predict
prioritise
prevent

TRESPASS

Technology-supported Risk Estimation by Predictive Assessment of Socio-technical Security

Deliverable D2.3.2

TRESPASS Social data and policy extraction techniques

Project:	TRESPASS
Project Number:	ICT-318003
Deliverable:	D2.3.2
Title:	TRESPASS Social data and policy extraction techniques
Version:	1.0
Confidentiality:	Public
Editor:	L. Coles-Kemp, C.P. Heath
Cont. Authors:	L. Coles-Kemp, J.-W. Bullée, L. Montoya, M. Junger, C.P. Heath, W. Pieters, L. Wolos
Date:	2015-10-30

Part of the Seventh Framework Programme
Funded by the EC-DG CONNECT

Members of the TRESPASS Consortium

1.	University of Twente	UT	The Netherlands
2.	Technical University of Denmark	DTU	Denmark
3.	Cybernetica	CYB	Estonia
4.	GMV Portugal	GMVP	Portugal
5.	GMV Spain	GMVS	Spain
6.	Royal Holloway University of London	RHUL	United Kingdom
7.	itrust consulting	ITR	Luxembourg
8.	Goethe University Frankfurt	GUF	Germany
9.	IBM Research	IBM	Switzerland
10.	Delft University of Technology	TUD	The Netherlands
11.	Hamburg University of Technology	TUHH	Germany
12.	University of Luxembourg	UL	Luxembourg
13.	Aalborg University	AAU	Denmark
14.	Consult Hyperion	CHYP	United Kingdom
15.	BizzDesign	BD	The Netherlands
16.	Deloitte	DELO	The Netherlands
17.	Lust	LUST	The Netherlands

Document History

Authors		
Partner	Name	Chapters
RHUL	Claude Heath, Lizzie Coles-Kemp,	1,9,10,A
GUF	Lars Wolos	5
UT	Jan-Willem Bullée	4, 6
UT	Lorena Montoya	2
UT	Marianne Junger	8
TUD	Wolter Pieters	7

Quality assurance		
Role	Name	Date
Editor	Lizzie Coles-Kemp, Claude Heath	2015-09-30
Reviewer	Axel Tanner	2015-10-23
Reviewer	Elmer Lastdrager	2015-10-16
Task leader	Lizzie Coles-Kemp	2015-10-30
WP leader	Michael Osborne	2015-10-30
Coordinator	Pieter Hartel	2015-10-30

Circulation	
Recipient	Date of submission
Project Partners	2015-10-30
European Commission	2015-10-30

Acknowledgement: *The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 318003 (TRESPASS). This publication reflects only the authors' views and the Union is not liable for any use that may be made of the information contained herein.*

Contents

List of Figures

List of Tables

Management Summary

This deliverable presents the short-listed collection of methods and data sets to be used by the TRE$_S$PASS model when building the database of social assets, producing the social dimensions of the Attack Navigator Map and undertaking the risk calculations that use social data.

Key takeaways:

- Identification of what social data is and how it contributes to the TRE$_S$PASS model.
- The data gathering tools and approaches are ordered in this deliverable by breadth, starting with the tools and approaches gathering the broadest range of social data and ending with the tools and approaches that gather the most granular types of social data.
- The data gathered through the tools and approaches outlined in this deliverable will be used to populate the attack pattern library and to contribute to the development of the narrative on which instances of the Attack Navigator Map are based.

The tools and approaches start with aggregation techniques used in Geographical Information Systems (GiS) to provide a high-level overview of risk hot spots. The aggregation techniques that gather data through questionnaires and present the analysis using visual mapping techniques enable analysts to ground user behaviours and practices related to information sharing and protection of particular spaces. Traditionally patterns of practice are linked to physical spaces as the visualisations illustrate but in TRE$_S$PASS could also be developed to link to digital and organisational spaces.

The next set of tools and approaches to be presented are those that form part of Stage Zero risk assessments. This involves participatory modelling techniques, designed to enable the different stakeholders to co-produce a model of the scenario. It also allows them to depict the different information-sharing and information-protection practices in operation within a particular scenario. Such modelling tools enable stakeholders to identify the goals and values of each community of practice, the potential conflicts between different information sharing and protection practices and the motivations behind information exchanges taking place within a scenario. The Stage Zero risk assessment approach can be usefully combined with the aggregation techniques to produce a more comprehensive map of information sharing and protection practices.

Social engineering success stories is an innovative narrative technique that contributes attack technique data to the attack pattern libraries. Such a technique analyses attacker stories and produces patterns of attack steps and attack motivation weighted according to

the frequency in the narrative corpus. This output can be used to overlay the information sharing and protection maps to better identify where there are gaps that might be exploited by attackers.

Information sharing and protection maps can be enhanced in different scenarios through organisational records. The Call-Detail record technique illustrates how this can be done in the telecommunications scenario.

As the introduction to this deliverable reflects, control strength is an important element of social data analysis. The practitioner survey of socio-technical cyber threats was used to illustrate how such surveys can be coded and analysed in varying ways to reveal different types of knowledge related to control strength and the threats to controls. These different coding and analysis approaches could be incorporated in the TRE$_S$PASS Attack Navigator Map. The Cues and Warnings and Security by Experiment techniques are particular approaches to testing the strength of controls and are valuable methods that can be used by security practitioners to refine a TRE$_S$PASS model.

1. Introduction

1.1. Goals

The focus of this deliverable is to address the topic of social data gathering that is needed for the TRE$_S$PASS model. This describes the varied data types covered by the term 'social data', and the chosen methods for processing this data. Moreover, it will be shown how this data and their associated research methods are to be encompassed by the TRE$_S$PASS analytical model, as inputs, and how they will contribute to the TRE$_S$PASS visualisation platform that accompanies the model. Furthermore, we also describe how TRE$_S$PASS modelling and visualisation outputs can be considered as an expression of theoretical interest in positive and negative security.

1.2. Motivation and challenges

D2.3.1 identified the different types of social data that are necessary in order to quantify information security successfully. This deliverable now focuses on a subset of tools and social data types shortlisted as likely candidates for the final version of the TRE$_S$PASS model.

The different candidates all reveal something unique about the context of the information security risk. Information security risk is situated within a specific context and the social factors that influence and, at times, give rise to a particular information security risk are therefore a part of that context. The work presented in this deliverable contributes to the TRE$_S$PASS model in two ways:

- Present techniques and methods that can be used by security practitioners to gather and make sense of social data. In particular these tools enable the security practitioner to develop a picture of the social context and to identify the practice and cultural groupings within that context.
- Present data, data types and data patterns that can be input into the TRE$_S$PASS model and used directly for calculation and visualisation.

1.3. Document structure

The remainder of this document begins by describing our frames of reference, giving definitions of key terms (Sect. 1.5). This is followed by detailed summaries and conclusions upon each of the data gathering methods and techniques which are ordered in terms of their breadth and granularity. Our conclusions upon the various approaches are finally tied into the wider aims of TREsPASS (Chapter 9).

Appendix A provides the context for this deliverable in the TRE$_S$PASS project. It describes the overall summary of the project and the TRE$_S$PASS workflow.

1.4. Foreground and background

For Chapter 3, all data analysis techniques are foreground, and *LEGO* techniques are background.

For Chapter 7, all included material is TRE$_S$PASS foreground, except the following:

- The contributions of Dechesne to the security-by-experiment work (33% of the corresponding papers; the other 67% is TRE$_S$PASS foreground);
- The contributions of external organisers to the Dagstuhl seminar and the corresponding report (60%; the other 40% is TRE$_S$PASS foreground).

1.5. Concepts

In order to correctly situate the methods that are presented in this deliverable, it is first important to explain a number of core concepts.

1.5.1. Definition of 'Social data'

Social data is understood primarily as data that feeds into our understanding of the human and organisational relationships in a given scenario, and is data that can be said to contribute to a model of the social relationships that are entangled within the scenario. Such a model explains or describes how the relational dimension interacts with the supporting technical infrastructures that enable information-sharing practices. These practices comprise an important part of social data, and the literature reflects this by emphasising the many different aspects of these practices as topics of study in their own right.

1.5.2. Definition of 'Control' and 'Control strengths'

A working definition of 'control strength' is discussed here, where this is understood primarily as the ability of a security control to withstand malicious attack. We have taken an industry standard approach to this core concept, accepting the definition supplied by *The Open Group* and the accompanying criteria they provide for measuring and rating control strengths (Tab. 1.1).[1]

> **Control Strength (CS) is the strength of a control as compared to a baseline measure of force.** A rope's tensile strength rating provides an indication of how much force it is capable of resisting. The baseline measure (CS) for this rating is pounds per square inch (PSI), which is determined by the rope's design and construction. This CS rating doesn't change when the rope is put to use. Regardless of whether you have a 10-pound weight on the end of the 500-PSI rope, or a 2000-pound weight, the CS doesn't change.
> Unfortunately, the information risk realm does not have a baseline scale for force that is as well defined as PSI. Consider, however, password strength as a simple example of how we can approach this. We can estimate that a password eight characters long, comprised of a mixture of upper and lowercase letters, numbers, and special characters, will resist the cracking attempts of some percentage of the general threat agent population. The password Control Strength (CS) can be represented as this percentage. (Recall that CS is relative to a particular type of force - in this case cracking). Vulnerability is determined by comparing CS against the capability of the specific threat community under analysis. For example, password CS may be estimated at 80 percent, yet the threat community within a scenario might be estimated to have better than average capabilities - let's say in the 90 percent range. The difference represents Vulnerability.
> **'Risk Taxonomy'**, *The Open Group*, Section 5.2.8, p.13.

Table 1.1.: *The Open Group* rating of control strengths.

Rating	Description
Very High (VH)	Protects against all but the top 2 percent of an avg. threat population
High (H)	Protects against all but the top 16 percent of an avg. threat population
Moderate (M)	Protects against the average threat agent
Low (L)	Only protects against the bottom 16 percent of an avg. threat population
Very Low (VL)	Only protects against the bottom 2 percent of an avg. threat population

According to the *The Open Group* schema, there is a four-stage lifecycle that controls follow: 1). the design of controls, and 2). their implementation, followed by 3). their use and maintenance, and finally 4). the disposal of controls no longer needed. Controls are also characterised with respect to three further dimensions, enabling the assessment of

[1] Accessed 19 October 2015: http://fairwiki.riskmanagementinsight.com/?page_id=37

control *capabilities*, or affordances, analytical categories intended to eliminate significant gaps that may occur in an organisation's risk management processes. These are:

1. **Forms**: policy, process, or technology (or a combination of these).
2. **Purpose**: preventive, detective, or responsive.
3. **Taxonomy**: an explicit description of control types, designed to ensure that gaps don't exist in the 'controls environment'.

To complete the array of analytical tools, the Open Group suggests that there are three primary control categories with which to calculate risk assessments based on this understanding of controls: **Loss event** controls, **Threat event** controls, and **Vulnerability** controls.

One Open Group author suggests that qualitative and quantitative approaches may be intermixed when applying the taxonomy to a given situation, and that 'the pertinent question isn't whether error or inconsistency exists, but whether the degree of error or inconsistency is acceptable.'[2] This mixed approach is a valued component of TREsPASS data gathering and modelling processes. In theory, then, several analysts evaluating the same scenario, should obtain reasonably consistent results.

1.6. How to approach and find social data

When gathering social data it is key to understand and use that data in the context in which it was gathered. It is also important to identify the different social practices at work as these form the communities in which information is produced, circulated, curated and protected. In this deliverable we present two techniques for linking social, physical and logical contexts.

The social practices are also used to agree on the goals and values of the community and these have a direct influence on the manner in which information is generated and managed. Social practices can be used to identify the controls and the control strength but also the modus operandi of the attackers and this deliverable presents techniques to do both.

When using analysis about social practices it is important to understand the concept of security in its broader context. Social practices both provide the freedom to do something (positive security) as well as the protection from harm (negative security) and a community's overall security is a combination of both.

[2] fairwiki.riskmanagementinsight.com

1.6.1. The importance of 'Context'

This brings us to the consideration of what constitutes a 'given situation', and how an appreciation of the importance of context may be incorporated into the tools that TREsPASS is designing. Here, context is understood primarily as the internal or external situation in which the information security risk is present.

Some social science theorists have pointed to the sheer complexity of social practices within and around organisations, suggesting that this presents a 'wicked problem' to any form of rigorous analysis (Rittel & Webber, 1973), which could be seen as a barrier to comprehensive understanding of social practices in their many forms, but which nevertheless presents an opportunity to create hugely 'rich pictures' of work context (Monk & Howard, 1998). Others have focused on this variety as a potentially rich source of highly contextualised data, capable of bringing the practices that surround information sharing practices to life for the analyst, giving the practices their rationale, or what we might call the internally consistent logic that allows them to be shared by specific communities of use (Dourish, 2004), a logic which may not be apparent from a viewpoint outside the practices. The importance of the participatory research methods is due to their ability to extract the highly specific narratives associated with these practices (Pentland & Feldman, 2007), some of which are described below (Chap. 3).

1.6.2. Social Practices

The intensive data-sharing in social practices, brings us to the question of how these situated accounts of practices can be scaled up, or given some degree of abstraction that will lend itself to being processed by TREsPASS tools that seek to analyse and represent social data and information security risk.

Social practices are understood primarily to be those iteratively developed patterns of human behaviour that are (in this case) associated with the sharing of data across a given organisational infrastructure, or 'place', for example (Harrison & Dourish, 1996). This also relates to how these internal patterns interact with other external patterns of practices. Social practices have been defined as recursive and cumulative temporal and spatial patterns (Giddens, 1984), or even as 'manifolds' of social practice (Schatzki, 1996). General patterns, at higher levels of societal analysis, have previously only been schematically visualised, creating pictorial metaphors for contrasting types of interlocking shapes and mechanisms that have been found in social practices (Shove, 2003).

1.6.3. Positive and negative security

Positive and negative security are concepts that have been widely accepted, even within the realm of international relations, where this is seen as 'the distinction between *freedom from* (negative) and *freedom to* (positive)' (Roe, 2008, p.778, our emphasis). Some writers have stated that rebalancing the kind of language used to describe the security landscape

requires that the 'security referent' is transferred from the state to the individual and in the process 'embodies a positive image of security' that is no longer 'focused upon the negative *'absence of threat'* approach (Hoogensen & Rottem, 2004, p.4, our emphasis). Some have argued that there is an ethical dimensions to this, that positive security 'defines liberation from oppression as a good that should be secured' (Huysmans, 2002, p.59).

The notions of positive and negative security can be regarded as complementary concepts, adding greater depth to one another, and furthermore, they can be developed as tools with which to elicit another closely related concept integral to the central aims of TRESPASS, that of resilience provided by context and human relationships. Security as resilience is a particularly strong theme in the work of security theorist Bill McSweeney (McSweeney, 1999) who outlines an argument for recognition of a form of relational security that supports the sense of everyday security where an individual feels safe and secure when going about their everyday activities (Roe, 2008). McSweeney says that positive security is necessarily tied to the 'more human' and identifiable 'property of a relationship'. Relational security depends upon such relationships, where they are part-and-parcel of effecting a service, for example. In such cases, the positive sense of security derived from trusted relationships (relied upon by individuals in order to carry out their day-to-day tasks and activities) both at work and at home. Positive security can therefore be a useful concept for the task of transforming what might be seen as a landscape of threat (metaphorically speaking), into one that represents the aspects of the landscape which lend themselves to every-day security.

Security has most often been referenced in the nominative, rather than the adjectival, says McSweeney. He suggests the importance of this distinction lies in the capacity to transform perceptions about the objects of security (their referents):

> There is a certain security, or confidence, in the fact that they are objects, tangible, visible, capable of being weighed, measured or counted. They protect things, and prevent things from happening. When we speak of 'secure', on the one hand, it suggests enabling, making something possible. (McSweeney, 1999, p.14)

This should be seen in contrast to the traditional focus on negative security, he says, quoting Arnold Wolfers: 'security after all is nothing but the absence of the evil of insecurity, a negative value so to speak' (p.14). McSweeney, on the other hand, argues that positive security creates a freedom to take part in the day-to-day events that are vital for the well-being of the individual (enables them), as well as the community and the wider society. Without relational security of this kind, a form of paralysis is experienced resulting from anxiety in the relationships that are fundamental to day-to-day experiences and practices. This aspect of security is highly relevant to cyber security because the mission of cyber security is, in part, about enabling the individual, the community and wider society (Von Solms & Van Niekerk, 2013) to conduct their everyday lives in environments that have been (and continue to be) transformed by a dazzling variety of digital media.

Positive and negative security can be further understood through the related concept of 'value-orders' (G. M. Smith, 2005) that operate within specific communities of practice (Wenger, 1998). In addition we have seen how the goal alignments that are found within

organisations can be traced in their action upon a service design for example (Heath, Coles-Kemp, & Hall, 2014). It is the aim of this deliverable to demonstrate methods that will elicit the values that constitute the basis of information-sharing social practices.

1.6.4. Personas and security

One possible way in which the data gathered by the techniques presented in this deliverable can be used is to develop personas that provide security practitioners with insights into the modus operandi and motivations of the different stakeholder groups involved in a risk scenario. Personas offer a means of illustrating the different stakeholders and perspective in a risk scenario.[3] They are particularly powerful in the context of information security risk which is greatly influenced by the risk perception of the individual. However the power of the persona is limited unless they can be situated within rich contexts and can be designed in such a way that they can be brought to life so that researchers and professional roles (viewers) can explore together how situations appear and feel from different perspectives.

1.6.5. Summary

The techniques presented in this deliverable work together to both develop the attack pattern libraries and to produce a narrative that situates the abstracted analysis presented in the Attack Navigator Map. Their focus is to provide the context, the data on social practices and the broad picture of what constitutes security for an attacker or a defender.

In year 4 there are many potential uses for the techniques and methods presented in this deliverable. Initially some of these techniques will be used to generate input to the attack pattern libraries. They will also be used to provide a richer version of the current Attack Navigator Map. They will also be presented as tools in their own right to be used by practitioners to refine and extend the TRE$_S$PASS model. The Persona is one possible artefact that might be created to link the attack pattern libraries with the Attack Navigator Map.

