased privacy risks related to the nature of the collected personal data or the processing that the data will undergo. The data is stored and processed exclusively in the EU and all third-party applications used have an appropriate data processing agreement with Utrecht University.

Processed data will be retained for at least 10 years for the purposes of research integrity. Before this archival, all personal information that can reasonably be traced back to you or your organization will have been removed or changed before the files are shared with other

[7] Haag, S., & Eckhardt, A. (2017). Shadow IT. Business & Information Systems Engineering, 59(6), 469–473.

researchers or the results are made public. The researcher will keep a link that identifies you and your organization with the information, but this link will be kept secure and only available to the researcher. Any information that can identify you will remain confidential. The information in this study will only be used in ways that do not reveal who you are. You and your organization will not be named or identified in publications about this study or in documents shared with other researchers.

WHAT ARE YOUR RIGHTS?

Participation is voluntary. We are only allowed to collect your data for our study if you consent to this. If you decide not to participate, you do not have to take any further action. You do not need to sign anything. Nor are you required to explain why you do not want to participate. If you decide to participate, you can always change your mind and stop participating at any time, including during the study. You will even be able to withdraw your consent after you have participated. However, if you choose to do so, we will not be required to undo the processing of your data that has taken place up until that time. The research data we have obtained from you up until the time when you withdraw your consent will be erased.

APPROVAL OF THIS STUDY

This study has been approved by the Science - Geo Ethics Review Board (SG ERB) at Utrecht University under #Bèta S-22770. If you have a complaint about the way this study is carried out, please send an email to the secretary of this Committee: etc-beta-geo@uu.nl. If you have any complaints or questions about the processing of personal data, please send an email to the Data Protection Officer of Utrecht University: privacy@uu.nl). The Data Protection Officer will also be able to assist you in exercising the rights you have under the GDPR. Please also be advised that you have the right to submit a complaint with the Dutch Data Protection Authority (https://www.autoriteitpersoonsgegevens.nl/en).

MORE INFORMATION ABOUT THIS STUDY?

In case you have additional questions, please contact Joost Gadellaa (researcher and data controller for the study) at j.f.gadellaa@uu.nl or Kate Labunets (project supervisor for the study) at k.labunets@uu.nl.

I have read and understood the study information dated ${date://CurrentDate/PT}, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction

Yes / No

I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.

Yes / No

I understand that information I provide will be used for the report and publications in academic venues (like conferences or journals).

Yes / No

I understand that personal information collected about me that can identify me, such as my name or email address, will not be shared beyond the study team.

Yes / No

I additionally agree that my information can be quoted in research outputs

Yes / No

I give additional permission for the pseudonymized interview transcript that I provide to be archived in UU's YourData so it can be used for future research and learning.

Yes / No

Your name:

Appendix D: Interview Protocol

Before recording

Thank the interviewee for willingness to participate, re-iterate the research goals, and set expectations for the duration of the interview (1 hour) and the topics that will be covered.

Have you received the informed consent form?

Do you have any additional questions about it?

Did you fill in the informed consent form?

Start recording

Introduction

What is your current position, and for how long have you been in this position?

What is your total work experience?

Have you had any security-related education, or do you have any certifications?

Can you characterize your institution for me?

How many students, what kind of education and research focus, etc.?

Occurrence of shadow IT

What is Shadow IT for you? How do you understand this concept?

Share this research's definition of shadow IT to delineate the topic.

Explicitly share the goal of identifying instances, not causes or consequences. Explain that the researcher takes notes to get back to consequences later.

What kinds of shadow IT do you see within your institute?

What kinds do you think are used the most?

Is it hardware and/or software?

How essential are these occurrences to the business?

By whom is it used, at what level in the organization?

Do you think this is different from non-HEIs?

Could you estimate how widespread the use of shadow IT is?

Cyber threats of shadow IT

Transition from identifying instances towards eliciting cyber threats.

Introduce the definition of cyber threat used in this research.

Does this shadow IT create cyber threats for your organization?

What security incidents are associated with shadow IT usage?

How could shadow IT lead to… (including an explanation if necessary)

… spoofing?

… tampering?

… repudiation?

… information disclosure?

… denial of service?

… elevation of privilege?

Interview closing

Would you like to add anything else?

Thank the interviewee for their time and explain further procedures of transcript review, member checking of codes and sharing of results.

General probes

Could you give an example?

Could you elaborate/explain?

What would be the shadow IT hardware/software/service here?

Threat scenario modelling-related probes

Who or what would do this? (asking for threat actor)

What enables it to do harm? (asking for vulnerability)

What would then go wrong? (asking for threat scenario)

What are the consequences? (asking for unwanted incident)

What damage is then done? (asking for asset damaged)

APPENDIX E: CODEBOOK

Final codebook used during the analysis of interview transcript. **Gr** is groundedness, the number of times a code occurred in total. **Cv** is the coverage, the number of different experts mentioning the concept. Sub-codes are sorted alphabetically.

Codes	**Gr**	**Cv**	**Description**
SQ3: Occurence	**157**	**11**	Folder for codes answering SQ3: “What types of shadow IT are observed in Dutch higher education?”
Self-acquired devices	**47**	**8**	Use of devices owned by employees. These devices are purchased directly from retail rather than being ordered through the official catalogue of the IT department. (Mallmann et al., 2019)
AV equipment	2	1	Audio and video equipment with a network connection
Control systems and OT	3	1	Hardware and software that monitors and controls operational equipment. E.g., HVAC, security hardware, excluding research or education hardware.
Devices managed by others	2	1	Devices that are not BYOD, not managed by the institution, but managed by another institution
Mobile phones and tablets	4	2	Use of mobile devices for work that is not supported/intended by central IT.
Networking devices	2	2	Unofficial WiFi access points, routers, switches etc. brought into the network by end-users
Research equipment	10	4	Hardware with embedded software used for research in laboratories etc. E.g.: fridges, microscopes, scanners.
Server hardware	9	4	Hardware bought to run server-type applications. E.g. websites, shared storage, ML, or automation scripts. Also includes NASes.
Storage devices	1	1	Usage of physical storage devices not approved by the institution. E.g. USB-harddrives and memory sticks.
Unmanaged PCs	14	6	Normal, but unmanaged, PC hardware brought into the institution by an end-user
Self-acquired software	**47**	**10**	Use of software installed by employees to perform their work tasks, excluding web-apps and SaaS services.
Locally installed apps	18	7	Apps, built by others, that are installed as an executable on a user’s machine and run locally
Unspecified software	29	9	Code used when the type of software is not specified.
Self-made solutions	**32**	**8**	Use of solutions developed by employees to perform their work tasks. (Mallmann et al., 2019)
Ad-hoc coupling of systems	3	3	When processes, for which usage of an official IT system or service bus would be preferred is instead handled by transferring data via email, scripting, scharing links etc.
Self-built websites	3	2	Self-built or comissioned, websites built without knowledge from the central IT department
Self-developed software	11	3	Software developed by staff or students for personal use or as part of research/education.
Spreadsheets and databases	15	6	Shadow-storage and (limited) processing of data. Often exports from official administrative systems.
Unapproved cloud services	**31**	**9**	Use of Internet-based Software and Software as a Service (SaaS) that are not approved or unknown by the IT department. These systems are also called Mobile Shadow IT once they can be accessed outside the workplace. (Mallmann et al., 2019)
Cloud productivity suite	11	6	Systems used by multiple users for ‘productivity’ tasks like email, contacts management, note taking, file storage and editing.
Cloud storage	9	6	Systems used by multiple user, primarily for file sharing and storing. When a system also provides an editing interface on top it is cosidered a productivity suite