Potential uses for the presented techniques and methods are given at the end of each chapter. However, in summary the contributions to the TRE$_S$PASS model can be characterised as follows:

- The techniques can be used individually or in combination to help modellers better understand socio-technical contexts.
- The contextual understanding gained from using these techniques help modellers to decide where to focus the analysis and also what aspects of a scenario to model.

[3] For example, see the 'Threat Assessment and Remediation Analysis' (TARA) methodology, a subset of 'Mission Assurance Engineering' Mitre Technical Report, which is being adapted to current work within WP2.

- The qualitative outputs obtained from these techniques can be used to focus surveillance and monitoring activities.
- The quantitative outputs obtained from these techniques can be used as input to the likelihood calculations performed by the model.

This summary list is broken down into more detail in the following table, where the TRE$_S$-PASS social data gathering tools and techniques describeds in this deliverable are compared (Tab. 1.2). This is with particular respect to the different dimensions of case studies and scenarios that are addressed by each, physical, digital and social/organisational. The tabs also shows how these tools can be integrated within the Attack Navigator Map (ANM), and thus mutually support one another across the TRE$_S$PASS visualisation and analysis platform.

Table 1.2.: *Table comparing TRESPASS social data gathering tools and techniques*

Ch.	Techniques + tools	Physical spaces	Digital	Social/Organi-sational	integration with ANM
2	GIS-ATM	Geolocation data	Aggregation of data contributing to understanding of context	–	Mapping of risk hot-spots
3	Stage-Zero	Rich picture contributing to an understanding of context and to the quantification of context	Rich picture	Rich picture	Mapping info. sharing
4	S.E. Success Stories	Contribution to a focus of analysis	–	Attacker modus operandi	Narratives to APLibrary
5	Call Details/Customer Relations	Cobtribution to a focus of analysis	–	Customer Records	Mapping info. protection
6	Socio-Technical Cyber Threats	Contribution to an understanding of context	–	New Threats/attack goals	Overlay patterns on ANM
7	Experiment/quantitative pen.testing	Physical trespass contributing to a quantification of context	Remote hacking	Social engineering	Model refinement
8	Cues and Warnings	Contribution to a quantification and qualifaction of context	–	Threat prevention	Model refinement

2. ATM and GIS

The ATM case study is a derivative of the IPTV case study in the sense that on the latter case study, ATM machines located throughout a city were identified as some of the locations where the financial abuse could take place. However, the ATM case study focuses on the risk of theft of the safe and malware, hence excluding other forms of crime such as robbery of ATM cards.

ATM risk modelling involves integrating technical data (logical infrastructure), physical data (physical location of the infrastructure) and social data (human factors). Until now, 'location' in the TREsPASS model has been used in the general sense. However, the ATM case requires extending the existing concept of 'location' to one that is also able to handle x, y (z) coordinates. An important issue that the ATM case highlights is that many aspects of the context k vary across a geographic area and as a result aspects of risk also vary. A summary of this case study can be found in Deliverable (The TRESPASS Project, D1.3.3, 2015).

2.1. Motivation

In the case of organisations which deploy infrastructure over large areas (e.g. banks), it is unfeasible to apply the TREsPASS methods and tools to analyse the infrastructure. The ATM case therefore highlights the need for an initial screening of risk, followed by application of the TREsPASS methods and tools only on the part of the infrastructure that was identified as being high risk.

Because the initial analysis is carried out at the macro-level, social data needs and extraction methods vary in relation to those used for the detailed analysis described so far in the TREsPASS project. Moreover, in the ATM case study the social characteristics of an area are used as control variables since it is unlikely that any countermeasure implemented by e.g. a bank will be able to modify the social characteristics of an area. In this sense, social data in the ATM case is used to depict the context. However, although unmodifiable by the organisation that controls the infrastructure, this social context is highly dynamic and it should be modelled.

2.2. Type of data

Due to the macro nature of the ATM case study, the source of social data is municipal or national census. That said, in some cities, community-based or expert-based mapping constitute potential sources of data in case census data is made available in units which are too large for meaningful conclusions to be drawn. Across Europe, basically three types of data collection is used:

- The Traditional census (i.e. full field enumeration) collects basic characteristics from all individuals and housing units (full enumeration) at a specific point in time. For example in the UK, France, Luxembourg, Italy, Austria, Portugal. France uses a variation of the traditional census (i.e., rolling census) which is a cumulative continuous survey covering the whole country over a period of time instead of on a particular day.
- Combined census (data from registers + field data collection). For example in Germany, Spain, Poland, Estonia.
- Register-based census. For example in all Scandinavian countries, The Netherlands, Belgium.

According to (Statistics Netherlands, 2014), the traditional census has a slightly greater level of detail but at a general level the results from these three types of censuses are comparable. The generation of census data will not be described in this deliverable as census bureaus across Europe document well how this data is generated. What is relevant for the purpose of this deliverable is to explain the preliminary processing that census data needs in order for it to be integrated with the technical and physical data.

2.3. Method

With the exception of victimisation data, most socio-economic data is aggregated. Aggregation is most often in the form of an administrative unit. However, sometimes this information is also found on the pixel level. Aggregation is an important topic from the point of view of the interpretation of the risk results in cases such as the ATM crime, which have source data with varying levels of abstraction. For example, the ATM data is represented as point data whilst the population data, which was collected at the household level, is represented as polygon data (i.e. non-standard shaped polygons) corresponding to a (4-digit) postal code. For data integration purposes, a first step consists of identifying a suitable unit of analysis. In practice, a geographic information system is able to display data at any level of abstraction since data can be aggregated and disaggregated easily. However, the selection of the unit of analysis is key in ensuring that reliable conclusions can be drawn while minimising processing time.

Extrapolations are often carried out but these come at the expense of reliability. Moreover, the typical timeline of 10 years implies that the boundaries of the units (e.g. neighbourhoods) might change over time since typically, municipal governments opt not for evenly

sized geographic units but for units with similar numbers of inhabitants or houses. This therefore means that as the population increases, the boundary of the administrative unit changes. This issue is known as a 'modifiable unit area problem' (i.e. MUAP) and it is an issue which has been widely acknowledged in the field of geographic analysis (Openshaw, 1984). The MUAP is an issue when a time-series analysis of socio-demographic data is performed. Often, the only solution to overcome this is to derive rates and convert the data into raster format.

Criminology theory shows which are the most common predictors of victimisation. In addition, criminology research shows that some of the relations between socio-demographic variables and crime are not linear and therefore such functions will have to be taken into account in the spatial modelling of ATM risk.

Most often census organisations provide socio-demographic information in the form of attribute data and a separate base map. The process of producing a geographically-referenced socio-economic database involves using Geographic Information System (GIS) software to join both datasets on the basis of a common key. GIS is then used to further process this data and then integrate it with the technical data (e.g. ATM machine characteristics) and physical data (e.g. location of ATM machines). Furthermore, if historical information of ATM victimisation is available, a procedure called Geographic-Weighted Regression can be used, which is a type of regression for spatially-varying relationships.

Since the ATM case relies on census data for controlling for socio-economic factors, the most important procedure involves data normalisation. However, since the census usually lacks populations broken down by e.g. time of day, techniques such as feature buffering and intersection as well as time-use budgets are options for deriving the necessary data.

2.3.1. Normalisation

There are two forms of normalisation needed in the ATM case, one has a broader focus and also applies to non-spatial data whilst the second one is more focussed and is relevant for the handling of spatially-referenced datasets:

- From a data management perspective, normalisation includes the procedure for organising, analysing, and cleaning data to increase efficiency for data use and sharing. This includes in chronological order: data structuring and refinement, redundancy checks, error elimination and standardisation.
- From a statistical point of view, normalisation handles the differences in values for areas that are unevenly-sized. For example, dividing total population by total area yields population per unit area, or density. This includes the procedure for dividing one numeric attribute value by another to minimise differences in values based on the size of areas or the number of features in each area.

2.3.2. Buffering and Intersection

Buffering is a proximity technique used to bound an area at a specified distance from the object (in the case of a point feature) or from all nodes along segments of an object (in the case of a line or a polygon feature). An example of the application of buffering in the ATM case relates to the modelling of the population scenarios. A model of ATM crime should control for population densities to account for the 'Eyes on the Street' (i.e. natural surveillance) effect. Although census data can be used to depict night-time population densities (i.e. number of persons in a household divided by area), day-time population is typically unavailable in census databases. However, a density map could be derived by using several concentric buffers around transport, commercial, educational and industrial establishments individually and then integrating the resulting four maps together in a procedure called intersection. Intersection builds a new area by feature class from the intersecting features common in both feature classes.

2.3.3. Time-Use Instruments

The sort of detailed population density information needed for the ATM case study could be generated by means of time use mechanisms, which systematically record how individuals use their time on different activities over a given period of time. Such research has appeared in several countries and similarly to censuses, some are repeated every five to ten years, for example, in Canada, Japan, the Netherlands and Norway (Harvey, 1999). Such national studies are typically used to find out the daily routines of inhabitants in terms of paid or voluntary work and in terms of recreational activities such as sport activities (Hoeben, Bernasco, Weerman, Pauwels, & van Halem, 2014). These studies have been used to identify a series of (space) time use instruments (known as 'activity-based approach') and mechanisms, which are derived from the work of (Hägerstrand, 1970) and are potentially useful in criminological research:

- Stylised questionnaires ask people to indicate how much time they spend in certain activities for a given time period (e.g. 'an average day').
- The time diary method (i.e. 'time-budget') asks people to record every major activity carried out during a given time period.
- In the experience sample method, respondents are given signals via e.g. a phone or electronic pager at random points in time and they have to note down their current activity. This method enables the recording of brief activities that are underreported in the time diary approach.
- Secondary data from the supply-side such as that of recorded attendance to events.
- On-site verifications count the number of people preset at a particular location at a given point in time.

- Direct observation involves following a person from a distance and annotating (i.e. observations) the activities and social contacts. A variant of this method is spot sampling where observations are made at random points in time.
- The data extracted by means of e.g. the time diary method can be aggregated at the level of a relevant spatial unit to generate population-density maps.

2.4. Envisaged use

This case study illustrates a risk analysis at a macro level for identifying priority areas in need of detailed analysis using the TREsPASS methods and tools. This approach is inspired in 'hotspot policing' which relates to strategies and tactics focused on small units of geography where crime is highest. The reasoning is that hotspots are small places in which the occurrence of crime is so frequent that it is highly predictable (Sherman, 1995).

2.5. Example input and output

In this section we provide an example of the output map resulting from referencing raw socio-economic data from the Portuguese census bureau. The procedure involved using a GIS to convert the data into rates (persons per hectare) per administrative unit (Figure 2.1). As an example, this social data was integrated with technical data to produce an ATM risk map for a given attacker profile (Figure 2.2). Finally we also suggest two methods two (i.e. Buffering/intersection and time use diaries) for deriving the necessary information to develop ATM risk scenarios broken down by relevant temporal classes.

The risk scenarios described above could be used to enrich the Attack Navigator Map and creates the possibility of linking with external data feeds such as crime databases for a particular area.

The second example presented involves the notion of the space-time prism (see Figure. 2.3) (Pacione, 2009) from which a time diary survey sheet can be developed to collect the necessary information to map population densities by hour of day, day of week or month of year. This procedure would allow to cover the social data gap explained before.

2.6. Summary

In summary, the techniques described in this chapter enable risk assessment practitioners to produce detailed spatio-temporal social data for carrying out an ATM risk assessment at the macro level. The idea behind this case study is to show how to make a quick analysis that allows the analyst to select high risk areas for further analysis using the TREsPASS methods and tools.

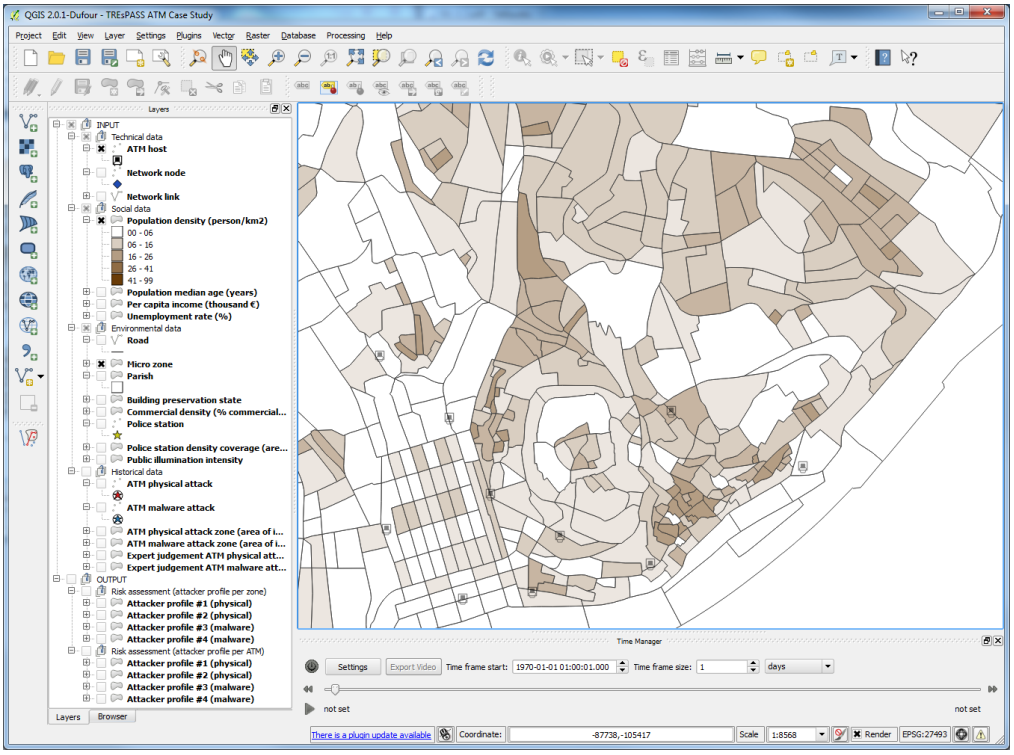

Figure 2.1.: The map shows the results of the normalisation procedure by converting census population density figures into rates (i.e. density by area).

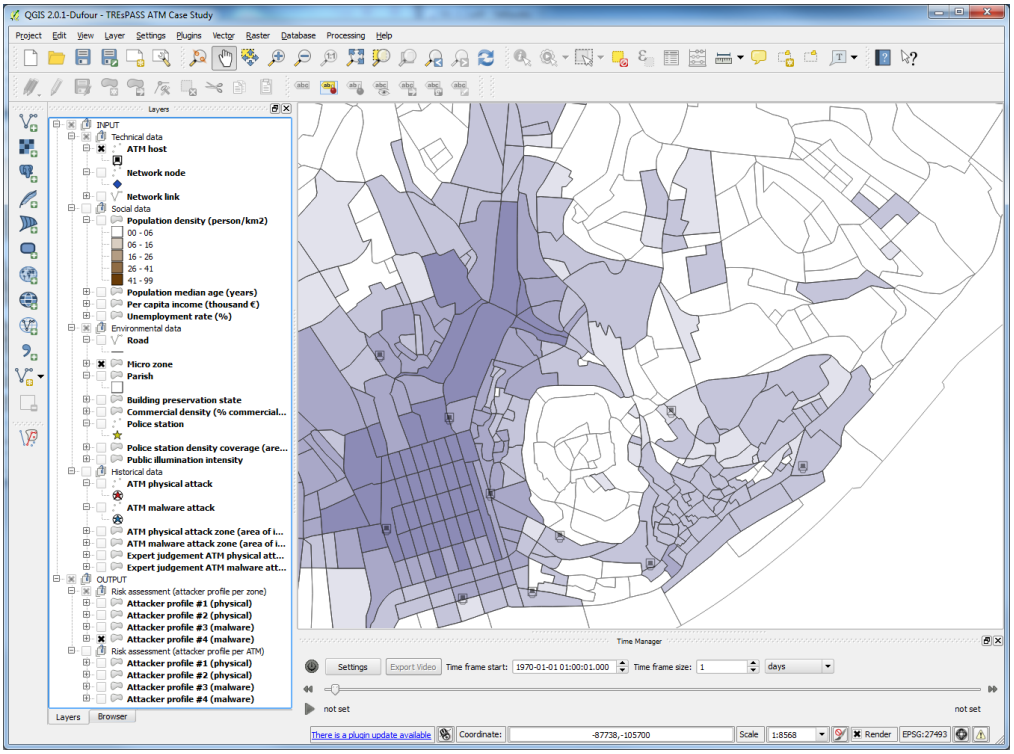

Figure 2.2.: The map shows the location of the ATM machines and the risk level of the various geographical units (darker colours denote higher risk).

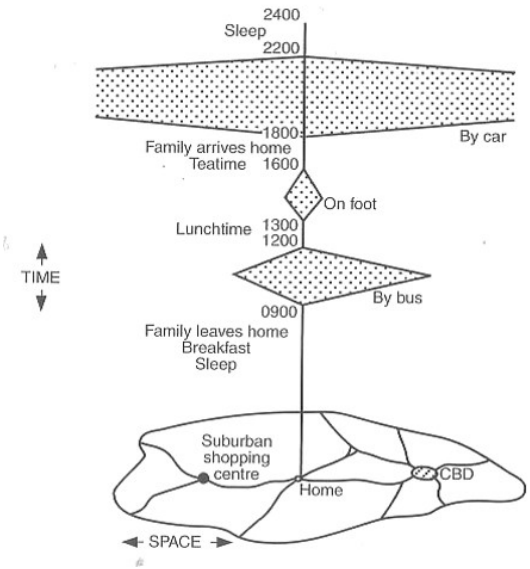

Figure 2.3.: The map shows population density originating from the census after it has been converted into rates (i.e. density by area). Source: (Pacione, 2009)

3. Gathering social data: Stage-Zero risk assessments

This chapter introduces a qualitative data gathering method using rich pictures to gather data about context and to identify the relationships between internal and external context. The chapter then goes on to show how the analysis of such data can be quantified.