Web-apps	12	5	Software, built by others, that runs in the browser.
SQ4: Cyber threat	**291**	**10**	Foder for codes answering SQ4: “Which cyber treats related to shadow IT are perceived by experts?”
Assets	**27**	**9**	An asset is something to which a party assigns value and hence for which the party requires protection.
Availability of data and systems	0	0	*Implicit code only used in analysis. A system working as intended and being reachable by users
Confidentiality of data	0	0	*Implicit code only used in analysis. Access to data being restricted to only those who should have access.
Continuity of primary processes	13	5	The education and research activities of an institution being able to continue unhindered
Employee time	4	3	Employees have to spend (more) time on something, decreasing the time they can spend on their primary tasks
Financial	4	4	Direct financial assets
Integrity of data	0	0	*Implicit code only used in analysis. The accuracy and consistency of data.
IP oppertunities	2	2	(Exclusivity of) valuable knowledge.
Reputation	4	2	What the institution is associated with by outsiders.
Threat scenarios	**75**	**10**	A threat scenario is a chain or series of events that is initiated by a threat and that may lead to an unwanted incident.
Access to employee account	14	6	Somebody can access an employee acount and perform actions on their behalf without permission. Could be seen as low-level privilege escalation.
CEO-fraud	2	1	Using social engineering and (possibly) access to an employee account, to lure a person with financial responsibilities into transferring money
Data accessed or exfiltrated	10	6	Accessing institution data and/or downloading it to outside of the organisation
Device is lost or stolen	5	2	A device is lost or stolen
Lateral movement	7	5	Taking steps to explore the network and possibly access other devices
Instal malware	14	8	Malware is installed as part of (or instead of) another application or introduced via a storage device, dropper or in-person
Perform phishing	6	3	Phishing or spearphising in order to extract login details from a target
Privilige escalation	2	2	Increasing user account privileges to gain administrative rights. The fist step of privilege escalation is coded as ‘Access to employee account’
Remote access	15	6	Being able to gain initial or persistent access trough a device connected to the network
Sabotage	2	1	Functionality of a system is intentionally reduced
Threats	**6**	**4**	Who/what may initiate threat scenarios and unwanted incidents?
APTs	3	2	Advanced persistent threats. Highly capable actors, often state-associated. Motivatied by espionage, militairy or political gains.
Cybercriminals	0	0	*Implicit code only used in analysis. Generic code for unspecified malicious actors looking for financial gain.
Hacktivists	1	1	Groups who carry out cyber-attacks in support of political causes (Fowler, 2016)
Students	2	2	Students from the institution themself.
Unwanted incidents	**67**	**10**	An unwanted incident is an event that harms or reduces the value of an asset.
Abuse of commputer resources	5	3	An institutions computing or storage resources are used for something else than it’s intended purposes
Commercial use of data	7	4	Student of staff data is used for commercial purposes, most notably advertising
Data breach	13	6	Data is published publically, if data is used for a different pupose, another code applies
Discontinuation of services	2	2	A (shadow) service is shut down by the (shadow) supplier

Leaked credentials	9	6	Specification of data breach: credentials are leaked causing immediate damage to account security
Ransomware and extortion	6	4	Data is encrypted and/or threatened to be leaked by an actor for financial gain, with the aim of getting a payment.
Student fraud	1	1	A student performs illegal actions in order to increase their academic perfomance
Unauthorized access	16	6	Somebody who should not have access to data or a system gains access to it
Unavailability of system or data	5	2	A system or data is not avaiable, potentially disrupting 'business' processes
Unexpected costs	3	2	Because of a shadow-IT related problem, the institution has to pay for a service that they did not plan for
Vulnerabilities	**139**	**10**	A vulnerability is a weakness, flaw or deficiency that opens for, or may be exploited by, a threat to cause harm to or reduce the value of an asset.
Firewall or antivirus lacking	3	2	Basic security measures are missing or disabled, allowing for exploitation of vulnerabilities in a system
Human error	3	2	A human error made it possible to perform malicious activity
Lack of access control	16	6	A system does not have (sufficient) access control, so unauthenticated users can use it
Lack of authorization policy	12	6	Authorisation control is not thought out, so that users might have access to parts of the system they shouldn't (even though a system has some authentication in place)
Lack of backups	3	2	Data is not backed up (properly), so it would not be protected in case of storage failure/ransomware/etc.
Lack of control of data	22	7	Data is stored outside of the control of the organisation (could be seen as a super-vulnerability, but is often not further specified)
Lack of data encryption	4	4	Data is stored unencrypted, making accessing it by unauthorized actors trivial
Lack of logging	3	3	(Security) events are not logged in the system, decreasing the institution's possibility to respond to problems/threats
No contract with the supplier	9	4	There are no formal agreements made with the (shadow) supplier. This is probably an overarching vulnerability.
No visibility of vulnerabilities	7	4	If a device is out of sight, so that vulnerabilities are not known or cannot be fixed.
Not actively managed	7	5	There is nobody responsible for a system's security. Probably a super-vulnerability
Not secure by design	6	4	Software is developed without following given principles that ensure an absence of vulnerabilities, probably a super-vulnerability
Outdated software	25	9	Software does not have the latest (security) updates Or security updates are no longer being supplied (legacy software)
Re-used password	7	6	The password used for a service is the same as for other services, e.g., the main organisation account
Remotely accessible	6	3	A device or endpoint that can remotely be accessed, increasing the attack surface of the organisation
Unsuitable hardware	2	2	System design issues, but on a lower level. E.g., untrustworthy hardware causing a risk of outages
Users have install rights	5	4	Users can install software on their managed device
SQ0: Codes for discussion	**278**	**11**	Folder for exploratory inductive codes regarding causes and culprits, governance, frequency, and treatments of shadow IT (threats)
Causal factor	**58**	**11**	A factor that causes Shadow IT, specific to HEI's according to the participant
Anticipated costs	4	2	Shadow IT is used because it is (thought to be) cheaper than the official alternative
Cooperation across organisations	8	5	Usage of Shadow IT is needed to work together (more easily) with staff or students from external orgainsations