3.1. Motivation

Recent research in the field of human-computer interaction (HCI) and in organisations, has involved groups of stakeholders engaged in participatory modelling, which in turn provides a description of information-exchange practices. This research on 'serious play' has informed the present chapter (Schulz & Geithner, 2013), and includes exploratory work that models organisational practices in some detail (Roos, Victor, & Statler, 2004) and strategic planning (Bürgi, Jacobs, & Roos, 2005).

The seemingly intangible aspect of social behaviour and of information-communication practices very often affect the core functioning of businesses. Yet the human dimension is very often glossed over in the study of cyber-security, humans sometimes being referred to as the 'weakest link' (West, Mayhorn, Hardee, & Mendel, 2009) in a chain of information custody. What can be easily observed is that differing degrees of trust and solidarity within an organisation, can lead to contrasting perceptions of security, and the values associated with it, and these are very difficult to visualise (let alone quantify) unless the input of stakeholders is explicitly sought through an active participation in studies.

It has been noted that the concepts of positive and negative security are a useful means of rationalising the varied types of social data that are gathered in a number of different ways (Chapt. 1). A participatory process such as the one described and advocated here, insists on continuous iteration within the information-visualisation process, and is ideally suited to the gathering of contrasting interpretations of a scenario, in effect brainstorming the full range of positive and negative implications of its facets. Furthermore, in the post-analysis of this data, there are inevitable difficulties concerning how to represent time and change in relation to vulnerabilities. The participatory process is also well suited to this issue, addressing it by insisting that the process remains recursive. Maintaining this tack should enable a security analyst to take account of the multiple perspectives of several actors and the nature of their relationships.

3.2. Methods

A specially developed form of participatory diagramming and physical modelling has been used, with a view to visualising networks of trust and solidarity, placing social data gathered directly from case-study participants at centre-stage. This has the effect of broadening the process of risk assessment, accessing social data as a starting point for identifying and then scoping the issues that are of paramount interest to the stakeholders. This has the effect of narrowing the field of enquiry, and refined technical types of data that can be used to reciprocally interrogate one another, in terms of both visualisation and analysis.

A four-stage case study was undertaken. The first stage used the *Archimate* framework (Lankhorst, Proper, & Jonkers, 2009) to traditionally model the risks to the design of a micro-payment service to be implemented using Internet Protocol TV (IPTV). The risks elicited in this stage did not reflect the networks of trust and solidarity that were very apparent in the security thinking when interviewing the service providers. In the next stage the service providers identified their core values and the basis for engagement with their customer base. In the last two stages of this process, the participants were given *LEGO* building bricks of given types and colours, selected so as to encode the movement of shared information and data, actors, and devices. The above-mentioned *Archimate* framework for enterprise and risk analysis is referred to here, using a similar colour coding (in terms of the colour of bricks) for the social, technical, and infrastructural dimensions of the scenario. At the same time, the organisational core values that had previously been mapped from early engagements were carried through the subsequent stages of analysis and interaction with the participants (Fig. 3.1).

3.3. Type of data

All data entries have purposely originated solely from the actions and utterances of the participants, and the entries for each data-line are restricted to what has been physically built and has been said by this group, grounding the categories for enquiry within the data rather than importing external criteria for these. This has been called a 'grounded' approach (Charmaz, 2011) to qualitative research methods for data gathering (Denzin & Lincoln, 2009).

The data can be managed in spreadsheet documents, for export to visualisation and other data management tools. The data fields are designed so as to capture:

a). the order in which elements of the representation are constructed by the participants,
b). the relative importance given to these elements within this representation, as determined by brick counts of the different colours, and the height of individual structures that represent the actors, data, and infrastructure,
c). the speech surrounding the co-construction of the models, so that the conceptual content of speech is traceable and accountable (to the group as

a whole rather than to individual members of the team, whose anonymity is protected).

This is based upon the assumption that valuable information about patterns of data sharing will be encoded within the representation, and that some part of this may be extracted by recording the physical layout of the model that participants co-produce, and tying this to the positive and negative concepts of security (see Sect. 1.6.3) invoked by the group at that time.

The data file contains 14 fields (Sect. 3.5):

```
Description, ID reference, Size, Participant speech, Timecode, Adverse Keyword,
Supportive Keyword, Green bricks, Blue bricks, Yellow bricks, Orange bricks,
White bricks, Pink bricks, Class
```

These are described in more detail below:

1. **Description:** a summary of speech and the overall discussion before and after the particular instance being referred to.
2. **IDs or reference points:** locations identified by name by the group on the physical model (Numeric).
3. **Size and height:** the quantity of bricks and other parts used to represent an element of the model (Numeric). Separate data columns also give the combined scale (mean of size and height).
4. **Participant's speech:** some data points refer to shared agreement on a specific point, at other times detailed speech is recorded.
5. **Timecode:** the timespan of the comment (Numeric).
6. **Keywords: Adverse.** Number of times occurring in speech (Numeric). Separate data columns also give the keywords themselves.
7. **Keywords: Supportive.** Number of times occurring in speech (Numeric). Separate data columns also give the keywords themselves.
8. **Green:** Count of *LEGO* colour used to represent infrastructure (Numeric).
9. **Blue:** Count of *LEGO* colour used to represent data and data-flow (Numeric).
10. **Yellow:** Count of *LEGO* colour used to represent actors (Numeric).
11. **Orange:** Count of *LEGO* colour used to represent the required innovations that have been identified (Numeric).
12. **White:** *LEGO* colour used to represent uncertainty or unknown entities (Numeric).
13. **Pink:** *LEGO* colour used to represent additional countermeasures (Numeric). The inclusion of this is with the second *LEGO* session with the case study participants in mind.

14. **Class:** basic types of entities represented, Actors, Infrastructure ('Infra'), Data, Social and organisational (abbreviated to 'Social'), and Countermeasures.

3.4. Envisaged use

In this section we present a number of potential use-cases, and some of their sub-tasks. These have been identified as being relevant to the social data and policies that have been extracted by the proposed techniques. As has been mentioned, this type of exercise is especially suited to stage-zero risk assessments where a new business or service is being designed and a rapid and insightful procedure is required in order to narrow down the field of enquiry to those areas that are deemed critical. Moreover, the methods described here could equally be relevant for existing businesses and services, especially in the case of a first risk analysis, or if a critical look at existing risk analysis is required. This could be described as an extension of the pen-and-paper 'brainstorming' group exercises carried out routinely by many organisations.

The data discussed here has been structured in such a way that these policies and other concepts can be visualised effectively within a graphical user interface, assisting with the wider project aim of providing analysts with thinking-tools on which to base their decisions, and making this type of complex social data amenable to the structured approach that this type of analytical tool requires. Repeating the insights gained from these analogue engagements, but this time incorporating them into a digital tool, adds a rich social dimension to the technical picture, and should on this basis help analysts to sift through a mass of undifferentiated technical data.

3.5. Inputs and outputs

In this section we describe specific data inputs for the proposed techniques, presenting a representative sample piece of input data, along with an example of a typical output that results from the application of the technique.

3.5.1. Example input

Throughout the engagement process leading up to and including the physical modelling, care was taken to interleave feedback regarding core business goals and concepts obtained during the early briefing stage, refreshing the enquiry with current and previous value and goal alignments within the organisation. The raw data obtained from our case-study fieldwork spans session recordings, hand-made notes and drawings, the physical models coded by reference points, infrastructural diagrams originating from the organisation, and our own diagrams made in order to encapsulate early findings (Fig. 3.1). As mentioned previously, the data is structured by the predetermined colour scheme for

the bricks (based upon the Open Group's 'Archimate' schema) (Lankhorst et al., 2009) representing different facets of the scenarios.

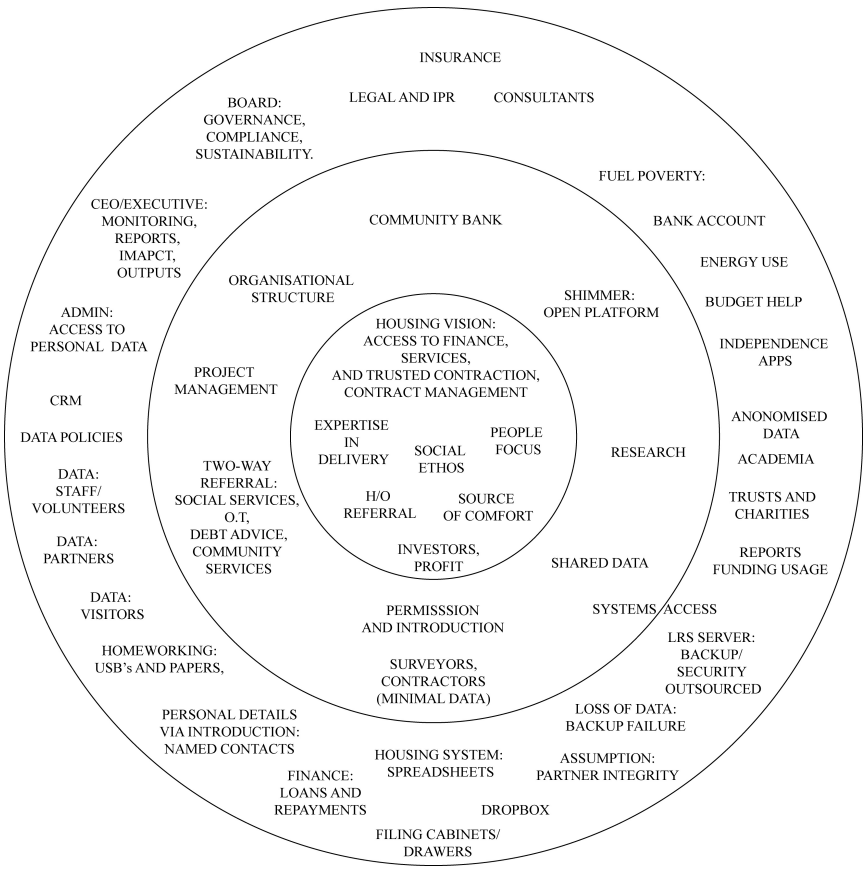

Figure 3.1.: Case-study: a picture of an SME's natural areas of interest, concern, and resilience as a social enterprise business based in London. The innermost circle shows the SME's central goals and values, the next concentric circle shows their tools collaborators and partners, and the outlying circle shows their out-facing components and partners that may on occasion be classed as potential adversaries or competitors.

3.5.2. Example output

The output would be a data file upon which keyword queries can be carried out, the search providing the reference number of the parts of the model associated with this keyword, plus descriptions of the social aspects referred to at these points, and detailed statements made by the participants directly related to this keyword. The level of detail required by a user can be filtered, so that less of statements are provided, for example. Queries on the output data result in clusters of data-points that contain values and goals obtained directly from participants.

To take one example, the data is queried for 'impact', resulting in a number of actors and other nodes that are implicated in this concept, and the associated statements also reveal where this concept has a bearing on relationships that exist between these actors and the supporting technology.

The output, a cluster of participatory data related by the concept of 'impact', can then be further analysed and have visualisation techniques applied to it, as needed, as part of a user interface. The keyword query is shown in blue here (Fig. 3.2) as a text highlighting keyword occurrences in the data (yellow), reference points on the physical model (green) and size of the referenced element (purple).

The resulting first-pass analysis is described below (Sect. 3.6). While images have been used in this Chapter to present the general approach to analysis of inputs, these images are explicatory only. A related research strand seeks to generate visualisations of the data complementing the analysis described here and communicating our findings to other Work Packages (Figs. 3.3 to 3.6). This two-strand approach will bring out the visual, qualitative aspects of the data (contained in the model itself and in the comments of the participants), and provide a representation that is richer and more heterogenous than those that formal methods are capable of providing.

3.5.3. Operationalisation

Lastly, a series of templates have been made, with a view to the continuation and extension of this work:

- A template for annotating upon recorded speech and/or video.
- A template for saving data in spreadsheet format and in comma-separated format.
- Protocols for translating the data into other forms, formal languages such as UML, Archimate, and formal graphing methods (a current line of development).

Subsequent transcriptions of the same group of participants, who continue their work on the initial model in a follow-up session, can in principle be incorporated into a growing and increasingly refined and focused data-set. Such a data-set can theoretically be used to carry out a sustained examination of complex security-related issues connected to a particular service.

3.6. Discussion

No analysis of the data can be achieved without ranking the data in some way, that is, by sorting it according to one dimension that is of special interest. In this case, the dimension of interest is the occurrence of positive and negative keywords or tags upon the data. These keywords and tags give a window onto the highly detailed data that is comprised of the transcribed participant dialogue, if such a window is desired. They also, importantly, allow any imbalances between positive and negative comments to be identified, enabling

```
Intervention and alerts may impact on Client housing,23;32,26.75,"that could be the housing
association, it could be, any, any of the partners",01:03:42.136 -
01:03:47.299,alert;impact,,intervention;partner,,,,,,,

Intervention and alerts may impact on Client housing,23;32,26.75,"yes, it could be",
01:03:43.590 - 01:03:44.973,alert;impact,,intervention;partner,,,,,,,

An event with impact on the family provokes further interactions concerning budgeting,
1;20,2,"like, something must happen to the family to make them need to interact and to budget
and things like this",01:10:52.258 - 01:10:57.503,impact;provoke,,family,,,,,,,

There is an impact from alerts on the Clients family from banking and further impacts of Client
circumstances behind this that the Partners need to be aware of,1;22,2,"is that we're talking
about an impact of something outside here, but there could be impacts here, that, drive that",
01:11:38.911 - 01:11:45.105,alert;impact,,family;partner,,,,,,,

There is an impact from alerts on the Clients family from banking but also impacts of Client
circumstances behind this that the Partners need to be aware of,0;23,18.5,"and that we need to
be, or Advice UK for example, needs to be cognisant of",01:11:46.669 -
01:11:51.237,alert;impact,,advice;family;partner,,,,,,,

Exorbitant Energy demand is the external driver for this scenario,36,16,"how about we just have
external event there, and a little link there, and that event could be anything?",01:12:00.706
- 01:12:05.684,impact,,,,,,,,,

The direct impact of Energy demand on the Client-side is seen in the red cones,37,2.5,and it's
directly impacted the house,01:13:25.767 - 01:13:27.271,impact,,,,,,,,,

Energy demand has a severe impact on the Client's family,37,2.5,it's a very very close
broadside as well,01:13:32.346 - 01:13:34.230,broadside;impact,,family,,,,,,,

Energy demand has a high impact on family,36,16,so it's a huge issue,01:14:01.156 -
01:14:03.127,impact,,family,,,,,,,

The impact of the Energy demand that is felt by household should be reflected in this
representation,36;37,18.5,this is a big issue because that's what it feels like to the
household,01:14:25.918 - 01:14:28.954,impact;household,,household,,,,,,,

Serious impact of income variation for Client,39,4.75,when something goes wrong the wheel falls
off,01:19:43.560 - 01:19:45.898,impact;wrong,,,,,,,,,

Discussion of how the complex relationships within Change Account will impact upon the Client,
0;16,27.75,"actually it's really interesting, thinking about how all this might work, hadn't
thought about it quite like that",02:03:24.540 - 02:03:30.736,impact,,,,,,,,

Discussion of how the complex relationships within Change Account will impact upon the Client,
0;16,27.75,"you know, how you, I hadn't really thought about that as a collaborative early
warning system",02:03:32.910 - 02:03:40.370,impact,,collaborate,,,,,,,
```

Figure 3.2.: Sample data, based upon a query upon **'impact'**, from an earlier version of the data-file. The resulting sample shows the reference points (marked in green), places at which this concept is invoked during the modelling process, and descriptions of those utterances (marked in yellow) made by participants that relate to the keyword, 'impact' (blue). The size of the modelling element is given (figure marked in purple), and fractions relate to the number of studs on each brick counted, where a unit of one is equal to 4 studs. For objects other than studs an approximate estimate is made based upon equivalent sizes.

us to locate potential 'blind-spots' concerning specific nodes within the representation. Finally, identifying these spots, and having these contextualised by a representation of the social and organisational practices that surround them, it can be ascertained whether

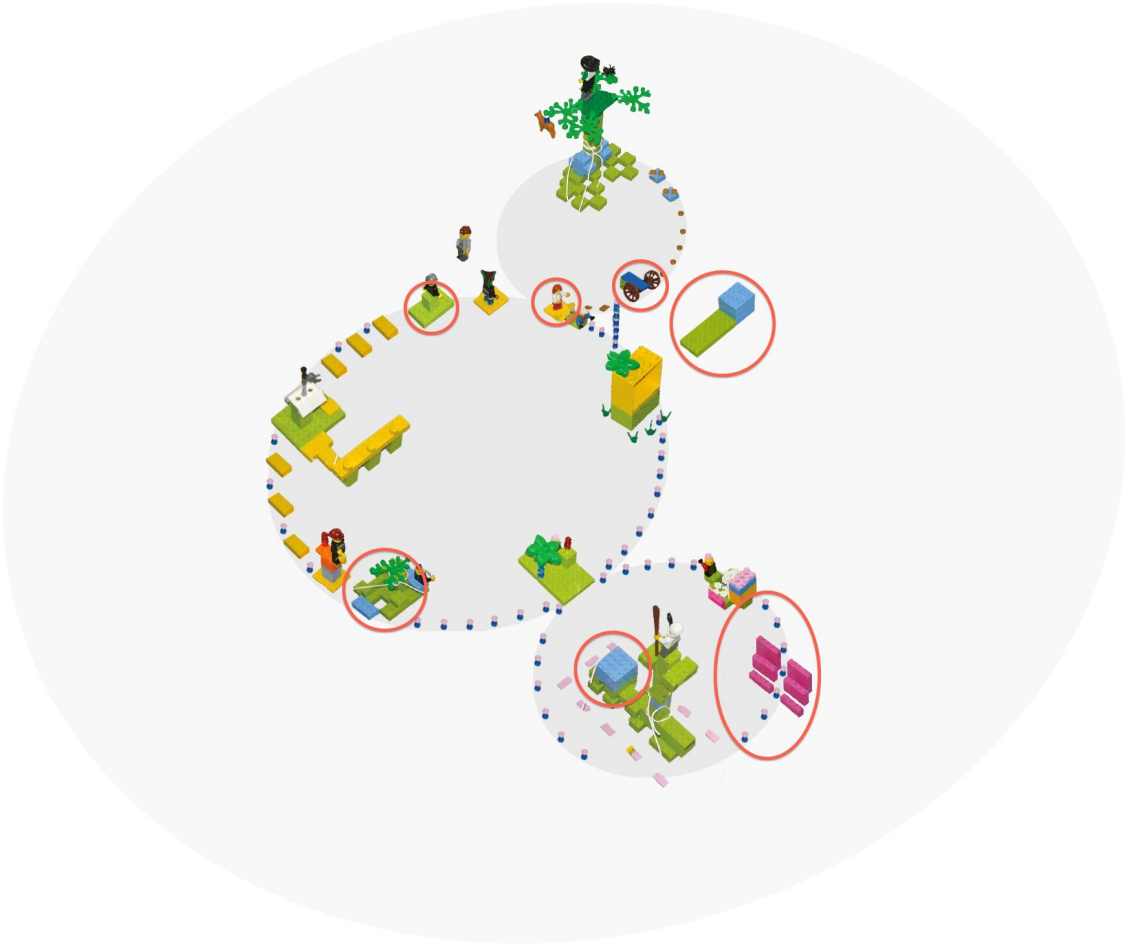

Figure 3.3.: The elements of the *LEGO* model rearranged in a digital collage, making it easier to see the flow of the relational service. The client (located between the upper and middle circles), has been connected to the notion of 'impact' in the data file, and is highlighted in red as are other nodes. The central area defines the essential relationships that are required for the smooth transaction of the service, and this is supported by the outlying banking (bottom) and state systems (top).

client data held at these places is at risk, as it is transferred between nodes during the provision of a service.