Culture of openness	3	3	Security is difficult, because institutions are (sometimes literally) open to outsiders and securing some aspects more is undesirable
Culture of personal freedom and autonomy	12	9	Staff or students are used to make their own decisions regarding their way of working, and they continue this tendency in choosing their IT setup, therefore not always using the official systems.
Drive to innovate	7	5	Staff or students want/need to try new tools, that are not yet procured/approved by the IT department
Federation approach	9	5	Responsibilities are left to lower level organisational units (faculties, institutes), leading to IT being initiated there as well
Findability of official solutions	4	3	Staff/students don't know where to find the official IT solutions, so they choose something themselves
Lack of consequences	1	1	There are no consequences to the usage of Shadow IT
Missing service	7	4	Shadow IT fulfils a need that is not fulfilled (completely) by an existing IT service
Use mandated by teacher or lecturer	3	2	Students or other lecturers use a certain system because somebody higher up mandates it.
Governance approaches	**24**	**10**	A way of dealing with Shadow IT, specific to HEI's according to the participant
'Free field' approach	5	2	Approach where the application landscape is devided in 'castle', 'city' and 'open field', with decreasing mandates from central IT
Control via budget	7	4	Controls are put in place when money is spend on IT to ensure compliance with the institutes standards
Pre-finance official systems	2	2	Official systems are being financed centrally, so that their use has no financial consequences for departments (unlike buying other software)
Three lines model	5	4	Seperating IT (security) governance into three. The first line is in the 'business' unit and responsible. The second line is advice available to the first line and helps to assess and monitor risks. The third line is independent and reports risks to the board.
Zero trust SaaS	5	4	Manage risks by primarily using SaaS applications with access management. Not relying on network peremiter secuirty. Includes many of the suggested treatment scenarios implicitly.
Indication of frequency or severity	**12**	**5**	When a participant expresses their suppositions about the extent of shadow IT in their institution
Privacy and Process	**33**	**10**	Incomplete, but some codes were moved to this category when they turned out to be out of scope after the coding. Many more were already disregarded as out-of-scope during the coding.
GDPR non-compliance	16	6	The event causes the institution to be non-conformant to the GDPR or other regulations
Incompatibility with other devices	3	3	Devices or OS'es of devices being incompatible with each other, causing problems
Shadow system contains wrong information	4	3	An unofficial system, possibly out of sync with the official system, is used as a source of truth for 'business' decisions
Unclear retention policy	5	5	It is not clear when or if data will be removed after storing it
Wrong data being used	7	3	Because of shadow IT or malicious intent, the wrong data is being used for business decisions 25/10/2022, 13:17,
Treatment scenario	**133**	**11**	A treatment scenario is the implementation, operationalisation or execution of appropriate measures to reduce risk level.
Add to SSO solution	2	2	Connecting the application/system to a shared authentication/authorisation service
Air-gapping	4	2	Disconnecting the device from the network. This code is for physical disconnecting, or completely removing access with policy. Less extreme forms are gathred under 'network segmentation'
Awareness of official solutions	10	4	Making staff or students more aware of official IT solutions, to decrease their tendency to use shadow IT