Moreover, alternating between atomised views ('static' visualisations that utilise the continuing metaphor of the building-brick) and views of wave-like transitions between states that the modelling passes through, given the above, allow a security analyst to pinpoint different types of data, events, and contexts, and to address the different kinds of risk to data that are present at each locale.

As the analyst drills down into the data they will see how the positive/negative keyword equation can be associated to certain nodes and places. it will also be apparent that this

will be a natural extension of the way in which participants have discussed problems and issues. For example, in the area of assessing potential 'impact' it was clearly natural to link positive mitigations to their own service design, while linking possible negative security keywords to areas of their client's lives that are subject to reverses in conditions, and over which the service providers have no direct control. The discussions in this case were focussed on how the design of their service might help to improve these unpredictable areas of their client's circumstances, and specifically in order to prevent any unnecessary and potentially harmful impacts from occurring.

Areas that are flagged up for attention in this way are the "excessive payment demands" sent by energy companies, "income variation" experienced by the client, and "interactions concerning budgeting" (see Fig. 3.6, and Tab. 3.1). The Table shown here gives the top-three risks for the client and their family in the IPTV scenario. This is assessed by means of the level of likely impact that unpredicted events in these areas might have, based upon the comments made during the session. The method trialled here is therefore a means of recording the everyday concerns and interests of actors, including clients and the service providers, within the context of the co-design work - where the question is asked: when building a representation of this service, what appear to be the areas where resilience is most fragile? This information is then gathered in such a way that is able to link values to these areas, and is made available to visualisation and business intelligence tools.

Table 3.1.: Top-three rating of risk/impact areas, specific to the IPTV client and their family. These risk areas were uncovered by co-design work with the service designers, and resulted in reinforcement of critical points in the system, by enhancing the breadth and refinement of controls at these points.

Risk rating	Description of risk
High (H)	Sudden and large energy demands
High (H)	Unplanned income variation impacting on client's resilience
Moderate (M)	Missing window to intervene in family budgeting interactions

Ranking the instances of positive security mentions, those cases where the engagement participants refer to topics that support social practices, allows for the remaining data to be seen against this backdrop (Fig. 3.5). Essentially the *LEGO* data can be used to add light and shade to the basic nodes of the representation, and by doing so the client's and other perspectives upon the scenario can be introduced and encoded within the formal output of the stage-zero risk analysis, such as one that can be constructed from the same data using Universal Modelling Language (UML), for example (Fig. 3.7). If the positive and negative values are combined, a tonal value can be derived. This value can be based upon the numerical aspects of the *LEGO* data, and also be made available to an interface user as a visual cue for decision-making, since this information is based upon a description of the areas of the model where 'impact' will be most keenly felt, to take one example.

It is possible to see how an overview of a scenario can benefit from a coarser-grained summary of these positive-negative points, as described above. However, if an interface

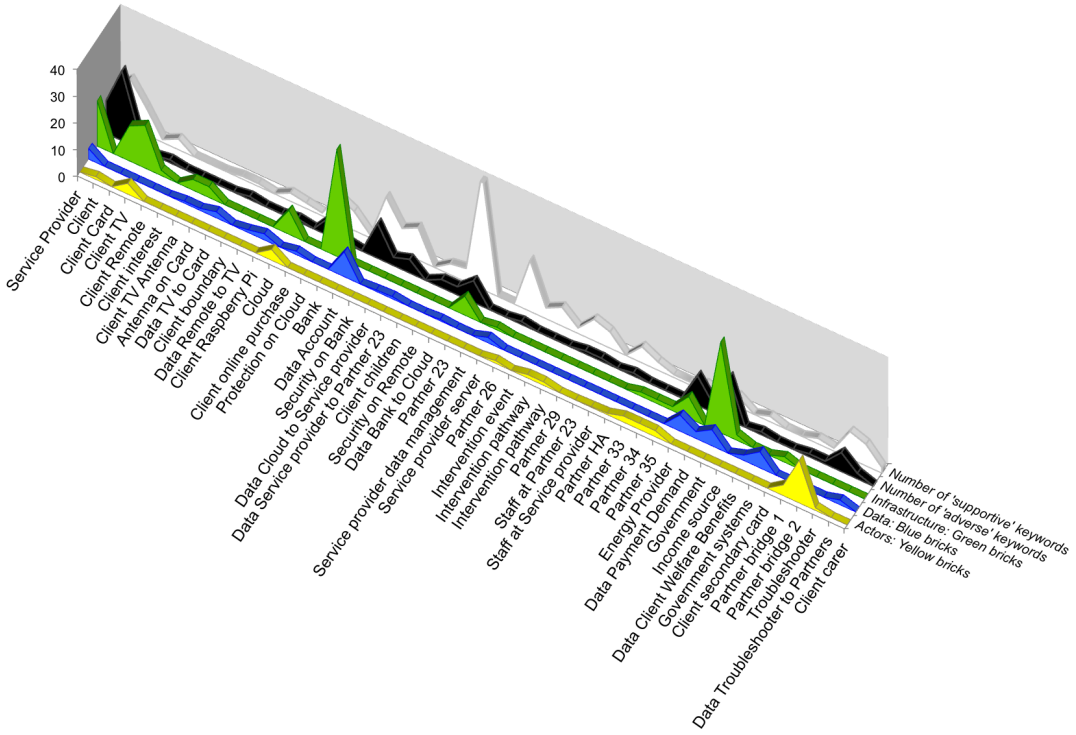

Figure 3.4.: Interim visualisation: a coarse-grained view, representing an entire physical modelling engagement (3 hours), and showing the sequence of elements as they were added by the group of participants. Counts of the occurrence of positive (white) and negative (black) keywords are indicated. The z axis presents actors (yellow), behind which data (blue) and infrastructure (green) can be seen. White and black are presented one in front of the other on the z axis, in order to better compare peaks in counts of positive and negative keyword occurrence. It is somewhat difficult to see overarching patterns in positive and negative security in this interim visualisation, and the following Figures address this difficulty by sorting data according to categorical and numeric values (Figs. 3.5, and 3.6).

user also wishes to drill deeper into the data, a finer-grain view will be needed. This should present detailed information about the components of a given scenario, as well as offer the interpretations that were gathered relating to certain of its aspects. This can be seen where the data has been queried regarding 'impact', revealing which nodes of the model have been linked to the concept, and which elements of the clients own data are linked to these nodes (Fig. 3.6). The client's data that resides or passes through the data-management systems of the Service Provider, and the Banking system, could be deemed vulnerable, according to this analysis of the participatory data. We can immediately see that the larger black areas at these nodes (referring to potentially adverse keywords that

were in use), are not counterbalanced by an equalising mass of white area (potentially supportive keywords in use).

This approach to sorting the data is necessarily binary (black and white) in the first instance, but it is intended to provide pointers towards areas of the data that require deeper analysis, and to begin the process of nuancing the zones of dark and light so that they become more graduated and well-understood.

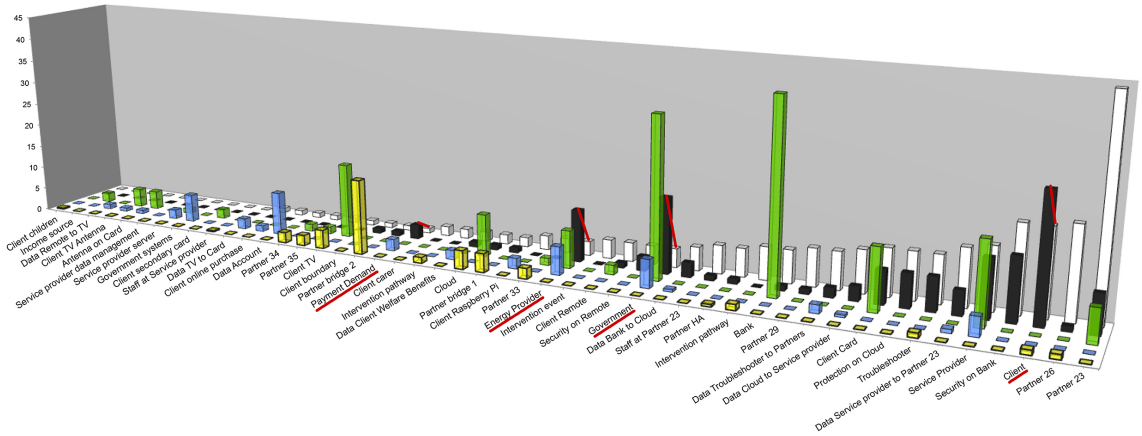

Figure 3.5.: Interim visualisation: the most obvious examples of imbalances towards negative security (the black columns) are in the areas of energy demands upon the client, the energy provider, data and policies originating from central government, and the client of the IPTV service. These areas have been highlighted with red. The white columns representing positive security comments are sorted in ascending order from left to right. The sorting relates to the sum of the count of positive keywords that are tagged to each ID. IDs are given along the y axis. This allows us to see the description of the particular elements and processes that were alluded to as the physical model was being built. Colour-code here follows the *LEGO* bricks, as in previous Figures.

3.7. The role of the stage zero approach

The particular contribution described in this chapter has been aimed at encapsulating the way in which case-study participants have represented scenarios in terms that they themselves would recognise and, importantly, take ownership of. In so doing, we have key identified social data that is related to values and goals that very probably would not have been derived from a traditional risk assessment. This could be described as an internally consistent view, rather than one imposed from an external source. The outcome of this approach is that analysts can identify key social and organisational policies and practices in existence and those could potentially be reflected in the Attack Navigator Map (ANM), either through annotations that are added to the models created in the ANM, or as a starting point for creating a map, perhaps through the facilities recently added

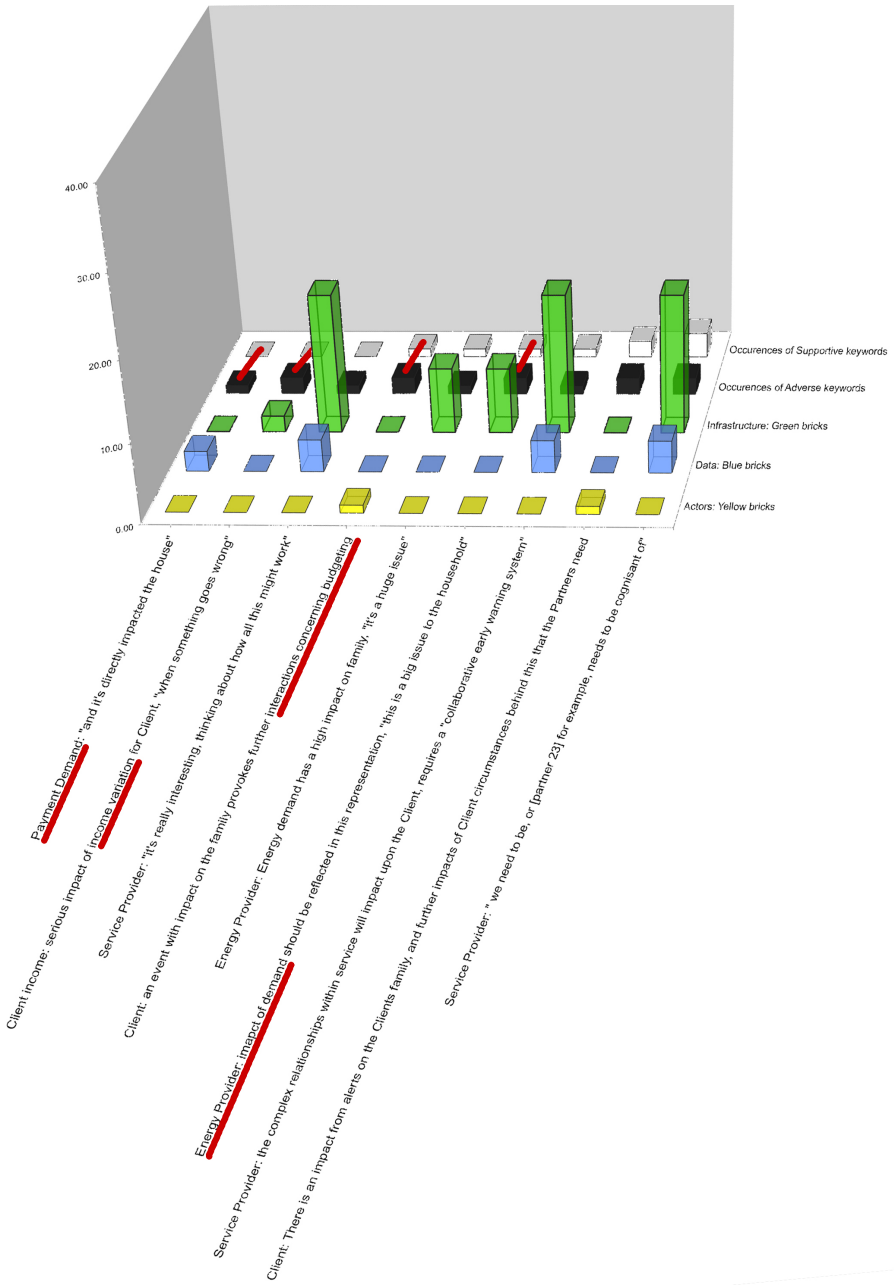

Figure 3.6.: Drilling down into the data, showing where potential 'impact' is not mitigated by positive security: where there are a higher number of negative keywords occurring (black columns) and where this is not counterbalanced by an equal number of positive keywords occurring (white columns), the data associated with this element of the *LEGO* model can be assumed to be vulnerable (in different degrees). In this chart specific information on 'impact' is gathered, concerning the client and the provider of the service. Colour-coded bricks, as in previous Figure.

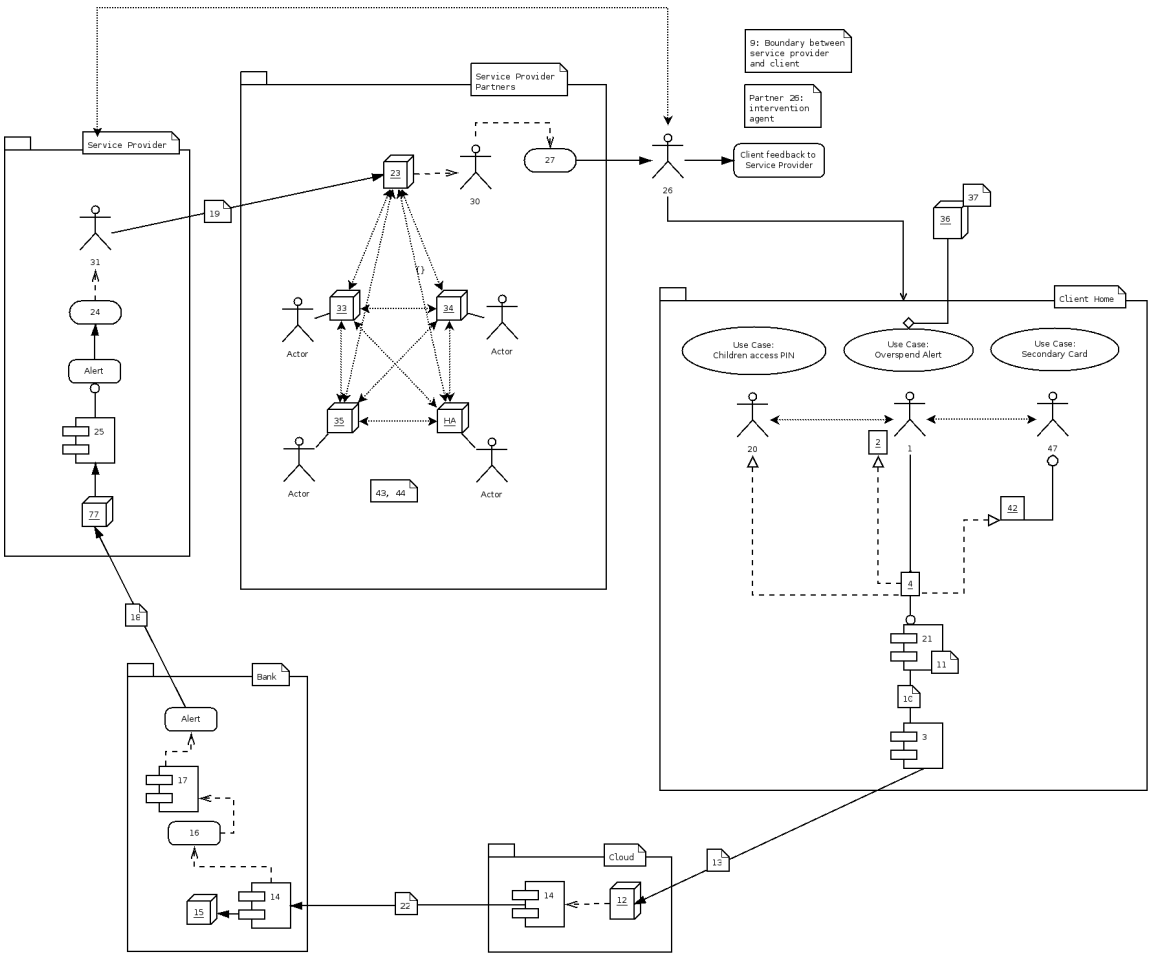

Figure 3.7.: Mapping the *LEGO* model's elements into UML format (Universal Modelling Language). This is a *UML use-case diagram*, representing three use-cases (represented by the convention of oval shapes). These were discussed at length during the session: 1. Children access PIN, 2. Overspend alert, and 3. Secondary payment card. Some locations are grouped as packages and connections between them are annotated with the IDs and descriptions of the corresponding parts of the *LEGO* model. This particular diagram has been arranged so that the relational service is seen as a rough circular clockwise movement of data from the client (right) to service provider and partners (left). This is a similar layout to the digital collage of the *LEGO* model (Fig. 3.3).

to the interface-design prototypes, whereby the user can edit the perimeters of attacker profiles, giving different aspects different weightings, or they can 'trace' or follow the points of an imported mapping, such as organisational floor-plans (Fig. 3.8), or indeed, the visualisations of co-design data such as those that have been described in this Chapter .

Why are the sequences, numbers and heights of brick assemblies a relevant metric? Because they encode the participants representation of relative importance, and the ordering

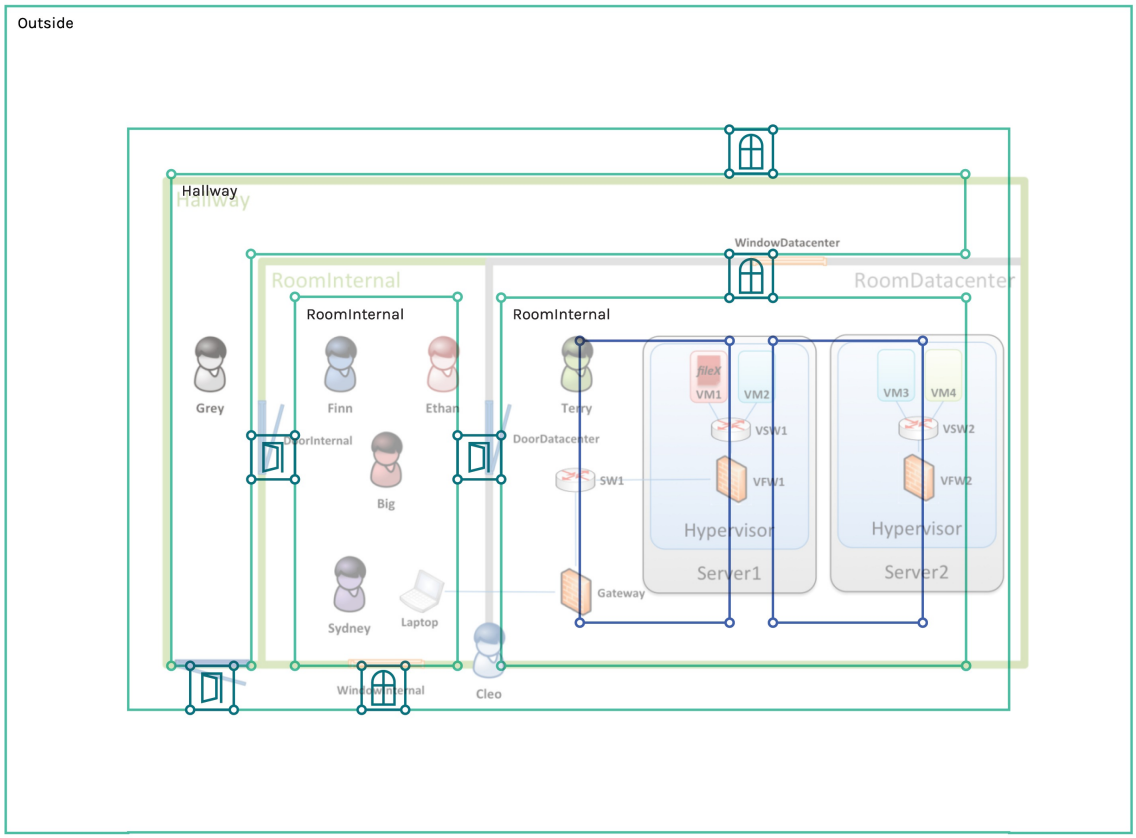

Figure 3.8.: **ANM**, showing the way that the the user can import a floor plan to work from, for example, and can trace over it, using the import to guide the construction of a new or existing model with the tools of the ANM. Locations have anchor-points that can be dragged around to match the locational object with the physical spaces on the floor plan. As well as using floor plans and technical diagrams, the user can also import photographs of rich-pictures models, including *LEGO* co-design work originating from the 'stage-zero' phase of risk assessment. Note that the visualisation of social and technical data, maps, scenarios, and countermeasures is afforded by the inclusion of this facility. Image from LUST, The Hague.

and type of connections or relationships that exist between actors for example. Moreover, the method also allows for the identification of omissions in the representation, since they may not be perceived as having a bearing on ontological securities around a service. Encoding in this way makes it potentially possible to connect the building of the *LEGO* model with the attack pattern library. In particular it provides data on control strength both in terms of social cohesion within a particular community and strength of a technological control.

When the numeric and ID values from the *LEGO* representation are added to the speech of participants (which have been coded for positive and negative keywords), it becomes possible to highlight certain spaces, where people feel secure, or otherwise, where people

feel there would be the greatest impact. This capability could be used by the Attack Navigator Map (ANM) to develop the narrative and better describe the context of the risk scenario.

The 'impact' example presented in the Sections above is not necessarily representative of all cases covered in the data but it does provide an example of a typically complex and interwoven issue that will be of interest to analysts. The operationalisation strategies mentioned above (Sect. 3.5.3) show how these methods can be transferred between case-studies, and how the granularity of the approach can be modified according to requirements. For example, the IPTV case study data described in the previous section, is reasonably detailed (including references to participants speech and also the physical models of participants).

By using the type of participatory data discussed here, we discover what constitutes a typical and routinely *'sufficiently secure'* state of affairs for the participants. This is data structured in such a way that it results in what philosopher Nelson Goodman called a 'graphically *replete* representation' (Goodman, 1976). In other words, this is a *rich* picture well-supplied (or 'filled') with appropriate information, without overwhelming the viewer or analyst with unnecessary detail. What matters with a representation, Goodman says, 'as with the face of an instrument, is how we are to read it' (p.170). An interface-design and visualisation strategy should naturally emerge as a readable solution for typical users. This requires a balanced immersion in both qualitative and technical data, straddling both the diagrammatic and the pictorial conventions, and taking the best of both worlds.

4. Social Engineering Success Stories

This chapter outlines the proposed methods for gathering social engineering success stories and how narrative analysis can be used to increase practitioner know-how, develop understanding about attacker techniques and point to additional variables that can be used to build an attacker profile.

4.1. Motivation

This section discusses the method that is used to obtain an insight into the *modus operandi* of a social engineer. There is a specific focus on the psychological mechanisms that are used to make targets comply the requests. The output data is useful for the development of directed and targeted interventions.

4.2. Type of data

The input data are qualitative in the form of written scenarios. The scenarios that are used in this analysis are from books on Social Engineering:

1). *The art of deception: Controlling the human element of security* (K. D. Mitnick & Simon, 2002),
2). *Ghost in the wires: My adventures as the world's most wanted hacker* (K. Mitnick, Simon, & Wozniak, 2011),
3). *Hacking the human: Social engineering techniques and security countermeasures* (Mann, 2008) and
4). *Social engineering: The art of human hacking* (Hadnagy & Wilson, 2010).

Other sources can be used as well, if they meet the following criteria: i) social engineering involves a non-technical social attack against the operator of a computer system and, ii) the narrative (contained for example in a book) contains case studies illustrating the use of social engineering.

The output data are mainly quantitative:

- "Persuasion Principles" (authority, commitment, liking, conformity, reciprocity and scarcity; discussed in detail elsewhere (Cialdini, 2009)). Six dichotomous variables were dummy coded, where 0 = not used, and 1 = used.

- "Modality" (telephone, face-to-face chat). Categorical variables regarding the medium used in during the attack.
- "Offender Goal". Open-ended question.

The variable that is qualitative:

- "Other methods" Open-ended question that can be used if non of the six persuasion principles fit adequately.

4.3. Method

To ensure agreement between multiple researchers, they:

1. processed a description of the Persuasion Principles.
2. performed coding on a training dataset containing 5 scenarios.
3. discussed the outcome of the training results.

After the training, all scenarios (Figure 4.1) were split into attack steps, each containing single interactions between two individuals. For example, if the offender first talks to EmployeeA and next to EmployeeB and finally to EmployeeC, the scenario is split into 3 attack steps (Figure 4.2). The persuasion principles used by the offender were coded for each attack step (Figure 4.3 which shows 3 interactions containing 1, 2 and 2 persuasion principles respectively).

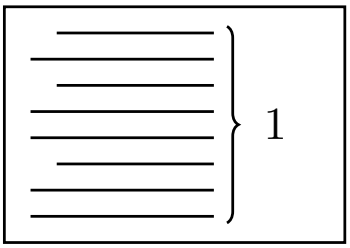

Figure 4.1.: Dissection of a social engineering scenario: One Scenario

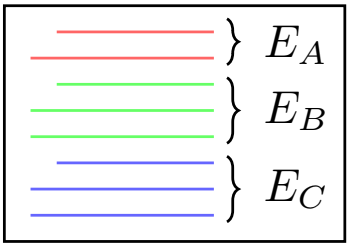

Figure 4.2.: Dissection of a social engineering scenario: Three attack steps

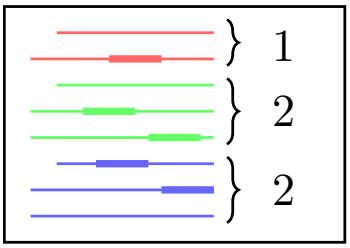

Figure 4.3.: Dissection of a social engineering scenario: Five persuasion principles were found in this scenario

Readers All scenarios were independently analysed by 2 researchers. An inter-rater reliability using the Kappa statistic was performed to determine the consistency among researchers. The researchers' inter-rater reliability was: κ = .909 (p < .000), 95% CI [.874, .944], N = 852. The N represents 142 attack steps times 6 possible persuasion principles per step. The results indicate there is an almost perfect agreement between the two researchers (Landis & Koch, 1977).

4.4. Proposed use

The proposed use for this technique is 2-fold:

i) This technique can be used to transform qualitative data (e.g. scenarios from a book, or stories based experience) to quantitative data that allows for statistical analysis. Furthermore, it allows to make more a systematic approach for combining, comparing and analysing scenarios.

ii) The results of the analysis can be used as a sub-attack description as part of the bigger picture. Since the data sources are 'performed perpetrator scripts' and only contain success stories, this gives some insight in what is used often and what is used incidental. This knowledge can be used in designing countermeasures. One practical outcome is that in a social engineering attack very frequently the authority principle is used, a countermeasure could be to make employees verify the legitimacy of an authority.

4.5. Example input and output

4.5.1. Input

Figure 4.4 shows the dissection of a 'real' scenario from *The art of deception: Controlling the human element of security* (K. D. Mitnick & Simon, 2002). The scenario contains two attack steps, where each attack step contains one persuasion principle. In both attack step 1 and 2 the offender uses impersonation together with the authority principle. In attack step 1, this was achieved by claiming to be an attorney, whilst in attack step 2 by

claiming to a member of staff from the R&D department. In both attack steps authority was operationalised by means of titles.

Cracker Robert Jorday had been regularly breaking into the computer networks of a global company. The company eventually recognized that someone was hacking into their terminal server, and could connect to any computer system at the company. To safeguard the corporate network, a dial-up password was required on every terminal server.

Robert *called*[a] the Network Operations Center `posing`[b] as an attorney with the **Legal Department**[c] and said he was having trouble connecting to the network. The network administrator explained that there had been some recent security issues, so all dial-up access users would need to obtain the monthly password from their manager. Robert wondered what method was being used to communicate each month's password to the managers and how he could obtain it.

It turned out that the password for the upcoming month was sent in a memo via office, mail to each company manager.

[a]Using the phone modality
[b]Impersonation
[c]Authority

Robert *called*[a] the company after the first of the month, and reached Janet, the secretary of a manager. He said, "*Janet, hi.* ***`This is Randy Goldstein`***[b] *in* ***Research and Development***[c]. *I know I probably got the memo with this month's password for logging into the terminal server from outside the company but I can't find it anywhere. Did you get your memo for this, month?*"

Yes, she said, she did get it.

He asked her if she would fax it to him, and she agreed. He gave her the fax number of the lobby receptionist in a different building on the company campus. He had already made arrangements for faxes to be held for him, and be forwarded to an on-line fax service. When this service receives a fax, the automated system sends it to the subscriber's email address.

The new password arrived at the email dead drop that Robert set up on a free email service in China. Best of all, he never had to show up physically at the location of the fax machine.

[a]Using the phone modality
[b]Impersonation
[c]Authority

Figure 4.4.: Example: 1 Scenario, 2 attack steps, 2 persuasion principles

4.5.2. Output

The sample consists of 74 social engineering scenarios and there are various outputs. The main outcome is a tree structure containing the psychological triggers for all crime scripts (see Figure 4.8).

Furthermore, the prevalence of each persuasion principle, the number of principles used per interaction and the number steps in an attack are discussed.

An analysis of variance (ANOVA) was used to show which persuasion principles are used in the context of social engineering attacks. The results showed that there was a significant difference in the use of principles during social engineering attacks [F(6, 134) = 39.92, p = .000]. The occurrence of the 6 persuasion principles is a) 63% for Authority, b) between 10 and 13% for Liking, Reciprocity and Commitment and c) under 2% for Scarcity and Conformity, refer to Figure 4.5.

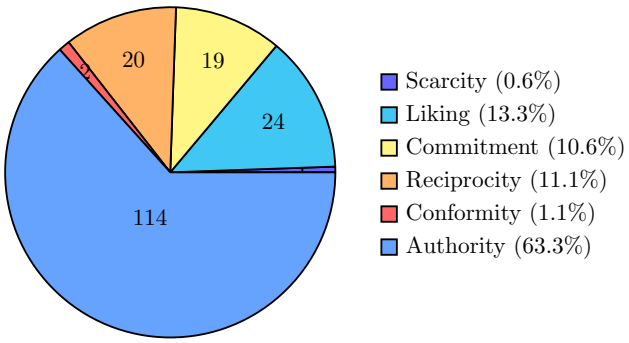

Figure 4.5.: Persuasion principles used

In total 142 attack steps contain psychological mechanisms, of which 125 include persuasion principles and the remaining 17 contain some other mechanism. The category of 'other psychological mechanisms' covers 12% of all psychological triggers used by offenders in their attack steps. This category contains: i) Act ignorant 1x (5.9%), ii) Creating curiosity 2x (11.8%), iii) Distracting 1x (5.9%), iv) Empathy/Pity 2x (11.8%), v) Just ask for it 9x (52.9%) and vi) Overloading 2x (11.8%).

Single attack steps contain up to four different persuasion principles. The average number of persuasion principles used per attack step is M = 1.44 (SD = .723). There was a statistically significant difference in the number of persuasion principles used in social engineering attack attack steps, [χ^2(3,N = 125) = 83.681, p = .000]. Regarding simultaneously used principles, single principles are used considerably more often whilst quadruple principles considerably less often, refer to Figure 4.6.

In total 74 scenarios contain 142 attack steps. The shortest attack path contained a single step, whereas the longest attack path consisted of six attack steps. On average the attack path has M = 1.92 (SD = 1.311) steps. There was a statistically significant difference in the number of attack steps used in social engineering attacks, [χ^2(5, N = 74) = 61.010, p = .000]. Regarding combining attack steps, single step attacks are used considerably more often whilst attacks containing six steps considerably less often, refer to Figure 4.7.