Awareness of risk and responsibilities	14	7	Increasing employee/student knowledge of what can go wrong with shadow IT, to decrease their tendency to do so.
Certificates for network access	3	2	Using certificates to authenticate a device on the netwerk, instead of, or additional to, username/password authentication
Disable mobile apps	2	1	Disable the possibility of institution software to be used on (private) mobile devices
Don't allow institution login for third parties	1	1	‘Login with Microsoft’ or ‘Login with Google’ is disabled for the institution, so users cannot share their data with third party apps
Embed security in existing regulations	11	4	Security checks are put in place with existing (privacy, ethics, financial) regulations and processes in the case that they are applied well
Endpoint monitoring, scanning and MDM	4	3	Devices are scanned for active software, vulnerabilities, logging, etc. On phones and tablet OSes this is called Mobile Device Management
Feedback conversations	8	5	Organising conversations with the first line responsible manager or local IT administrators (not end-users)
Have agreements with un-supported software suppliers	5	4	Even if a supplier is not officially procured by the institutions, agreements are made to ensure security standards for shadow use
Have antivirus	5	4	Installing antivirus on local devices to prevent malicious code from executing
Internal network scanning and monitoring	6	5	Scanning the internal network and servers for known vulnerabilities, and notifying the responsible administrators
Managed devices	14	8	Issuing managed devices to employees (and possibly students) to have more control over software and monitoring installed/allowed.
MFA	9	6	Adding a second or more factors to the authentication process to increase resilience to attacks. Also includes PAM and Passwordless authentication
Network perimeter scanning	3	3	Scanning the perimeter of the institution network for, for instance, open ports and unknown API endpoints.
Network segmentation	22	9	Seperating parts of the network in order to decrease the scope of lateral movement
Patch Management	3	2	Installing software updates on (shadow) hardware
Take over management	2	1	Taking over the management of self-aquired devices
Unique password requirements	4	3	Having password requirements that are stricter or different than other services, in order to discourage re-use of passwords
Zero footprint design	2	1	Software is designed in such a way that it does not leave traces/data on the host device it runs on
Who	**24**	9	Category used to code the type of actor that initiated the use of shadow IT
Decentral IT department	2	2	Another level of IT management in the organisation, often an institute or faculty who partly manages their own IT
Educators	12	5	Employees in their role as teaching staff
Researchers	9	4	Employees in their role as researchers
Students	1	1	Students of the institution

APPENDIX F: ANONYMIZATION GUIDELINES

In general, we take an extensive approach to person and institution anonymization, removing everything that can possibly be traced back to a person or institution to the best of our ability. We try to preserve grammatical correctness for readability. Pseudonymization only used within a paragraph, where it can be used to indicate the interviewee is still talking about the same institution, service etc. Apart from that, we anonymize al sensitive information to minimize the risk of re-identification.

Description	Original (example)	Transcription
The institution is always completely anonymized.	Utrecht University	[institution]
All cities are anonymized.	Utrecht	[city]
All names are anonymized.	Joost Gadellaa	[name]
For background information, we take a conservative approach, completely anonymizing paragraphs.	I studied Business Informatics from 2020 to 2022	[personal background information]
We do the same for institutions, to minimize risk of re-identification.	UU has over 35000 students in 7 faculties	[institution background information]
For this, it is often needed to obfuscate years and dates	Four years ago, we restructured the IT responsibilities	[Several] years ago, we restructured the IT responsibilities
Important to this regard is the number and local naming convention of departments since this is unique for institutions.	This user studied at two of the seven existing faculties during their studies.	This user studied at two of the [departments].
We remove URL's and email addresses, preserving details such as "the official domain was used" when relevant.	They just used their initials.surname@uu.nl to log into uu.shadowit.nl	They just used [institution email format] to log into [institution].[URL]
Because used systems can be sensitive, both in terms of current security profile as well as future procurement processes, software and hardware is reduced to its type and whether it is official.	UU uses Office365, but I prefer Nextcloud.	[Institution] uses [official cloud productivity suite], but I prefer [shadow cloud productivity suite].
If necessary for the context or to preserve the function of software in the sentence, the type of software or service can be specified beyond the categories used within the thesis.	They couldn't get Thunderbird as an application on their laptop, so they get it directly from Mercurial.	They couldn't get [an email client] as an application on their laptop, so they get it directly from [a source code management tool].
Within a few paragraphs, a local reference can be used to clarify and shorten the anonymization. We do not use global pseudonyms.	They used Nextcloud as their self-hosted personal cloud for work purposes, so they wanted our storage solution to integrate with Nextcoud as well.	They used [a shadow cloud productivity suite] as their self-hosted personal cloud for work purposes, so they wanted our storage solution to integrate with [that suite] as well.
References to earlier in the conversation might need more context. If the paragraph does not make the context clear, a reference to a line can be used.	That Nextcloud I talked about earlier worked much better than the cloud solution we procured ourselves.	That [shadow cloud productivity suite from line 24] I talked about earlier worked much better than the cloud solution we procured ourselves.