4.6. Summary

In summary, the techniques described here provides a methodology to extract quantitative data from qualitative sources. The example illustrates how this methodology was used to

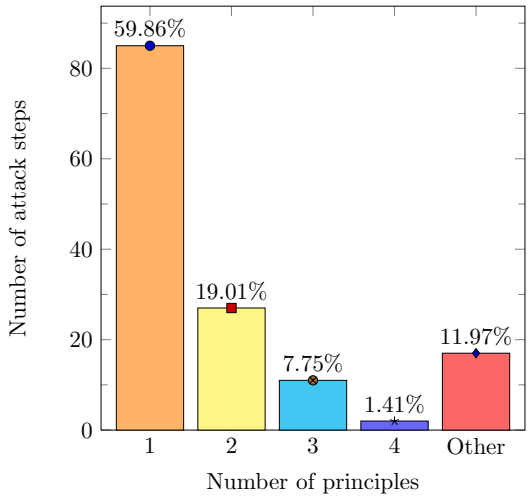

Figure 4.6.: Number of principles used per interaction

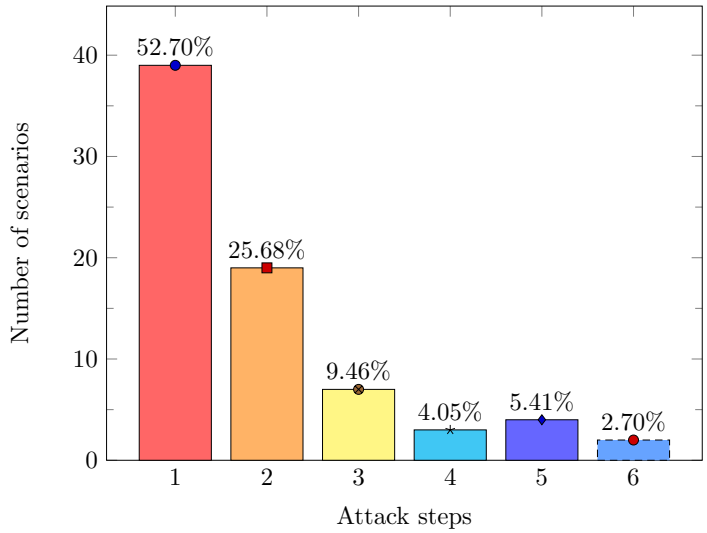

Figure 4.7.: Number of steps in an attack

explore the extent to which persuasion principles are used in social engineering attacks. 74 scenarios were extracted from 4 books on social engineering and were analysed. Each scenario was split into attack steps, containing single interactions between attacker and target, and persuasion principles identified.

The main findings are that:

1. **Persuasion principles** are indeed used in social engineering attacks.
2. **Authority** (one of the six persuasion principles) is used considerably more than others.

3. **Single-principle attack steps** occur more often than multiple-principle ones.

The social engineers identified in the scenarios more often used persuasion principles compared to other psychological mechanisms. The scenario analysis illustrates how to effectively exploit the human element in security. The findings support the view that security mechanisms should include not only technical but also social countermeasures. Besides the current usage of this method, it can also be used to dissect e.g. emails and transcribed interviews. Furthermore this technique is not limited to only the application of the persuasion principles, and other frameworks would work equally as well. A potential future direction could be to automate the narrative analysis using a narrative analysis tool. This would enable greater coverage and the ability to generate statistical output from this approach that could be used as input to the attack pattern library.

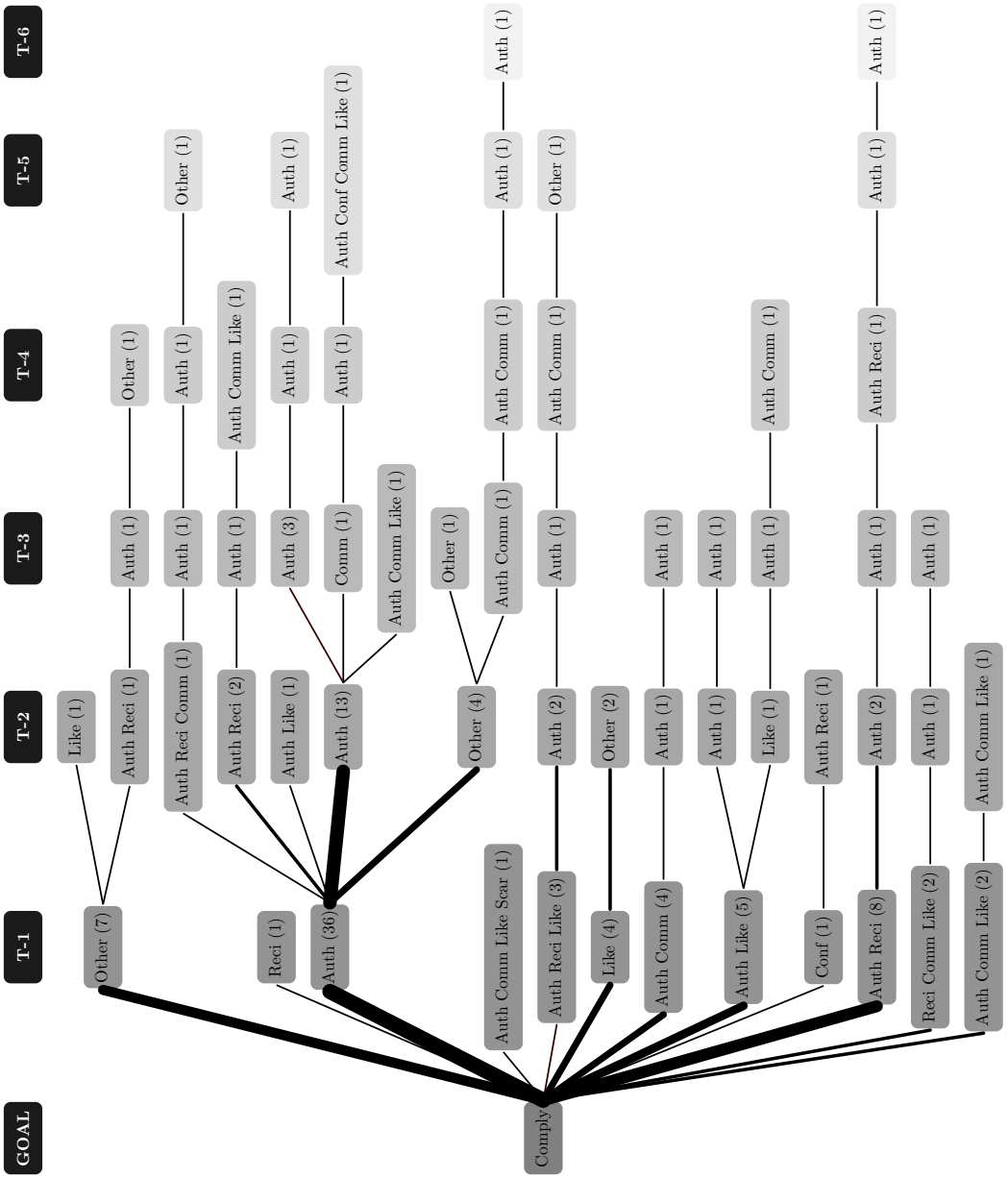

Figure 4.8.: Tree structure of social engineering scenarios. The line width denotes the frequencies, which are specified inside the brackets.

5. Telecommunication Services

This chapter provides details of the type of data that can be gathered in the field of telecommunication service providers. This data can be used to focus analysis and to determine the scope of the TRE$_S$PASS model.

5.1. Motivation

This chapter aims to provide a summary as well as background information on the way data collection is done for TREsPASS in the field of telecommunication service providers (TSP). This description can in itself be considered beyond the state of the art in science as it is not reflected in current research literature. In fact, in this regard neither real data nor any kind of deeper insight into TSPs' working methods and practices may be considered generally available to the research community to date.

5.2. Type of data

Possible sources of data collection include

- Results from early-stage fraud risk assessments of new services or products that a TSP plans to launch,
- Real-time fraud management systems, which are capable of detecting fraudulent activities based on e.g. Call detail records (CDRs), and
- Data from Customer-Relationship-Management (CRM) systems.

Being interconnected to other major telecom networks, a TSP may also gather data from the supervision of boundaries of its respective network (monitoring of traffic from/to other TSPs' networks). In so doing, a TSP will deal with both customers' personal data and their CDRs, and these types of data can both be considered ”social data”.

5.3. Method

In the telecommunications context, knowledge with regard to existing commercial, social and technical data is acquired based on observations and interviews with industry practitioners from a major TSP. For TREsPASS, this domain-specific insight is introduced, aggregated and utilised by project partner GUF.

At TSPs, CDRs are generated in network nodes and forwarded to billing systems while, in theory, customers' personal data can be pulled from CRM systems. However, there are notable restrictions which apply equally to research purposes: Data protection and other company policies of the respective TSP prevent data from being used for any arbitrary purpose, and notably such policies prevent data from crossing the company's boundaries or, in case of multi-national corporations, even the boundaries of a national company's (NatCo) national jurisdiction.

5.4. Envisaged use

Observations and interviews with practitioners from the telecommunications field help identify, structure and process relevant enterprise data, notably commercial and technical data which TRE$_S$PASS needs to get a full picture of in order to develop the capacity to consider and integrate domain-specific requirements within the scope of the project. These observations and interviews are done with a view to particular business activities of a TSP, relevant market characteristics, and other aspects. Aggregated findings then find their way into the respective case study in time. This example is also illustrative more generally of how organisational records might be used by security practitioners to develop the TRE$_S$PASS model over time.

5.5. Example input and output

A CDR is a data record which documents the details of a telephone call or other communications transaction passing through a piece of telecommunications equipment. Attributes contained in a CDR may not only include time, duration, completion status, source number, and destination number, but can in fact be much more detailed. A CDR may e.g. contain attributes such as:

- phone number of the subscriber originating the call (calling party, A-party)
- phone number receiving the call (called party, B-party)
- starting time of the call (date and time)
- call duration
- billing phone number that is charged for the call

- identification of the telephone exchange or equipment writing the record
- a unique sequence number identifying the record
- additional digits on the called number used to route or charge the call
- disposition or the results of the call, indicating, for example, whether or not the call was connected
- the route by which the call entered the exchange
- the route by which the call left the exchange
- call type (voice, SMS, and so on)
- any fault condition encountered

CDRs provide a wealth of information that can help to identify suspects, in that they can reveal details as to an individual's relationships with associates, communication and behaviour patterns, and even location data that can establish the whereabouts of an individual during the entirety of the call.

5.6. Summary

In summary, the techniques described in this chapter enable TSPs to manage multifaceted fraud threats, safeguarding their own interests as well as those of their customers, while protecting the viability of core business models. Understanding the industry's specifics, notably with regard to the processing of relevant data, as well as continuous analysis during the course of the project forms the basis for the aim to make TRESPASS methods and tools useful instruments to address relevant problems in step with actual practice in the industry.

6. Socio-Technical Cyber Threats

This chapter outlines some techniques for gathering and analysing qualitative data from security practitioners. These techniques can be used to help to determine the scope of the TRE$_S$PASS model.

6.1. Motivation

The motivation for developing these techniques is to develop structured methods for extracting the expert knowledge from security practitioners so that this knowledge can be used in designing and refining the TRE$_S$PASS model.

6.2. Type of data

The variables in the analysis were: i) Socio-Technical Cyber Threats is from the past and ii) Socio-Technical Cyber Threats of the future. The variables were open questions and of a qualitative nature.

6.3. Method

The sample consisted of 44 subjects of both sexes who were present at the CSP-forum conference, April 2015, in Brussels.

6.3.1. Procedure

The subjects at the CSP-forum conference were approached by a researcher and were asked if they could help out by filling in a short questionnaire. All subjects were chosen at random and none of the approached subjects refused to help. Each subject was asked some demographic information, what they perceived as Socio-Technical Cyber Threats in the past 15 years and how they think these threats will evolve in the next 5 years. For both past and future, they were asked to provide 3 answers.

6.3.2. Subjects

The subjects attending the CSP-forum conference 'suffer' a self selection bias, where they all i) have an interest in or ii) are related to cyber security. All subjects were employed somewhere in Europe[1] and have between the 0 and 20 years of experience in the field of cyber security. The job titles of the subjects were divided over 10 categories, refer to Figure 6.1. An overview of the industries, where the subjects are active, is shown in Figure 6.2. In total 70% (31 out of 44) of the subject answered to be familiar with Socio-Technical Cyber Threats .

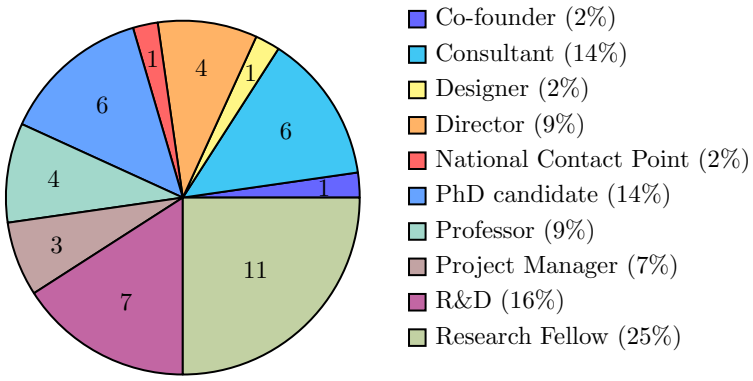

Figure 6.1.: Job title given.

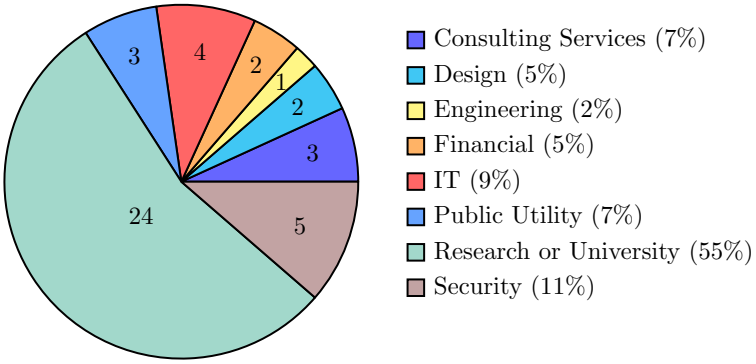

Figure 6.2.: The industry the subjects were employed in.

6.3.3. Analysis

Thematic analysis was used to analyse the answers. Thematic analysis coding system by Boeije (Boeije, 2009) consists of 3 steps:
i) '*Open Coding*' is the process of breaking down an item into fragments, examining each

[1]The subjects were employed in: AT, BE, BG, DE, DK, ES, FR, GR, IR, IT, LU, NL, NO, SE, TR and UK.

fragment, conceptualising each fragment with a code and categorising the data. In this stage, no selection of relevance is made. Using open coding contributes to the organisation and structuring of the data. The output of Open Coding is a list of codes and annotations (Boeije, 2009, pp. 96-108).
ii) '*Axial Coding*' or 'focused coding' re-assembles the codes from Open Coding into different categories. A distinction is made between important and less important categories, this process also reduces and organises the data set, by crossing out or merging synonyms and redundant codes. Axial Coding therefore achieves categorisation, such categories to be described and distinctions made between main and sub-categories (Boeije, 2009, pp. 108-115).
iii) '*Selective Coding*' or 're-assembling' finds connections between dominant categories and makes a model. One possible approach is to define a 'core-concept' which constitutes the heart of the model, and as such this concept is seen to appear frequently in the data, with other categories often linked to it. Selective coding is the final phase of the research and results in a description of the most important concepts, providing a coherent story where the relation between the concepts is described and provides the answer to the initial research question (Boeije, 2009, pp. 115-118).

6.4. Envisaged use

This methodology enables us to gather, group and structure essence of unstructured inputs from questionnaires, brainstorm sessions and other qualitative inputs. On a more practical note, this method is used in D133, emerging Socio-Technical Cyber Threats . The described technique was used to see how the Socio-Technical Cyber Threats , have emerged over the past 15 years and what this tells us about the next 5 years.

6.5. Example input and output

Q: What is perceived as the biggest threats in Socio-Technical Cyber Security in the past 15 years (since the year 2000)?

There were 103 (*data items*) Socio-Technical Cyber Threats reported by the subjects, containing 88 unique entries. There were 3 main themes identified: i) What does an offender use, ii) What makes an attack succeed and iii) The goal of the offender. An overview of all themes is provided in Figure 6.3.

'**What does an offender use**' includes 4 sub-themes: i) *Contact Target*, ii) *Deception*, iii) *Impersonation* and iv) *Increase Compliance*. This theme describes what tricks an offender can use to accomplishment the goal of the attack. The first sub-theme 'Contact Target' describes how an offender contacts the targets. The subjects often mention contact via email messages, either in targeted or mass form, the former is described in the next answer: "*Spear Phishing (Tailored attacks)*". Another method mentioned in the survey is that of the use of the telephone by the offenders: "*Phone caller requests credit card*

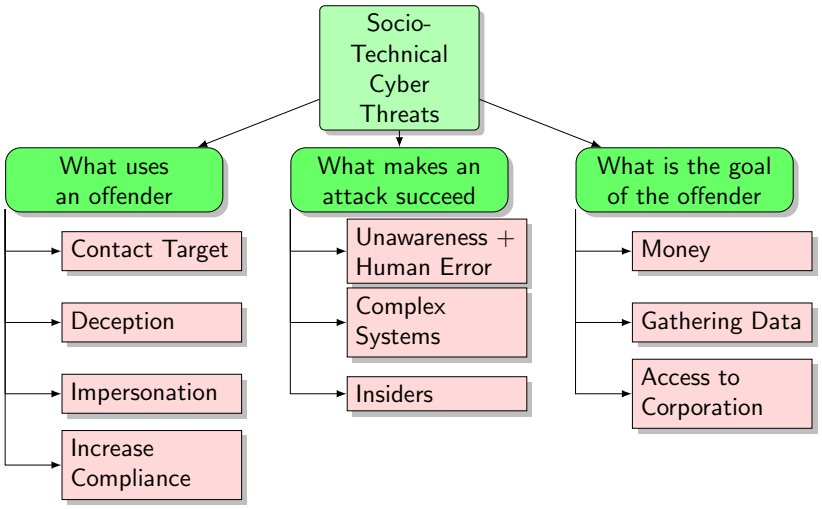

Figure 6.3.: Overview of socio-technical cyber threat themes of the past 15 years (2000 - 2015).

details". Furthermore, the use of dating sites or chat rooms to find and contact targets is also mentioned.

The second sub-theme is called deception, this refers to the offender using lies and mystification to make the victim comply. The use of fake emails is most often mentioned as method for the offender to deceive a target, refer the the next answer "*Fake emails*". Related to this are the unwanted email (or phone) messages that advertise products, illustrated by the following answer: "*False publicity email or phone call*". This is also referred to as spam messages, a special category where the receiver is asked to perform an action, an answer given by the subjects: "*Spam emails asking for help*". Furthermore, fake content on websites or spoofed websites are mentioned as a deception used by the offender to make the target comply, refer to the next answer: "*Social Manipulation by fake URL or online content*". Finally, deception by physical objects is also mentioned, like the Torjan Horse. The analogy is malware hidden inside a USB devices and dropped in a place likely to be taken by someone and put in a PC. An answer referring to this attack is as follows: "*Dropping USB key (with malware)*".

The third sub-theme 'Impersonation' refers to the offender claiming to be someone else to make the target perceive the story as more convincing. Impersonation is mentioned in different degrees. The easiest way to do this is by simply saying that you are someone else; by selecting the right person to impersonate, the strength of an argument can increase. This can both be done over the telephone and in an email as mentioned in this answer: "*False authority email or phone call*". The second approach is to take over the identity of someone, referred to as "*Identity theft*" or "*profile stealing*" Finally, a ghost identity can easily be created on social media and social networks in order to come close to the target. The next answer illustrates this attack: "*Persons creating fake identities to intercept user credentials*".

The fourth and final sub-theme relates to what an offender can use to increase compliance. An offender can use the six persuasion principles to strengthen the argument (Cialdini, 2009): "*Peer pressure*" (also known as conformity) and authority are mentioned as persuasion principles helping to increase the strength of an argument. Furthermore, acting helpless relates to the willingness to help other people, in the survey mentioned as: "*Spam emails asking for help*".

6.6. Summary

In summary, the techniques described here provide a methodology to structure and summarize qualitative input. An example is shown for open-ended questions regarding Socio-Technical Cyber Threats . The main themes and their sub-themes are distilled out of the answers of 44 subjects. Besides the current usage of this technique, it can also be used in interviews, chat-transcripts or message boards. As such the techniques described in this chapter, expert opinions offer additional insights can be obtained in order to build a TRESPASS model and Attack Navigator Map.

7. Security-by-experiment and quantitative penetration testing

This Chapter focuses on getting security data out of different forms of experiments in socio-technical systems, in particular concerning possibilities for security learning after deployment of a technology. Qualitative data, such as possible attack scenarios, provide input for the structure of the navigator maps, such as agents, objects, and their connections. Quantitative data, such as time required for certain attack steps or their success rate, provide input for annotating the navigator maps with values for quantitative analysis. In particular, this chapter proposes quantitative penetration testing as a means for acquiring annotations. The method proposed allows for estimating both the difficulty of the attack steps and the skills of the penetration testers. In our discussion, based on the Dagstuhl seminar we organised (Gollmann, Herley, Koenig, Pieters, & Sasse, 2015), we emphasise the importance of the different types of socio-technical security metrics and their context of use.

7.1. Motivation

The following methods are discussed:

- agenda-setting for availability of security-relevant data from pilots
- quantitative penetration testing
- reflection on socio-technical security metrics

We believe that more security-relevant data should become available from pilots with new technologies. So far, many pilots with emerging technologies such as smart grids do not evaluate security aspects at all (Dechesne, Hadziosmanovic, & Pieters, 2014). To this end, we invested in agenda-setting research on “cyber security as social experiment” or “security-by-experiment”, emphasising the need for learning from pilots, next to development approaches such as the security-by-design paradigm.

A key focus of the project is extracting relevant data from penetration tests, which can be done as part of piloting. As penetration tests inherently have an attacker point of view, they can provide the right kind of data for navigator maps and attack trees. Such tests may include digital, physical as well as social elements. For example, they may include remote hacking, physical trespassing, and social engineering. In particular, quantitative approaches can deliver data such as likelihood of success of attack types, or time needed for specific steps. We conducted practical experiments (Bullée, Montoya, Pieters,

Junger, & Hartel, 2015), but also contributed to methodological innovations on separating attacker properties and system properties in such tests (Arnold, Pieters, & Stoelinga, 2013, 2014).

7.2. Type of data

The data targeted by the techniques are (a) qualitative data describing the actual security architecture being piloted and possible alternatives, and (b) quantitative data about the resistance of such an architecture against executed penetration tests, serving as a proxy for the resistance of such an architecture against actual attacks in a real-life threat environment. This chapter does not cover attacker profile data.

7.3. Method

The details in this chapter outline how experiments can act as data sources for focusing and scoping the model and for providing feedabck as to how to interpret the model's outputs. This section then goes on to show how penetration testing can be used to gather data and provide quantitative input to the attack pattern library.

7.3.1. Data from responsible piloting

The security-by-design paradigm has gained significant popularity. However, this paradigm seems to abstract away from the question how to get meaningful security data about systems once they have been deployed. In a line of research on "cyber security as social experiment", or "security-by-experiment", TRESPASS researchers have aimed at identifying (a) conditions under which social experimentation with deployment of security-sensitive technologies is acceptable (Pieters, Hadžiosmanović, & Dechesne, 2014), and (b) which existing techniques can serve as data sources on running pilots or fully deployed systems (Pieters, Hadžiosmanović, & Dechesne, 2015). The latter paper suggests the following existing techniques and data sources that can assist in developing navigator maps of systems:

- Feedback on beta versions
- Feedback on open source software
- Bug bounty programs
- Red-team/blue-team exercises
- Honeypots
- Penetration testing

In this Chapter, we focus specifically on penetration testing as a source of data. Methods for other data sources could be developed in follow-up activities.

7.3.1.1. Cyber security as social experiment

Lessons from previous experiences are often overlooked when deploying security-sensitive technology in the real world. At the same time, security assessments often suffer from a lack of real-world data. This appears similar to general problems in technology assessment, where knowledge about (side-)effects of a new technology often only appears when it is too late. In this context, the paradigm of new technologies as social experiments was proposed, to achieve more conscious and gradual deployment of new technologies, without losing the ability to steer the developments or make changes in designs (Pieters et al., 2014) . We propose to apply the paradigm of new technologies as social experiments to security-sensitive technologies. This new paradigm achieves (i) inherent attention for the ethics of deploying security-sensitive systems in the real world, and (ii) more systematic extraction of real-world security data and feedback into decision making processes.

7.3.1.2. Security-by-experiment: Lessons from responsible deployment in cyberspace

Conceiving new technologies as social experiments is a means to discuss responsible deployment of technologies that may have unknown and potentially harmful side-effects. Thus far, the uncertain outcomes addressed in the paradigm of new technologies as social experiments have been mostly safety-related, meaning that potential harm is caused by the design plus accidental events in the environment. In some domains, such as cyberspace, adversarial agents (attackers) may be at least as important when it comes to undesirable effects of deployed technologies. In such cases, conditions for responsible experimentation may need to be implemented differently, as attackers behave strategically rather than probabilistically. In this contribution, we outline how adversarial aspects are already taken into account in technology deployment in the field of cyber security, and what the paradigm of new technologies as social experiments can learn from this (Pieters et al., 2015). In particular, we show the importance of adversarial roles in social experiments with new technologies.

7.3.2. Quantitative penetration testing

As an example, consider results from Deloitte's "Phishing-as-a-Service", where fake phishing mails are sent within client companies to test employee responses and provide embedded training. The socio-technical system under consideration is the company ICT, with the specific target being the vulnerability of the employees to phishing attempts. We are currently analysing the results of Phishing-as-a-Service exercises in relation to properties of the phishing e-mails. Another example is the key experiment discussed below.

Next to data on individual attack steps, data can also be gathered from multi-step penetration testing, where the penetration testers determine their attack vectors themselves. To this end, we developed methods for quantitative penetration testing using item response theory.

7.3.2.1. Quantitative penetration testing with item response theory

Existing penetration testing approaches assess the vulnerability of a system by determining whether certain attack paths are possible in practice. Thus, penetration testing has so far been used as a qualitative research method. To enable quantitative approaches to security risk management, including decision support based on the cost-effectiveness of countermeasures, one needs quantitative measures of the feasibility of an attack. Also, when physical or social attack steps are involved, the binary view on whether a vulnerability is present or not is insufficient, and one needs some viability metric. When penetration tests are performed anyway, it is very easy for the testers to keep track of, for example, the time they spend on each attack step. Therefore, we have proposed the concept of quantitative penetration testing to determine the difficulty rather than the possibility of attacks based on such measurements (Arnold et al., 2013, 2014). We do this by step-wise updates of expected time and probability of success for all steps in an attack scenario. In addition, we show how the skill of the testers can be included to improve the accuracy of the metrics, based on the framework of Item Response Theory (Elo ratings). We prove the feasibility of the approach by means of simulations, and discuss application possibilities.

7.3.2.2. The persuasion and security awareness experiment: reducing the success of social engineering attacks

Objectives: The aim of current work is to explore to what extent an intervention reduces the effects of social engineering (e.g. the obtaining of access by persuasion) in an office environment (Bullée et al., 2015). In particular, we study the effect of authority during a 'social engineering' attack.

Methods: 31 different 'offenders' visited the offices of 118 employees and on the basis of a script, asked them to hand over their office keys. Authority, one of the six principles of persuasion, was used by half of the offenders to persuade a target to comply with his/her request. Prior to the visit, an intervention was randomly administered to half of the targets to increase their resilience against attempts by others to obtain their credentials.

Results: 37.0% of the employees who were exposed to the intervention surrendered their keys whilst 62.5% of those who were not exposed to it handed it over. The intervention has a significant effect on compliance but the same was not the case for authority.

Conclusions: Awareness-raising about the dangers, characteristics and countermeasures associated with social engineering proved to have a significant positive effect on neutralising the attacker.

7.3.3. Reflection on socio-technical security metrics

In the report of our Dagstuhl seminar (Gollmann et al., 2015), we have distinguished between type 1 and type 2 security metrics, related to adversarial roles in experiments. We have followed up on this idea by distinguishing between counterfactual and non-counterfactual approaches to security argumentation and security metrics (Herley & Pieters, 2015). In this framework, penetration tests are counterfactual approaches (type 1; because the attackers are controlled rather than real threat environments), whereas data on actual incidents are non-counterfactual (type 2; as they include the influence of the real threat environment). This distinction is important: if one wants to measure the difficulty of an attack step, the appropriate metric is type 1. If one wants to measure the expected loss, the appropriate metric is type 2. This type 2 metric can be either historical data, or analytic results from combining type 1 metrics with adversary profiles in order to predict adversary behaviour.

7.4. Envisaged use

The methods presented in this section can be used to guide pilots or experiments with new technologies, in order to increase the availability of security-relevant information from the pilots. This can be qualitative information related to security architectures and possible changes thereof, but also quantitative information concerning security measurement in the pilots.

We have suggested quantitative forms of penetration testing as a specific method that can be used in the pilot phase of new technologies. Within this domain, we have suggested a specific innovation in which the skill of the penetration testers is factored in in the measurement obtained from the tests. This particular innovation has not been tested in practice yet.

7.5. Example input and output

In this section we describe specific data inputs for the proposed techniques, presenting a representative sample piece of input data, along with an example of a typical output that results from the application of the technique.

Based on responsible deployment, targets should be set for what we wish to learn about the security of a new technology in the pilot stage. Pilot design should describe which security features should be evaluated in the pilots and how. Based on this input, data can be gathered. Of this data, identified vulnerabilities and possible attack scenarios can be used as input for the Attack Navigator Maps. If the map does not generate certain proposed scenarios, or does not represent exploiting specific vulnerabilities as attack steps, the map can be changed accordingly. Data from red-team exercises and penetration tests should be represented in a quantitative form, not only reporting success, but also time

taken and properties of the participants/testers, such as skill levels. This can be done for new technologies, but also for existing organisations. Based on this data, difficulty levels for specific attack steps and scenarios can be calculated and used as annotations for the maps.

For example, assume that penetration testers or ethical hackers report a possible attack scenario on a new cloud service in which they social engineer an administrator, taking them 10 minutes by phone. If the administrator did not occur on the map yet, s/he should obviously be added. Additionally, a general social engineering attack pattern should be instantiated for this particular situation and annotated with the measured time, or, if it already exists, it should be updated based on the measured time. Additional follow-up experiments may be conducted to evaluate the likelihood of success and average time-to-success more precisely.

7.6. Summary

In summary, the techniques described here provide (a) a paradigm for responsible piloting with new security-critical technologies, providing more data on possible attacks and their properties, (b) methods for quantitative penetration testing for individual attack steps as well as complete systems, yielding difficulty (control strength) values for annotation of navigator maps, and (c) a critical reflection on different types of security metrics and their use. As such the techniques described in this chapter, offer additional means by which data gathering can be planned and executed in order to build a TRE_SPASS model and Attack Navigator Map.

8. Cues and warnings against phishing may not be effective

8.1. Introduction

The present study investigates whether users can be protected from attempts to illictly gather personal information. It tests the effectiveness of two interventions that aim to protect users against social engineering attacks, namely 1). Cues to raise awareness about the dangers of online activities and 2). Warnings against disclosing personal information.

In this deliverable we continue within the framework that was developed in D2.3.1 and present new research findings that shed a new light on previous findings. Specifically, we present results from social science that can serve as input for the socio-technical system models (e.g. social engineering experiments) and explanations of behaviour that help account for the effect of countermeasures on the vulnerability for social engineering attacks.

Risk management can be seen as a decision theory problem, as it aims at optimising investment decisions (see D2.3.1 Section 7.1.4). However, the adversarial aspects in security point to the possibility that threat agents adapt to decisions made, thereby invalidating decision theoretic assumptions. It is often assumed that putting a countermeasure in place may simply cause the threat agents to attempt different scenarios. In this line of reasoning understanding what countermeasures can do is essential.

Knowledge on the effectiveness of counter measures to prevent cyber-security breaches illustrates the need to make models flexible. In the present deliverable we showed that in some contexts countermeasures can be effective, but in others they do not. In the key-experiment (D2.3.1) we showed that a prevention campaign can reduce significantly the vulnerability of humans in an office environment. Specifically, information and a reminder, namely a key-fob reduced handing over a key to an attackers from 62.5% to 37%.

In the ‘cues and warnings experiment’ (present deliverable) we show that similar types of counter measures had little to no effect: a) cues to cybercrime did not work to reduce members of the public to reveal an email addresses, half of the digits of their bank account number of information on their online shopping; b) A warning in the form of a leaflet decreased the number of emails addresses revealed to an ‘attacker’ but did not affect the reporting of bank account information or online shopping. It is suggested that changes in context - from a relatively quiet professional environment (key-experiment) to square in a shopping Centre (cues and warning experiment) - may explain the difference in the

effectiveness of similar countermeasures. These findings suggest that models may need to include contextual information in order to be able to predict how countermeasures affect human behaviour in cyber-security.

8.1.1. Cyber-attacks are common

Many cyber-attacks start with users who unknowingly disclose personal information to attackers. This is especially true for spear phishing, or, targeted attacks, which are an increasingly popular form of attacks (Hong, 2012; Wueest, 2014). Personalising seems to increase the success of phishing. Targeted attacks are relatively successful (Jagatic, Johnson, Jakobsson, & Menczer, 2007). An experiment using social network information, showed that people 4.5 times more likely fall for phish sent from an existing contact over standard phishing attacks. Of 512 students at the Corps of Cadets at the West Point receiving a spear phishing email, mentioning a problem with their Grade Report, 80% clicked on the link in the email (Ferguson, 2005).

8.1.2. Anatomy of an attack

In a targeted attack, the first step in the process is the reconnaissance phase, were the aim is to learn as much as possible about a targeted person (Wueest, 2014). Attackers can attempt to find information through public sources – such as social media (Hong, 2012), they can steal it (Bursztein et al., 2014) or they can ask it directly to their potential victims, for instance, by sending phishing emails or by calling them, as Mitnick did (K. D. Mitnick & Simon, 2002, 2005). Other studies also reported the occurrence of 'voice phishing, (Maggi, Sisto, & Zanero, 2011; Salem, Hossain, & Kamala, 2010). Many studies showed that users are very vulnerable to phishing attacks (Abraham & Chengalur-Smith, 2010; Dodge & Ferguson, 2006; Hong, 2012; Jansson & von Solms, 2011; Wright, Jensen, Thatcher, Dinger, & Marett, 2014).

8.1.3. Origins of success of phishing

Many security offices have lamented on the vulnerability of users to phishing attacks and their tendency to disclose information (Adams & Sasse, 1999; Kirlappos & Sasse, 2012). These security managers tend to forget about human nature. Part of the success of social engineering may not have to do with clever offenders but characteristics of humans. Below we state that humans tend to trust and they tend to disclose information to other at relatively high rates.

Trust determines the way that an individual approaches other people. Trust usually is defined as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or the behaviour of another" (Rousseau, Sitkin, Burt, & Camerer, 1998). Many studies show that humans tend to trust each other

(Fetchenhauer & Dunning, 2009; Glanville & Paxton, 2007). Trust has evolutionary advantages for humans (Ostrom, 1998; Penner, Dovidio, Piliavin, & Schroeder, 2005). Trust has been studied for individuals, and exists at the organisational and country level as well.

In general, having trust in others has positive outcomes for individuals (Dohmen, Falk, Huffman, & Sunde, 2012; Fetchenhauer & Dunning, 2009; Frattaroli, 2006; Glanville & Paxton, 2007; Ostrom, 1998). Trust is also crucial for the 'health, harmony, and growth of any organization' (Fetchenhauer & Dunning, 2009; Tyler, 2003). Trust is essential for maintaining electronic commerce (Ratnasingham, 1997). The level of trust has also been shown to differ across countries and is related to differences in outcomes such as economic growth, volume of trade, and institutions, (Cozby, 1973; Dohmen et al., 2012, p. 646). Trust is crucial for maintaining a fully-functioning democracy, (Fetchenhauer & Dunning, 2009, p. 263) and economic development (Zak & Knack, 2001).

Besides having relatively high trust, humans seem to have low thresholds to disclose personal information and disclose personal information relatively often. Disclosure or self-disclosure can be defined as the process of communicating information about oneself verbally to another person (Cozby, 1973). Most of the studies on disclosure have been done in the field of psychology and mental health. These studies showed that self-disclosure has positive outcomes (Dindia, Allen, Preiss, Gayle, & Burrell, 2002). Persons tend to disclose to people they like, and they also tend to like people who disclose to them (Omarzu, 2000; Sprecher, Treger, & Wondra, 2013). In personal relationships disclosure is often reciprocated (Cozby, 1973; Omarzu, 2000), and it has been described as 'an exchange process', and as 'rewarding' (Cozby, 1973; Worthy, Gary, & Kahn, 1969). Many studies also show that the level of disclosure depends on both the context and on individual differences (Dindia et al., 2002). Disclosing personal information in part the result of trust. Trust and self-disclosure have a relatively strong positive relationship (Wheeless, 1978; Wheeless & Grotz, 1977).

Much research has also been performed from the point of view of the attacker (Abraham & Chengalur-Smith, 2010; Hong, 2012; Jagatic et al., 2007). A number of studies also looked at disclosure by potential victims. John, Acquisti, and Loewenstein (2011) presented experiments on disclosure. Disclosure was measured when respondents answered questions on deviant behaviour such as 'Having sex with the current husband, wife, or partner of a friend', or 'Making a false insurance claim'. Three experiments showed that users disclosed more personal information on unprofessional looking websites - which are arguably more likely to misuse it than on professional looking websites which were less likely to misuse it.

In other words, individuals are prone to disclose in contexts that downplay privacy concern, ironically, even when such contexts are likely higher in both objective and perceived disclosure danger (John et al., 2011, p. 868). Interestingly, in a fourth experiment, when users are cued to think about privacy, these differences in the context, type of website- disappear and all users have similar rates of disclosure. John et al. (2011, p. 868) conclude that their results 'stand in contrast to the considerable body of privacy research that is premised on the assumption of rational choice', which states that people make trade-offs between privacy and other concerns, implying that disclosure is the result of this rational

weighing of costs and benefits, in which objective costs, such as an unprofessional looking website should prevent or at least decrease disclosure.

In a later study, Acquisti, John, and Loewenstein (2012) looked at the effectiveness of raising privacy concerns on the disclosure of information. Cues to think about phishing consisted a number of pages with picture of phishing email and the request to categorise them as phishing or not-phishing. In this study, cues about phishing led to a decrease in disclosure (Acquisti et al., 2012).

In another study, Joinson, Reips, Buchanan, and Schofield (2010) also presented experimental data. They reported that self-disclosure was reduced when a weak privacy policy was combined with context of low trust. In all other combinations of trust (high versus low) and privacy (High versus low), self-disclose was higher. It is notable that their respondents had relatively high rates of disclosure. This study used similar questions on deviance as (John et al., 2011).

8.1.4. What can be done about it?

Because users are easily tricked into giving personal information, security managers and researchers have been looking for preventive measures (Abraham & Chengalur-Smith, 2010). Below education and warnings are discussed.

8.1.5. Education

Education has been evaluated in a number of experimental studies. Some experiments evaluated training users and showed that this helped them to fall less often for phishing mails. For instance, the “School of phish” (Kumaraguru et al., 2009) compared to what extent three groups of about 170 participants each fall for phishing scams. The control group received no training, one group was trained once and the last group received training twice. The results suggest that training reduces the likelihood of participants falling for phishing scams. However, even after training the number of participants who fall for phishing scams remains of the order of 20%. Similar experiments also produced positive findings (Kumaraguru, Sheng, Acquisti, Cranor, & Hong, n.d.; Sheng et al., 2007). Other studies also mentioned positive effects of education/training via email or online (Alnajim & Munro, n.d.; Dodge, Coronges, & Rovira, 2012; Greis, Nogueira, & Kellogg, 2012; Huang, Shen, Doshi, Thomas, & Duong, 2015; Jansson & von Solms, 2011). However, other studies reported no positive effect from training (Caputo, Pfleeger, Freeman, & Johnson, 2014; Davinson & Sillence, 2010).

8.1.6. Warnings

Warnings have also been tried, generally to warn for a lack of website safety. Kirlappos and Sasse (2012) used a warning to inform about website safety. Warnings helped improve user behaviour, but again, a relatively large part of the users did not adjust his/her

behaviour when monetary rewards were at stake. A study by Zhang (2014) found an adverse effect of warnings. This study investigated user's behaviour in the presence of warnings for mobile sites. The security cue was operationalised by the presence or absence of a security warning banner showing that a trusted security certificate could not be detected. In contrast to expectations, participants disclosed more social media information, i.e., number of Facebook friends, Twitter ID, number of Twitter followers, number of people followed on Twitter, on the stimulus website when the security cue was present.

In conclusion, there is not a large experimental literature on the prevention of social engineering attacks, and their results have been ambivalent. Still, there is a shortage of research showing the practical effectiveness of information security preventive intervention their impact on end-user's behaviour (H. J. Smith, Dinev, & Xu, 2011). The present study contributes to the field and tests two interventions that aim to prevent disclosure of personal information. The main question is whether an awareness intervention and a warning prevent users to disclose personal information in a social engineering attempt?

8.1.6.1. Increasing awareness.

The awareness intervention consists of 4 questions meant to raise awareness of privacy issues and cybercrime. This corresponds to the 'salience principle' that states that our attention is drawn to what is novel and seems relevant to us, and this is an important determinant of behaviour (Dolan, Hallsworth, Halpern, King, & Vlaev, 2010). The four questions can also be conceptualised as 'priming'. Priming activates knowledge of certain goals and makes them ready for use (Kenrick, Neuberg, & Cialdini, 2005). Research shows that people's behaviour can be altered when they are exposed to certain sights, words or sensations (Dolan et al., 2010; Kenrick et al., 2005). Priming often works outside conscious awareness (Dolan et al., 2010; Kenrick et al., 2005).

8.1.6.2. Delivering a warning.

Warnings are a much more direct way of communicating a message. Warnings can be successful in influencing behaviour (Argo & Main, 2004; Wogalter, Laughery, & Mayhorn, 2012). There are guidelines for effective warnings, which were summarised by (Wogalter et al., 2012). Principles that were relevant for the present intervention were: 1). Brevity, warnings should be as brief as possible, 2). Design for the low-end receiver, meaning that warning should not be directed at an 'average person' but for people who have lower competence, education, knowledge, and/or the elderly or the disabled, for instance.

8.2. Method

8.2.1. Sample

The data were collected by approaching people at several locations in the commercial area of the city centre of Enschede (the Netherlands). Enschede is a city at the East of the Netherlands of 158,586 inhabitants (Netherlands, 2015). One of the researchers approached the members of the public in the following way: 'Can I ask you something?' Is the person agreed, he would continue: 'I am a student of the University of Twente, I work on a study on Phishing and Cyber security. Do you have time to complete this survey, it only takes a few minutes.' The research was carried out for five days between May 18 and May 26, between 10AM and 5 PM. Attention was paid to spread the different conditions over different days and times of the day. We aimed to interview 100 respondents for each of the three experimental conditions.

8.3. Measures

8.3.1. Experimental condition

The questionnaire was based on Beunder, Kerkers, and Orij (2014). It consists of two parts: 1) a common part that all respondents answered, and 2) four questions on cybercrime and privacy on the internet. There were three conditions: two experimental conditions and a control condition.

1. The awareness condition. In this condition respondents had to answer 4 questions about cybercrime, namely: 1). 'Are you familiar with the term phishing?' 2). 'Are you aware of the amount of personal information you share on the Internet and that is publicly accessible?' 3). 'Do you use Facebook? If so, what are generally your privacy settings?' 4). 'Have you ever been scammed on the Internet (for example through phishing)?'. These questions were placed in the middle of the questionnaire.
2. The warning condition. In this condition, prior to getting the questionnaire, respondents were handed over a leaflet of A4 format (Figure 8.1) before they received the questionnaire. In addition, a small part of this leaflet was placed at the top of each page, as a reminder (Figure 2). This leaflet was inspired by Bullée, Montoya, Pieters, Junger, and Hartel (2015) who created a poster for an intervention in a study aiming to reduce the success of social engineering attacks (Bullée et al., 2015). The leaflet attempted to be brief, focussed on the exact right issues and attempted to be as simple and direct as possible, following suggestions on successful warnings (Wogalter et al., 2012).
3. A control condition, in which respondents answered only the questions of the common part of the questionnaire and did not receive the leaflet.

Figure 8.1.: Warning

8.3.2. Measures of disclosure

Disclosure is measured with four questions and a sum score of the positive scores ('total risk'):

1) **Email address** A request to write down one's e-mail address was asked at the beginning of the questionnaire. Respondents were first asked if they wanted to receive a copy of the results of the study, and then if they could fill out their email address. As this question was presented at the beginning of the questionnaire, the email address differentiates the warning condition and the combined the awareness and control condition, as the awareness raising questions were asked later in the questionnaire.

Figure 8.2.: Small Warning Message

2) **Bank account information** Only a part of the bank account number was asked, to protect respondent's privacy: respondents were asked to fill out 9 digits from their 18 digit bank account number Respondents needed to show if they knew their bank account number by heart. In many European countries, the bank account number consists of letters for a specific country, a control number and an abbreviation of the banks' name (see figure 3).

Figure 8.3.: Bank account number, the respondent was asked to fill in the squares

3) **Product** If respondents had bought something online, they were asked which product they had bought. A broad range of options were offered and an option was added to fill out something else.

4) **Online shop** Besides what the respondents bought online, there was a question about the name of the web shop where they bought their products. The answer categories were some major Dutch online shops and the category additional.

For these four measures of disclosure, answer categories were: filled in (1) or not filled in (0). No specific information was coded or used, to preserve respondents' privacy. 5) Total risk. A total risk variable is the sum of the positive answers on the previous questions, namely the number or times respondents did disclose personal information on each of the four measures. Total risk is a continuous variable with scores from 0 to 4. In figure 4, we differentiated between the highest score (4) and the lower scores combined.

These questions, with the exception of the email address (see above), were all placed at the end of the questionnaire.

8.3.3. Control variables

Information on age, sex and educational level was recorded. Besides socio-demographic variable, two questions were asked about computer knowledge (see Table 8.1) and the

amount of time respondents spend online (in term of days a week and hours a day). As a non-linear relationship was found between age and the measures of disclosure, an age-square variable was also constructed and used in the multivariate analysis.

8.3.4. Analysis

Description of the data was done using frequencies and cross tables. To analyse the effectiveness of the intervention, logistic regression and regression analysis were performed. Because the three research groups differed on age, these analyses control for age and, as mentioned above, age square.

8.4. Results

8.4.1. The sample

A total of 290 persons filled in the questionnaire. Two respondents were not online at least once a week, and accordingly, these respondents were deleted from the analysis. Due to 10 missing values (6 on age and 4 on education), 278 respondents remained in the analysis. 256 respondents shopped online. The analyses for disclosure in relation to online shopping were executed on this subgroup only.

58% of the respondents were females, which is more than the general population. The sample is slightly younger than the general population. In Enschede, 22.2% of the population is younger than 20, in the present study this is 26.3%; and 16.2% of the Enschede population is older than 65, in the present study this is 5.4%. This means that the sample, without being representative of the population, does represent a reasonable representation of it, younger than 65 segment.

The control group consisted of 98 respondents, the awareness group of 97 respondents and the warning group of 95 respondents. Table 8.1 shows the characteristics of the respondents on the control variables. The three groups contain equal number of males and females and they do not differ with respect to educational level, computer knowledge, and on time spend online. However they do differ in terms of age. The mean age is 30 years old, however, the warning group is somewhat older (mean age=34.2) and the awareness group is somewhat younger (mean age=26.3). Accordingly, to assess the effectiveness of our interventions, controlling for age is necessary.

Additional analysis also showed that age is also related to the measures of disclosure as seen in Figure 8.4, which shows that overall, older respondents less often disclosed personal and identifiable information. This is most obvious for reporting an email address. It is notable that there is a non-linear trend for disclosure by age. The disclosure of information increases by age, to decrease again for the eldest age-group. The 20-29 year olds and the 30-39 year olds have the highest disclosure levels.

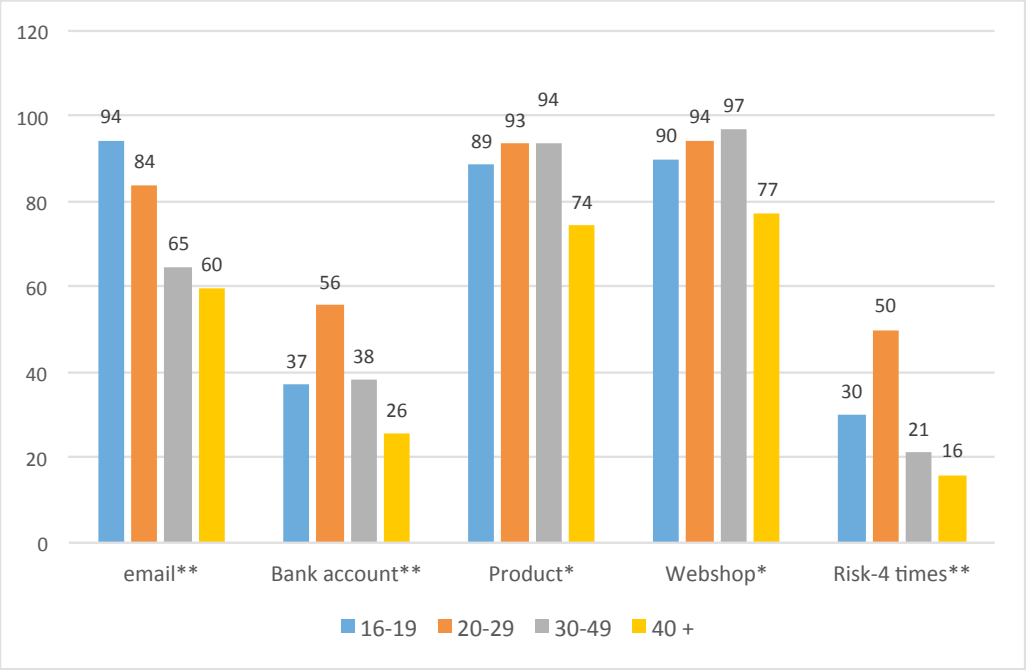

Figure 8.4.: outcome by age, in %, and % of the highest risk score, selection: all for email and bank account and online shoppers for product, online shop and risk score.

8.4.2. Effectiveness of the interventions

Table 8.2 and 8.3 show the bivariate relationship between the experimental condition and the measures of disclosure. Only one of the five outcomes is significantly related to experimental condition. In the warning condition, a smaller number of respondents report their email address, namely 67%. In contrast, this was 88% in the awareness-group and 81% in the control group ($p<.01$). 43.5% of all respondents filled in information on their bank-account, there were no differences between the three groups. Among the online shoppers, 89.8% of the respondents filled in what kind of product(s) they purchased and 91.4% filled in the name of the online shop where they did these purchases. Again, no differences were found between the three groups.

Multivariate analyses were executed to study the effectiveness of the interventions, while controlling for differences in age and age square (Table 8.4 and 8.5). The findings show that a warning decreases disclosure of one's email address, although this result is on the verge of statistical significance ($p=.08$). The interventions do not affect the number of respondents that reporting their bank account, the type of product one bought online, and the total risk score.

Reporting the online shop is related to the experimental condition but in an unexpected way. The logistic regression shows that age, age square and the presence of a warning are related to disclosure. To describe the findings in more detail a cross tabular analysis was created of disclosure of online shop, by age and experimental group (Figure 8.5). In the natural situation, for the control group, there is a non-linear relationship between

Table 8.1.: Characteristics of the respondents, means and % (N=278)

All	Control	Awareness	Warning	%. Mean
Sex. % females ns	56.3%	57.4%	61.4%	58.3%
Mean age (years) p=.003	30.2	26.3	34.2	30.1
Mean number of hours online ns	4.6	5.3	5.5	5.1
Mean number of days a week online ns	6.5	6.6	6.4	6.5
Education. high. in % ns	18.8%	22.3%	28.7%	23.1%
How well you can manage a computer?				
[Well. to very well.] ns	54.2%	53.2%	55.7%	54.3%
Total	96	94	88 (1)	278
Online shoppers only				
Sex. % females ns	56.3%	57.6%	61.0%	58.2%
Mean age (years) p=.05	27.1	26.5	31.3	28.2
Mean number of hours online ns	4.8	5.2	5.4	5.1
Mean number of days a week online ns	6.6	6.6	6.5	6.6
Education. high. in % ns	19.5%	22.8%	32.9%	24.7%
How well you can manage a computer?				
[Well. to very well.] ns	57.5%	54.3%	63.6%	58.2%
Total	87	92	77 *(2)	256

* p<.05, ** p<.01
(1) N=87 for education
(2) N=76 for education

age and disclosure: The youngest (the 19 and younger) and the eldest respondents (the 40 and older) tend not to disclose as much information and the two other age groups. However, in the warning condition, these age differences disappear and all respondents have high disclosure rates. In the awareness condition, only the 40 an older have low disclosure rates, however, this is not a statistical significant result. In other words, the experimental intervention seem to equalise groups and lead to higher disclosure in each age group.

Table 8.2.: Respondents providing personal identifiable information, in %

All	Control	Awareness	Warning	Total
Email filled in **	81.3	88.3	67.0	79.1
Bank account number ?ns?	49.0	40.4	40.9	43.5
N	96	94	88	278
Online shoppers only				
What kind of product you purchased? ns	92.0	84.8	93.5	89.8
Filled in the name of the web shop ns	87.4	89.1	98.7	91.4
N	87	92	77	256

* p<.05, ** p<.01

Table 8.3.: Number of times respondents provided personal identifiable information, selection: online shoppers, in %

	Control	Awareness	Warning	Total
0	2.3	.0	.0	.8
1	3.4	3.3	2.6	3.1
2	12.6	21.7	24.7	19.5
3	37.9	42.4	39.0	39.8
4	43.7	32.6	33.8	36.7
N	87	92	77	256

Pearson Chi-Square=9.4, df=8, p=.31

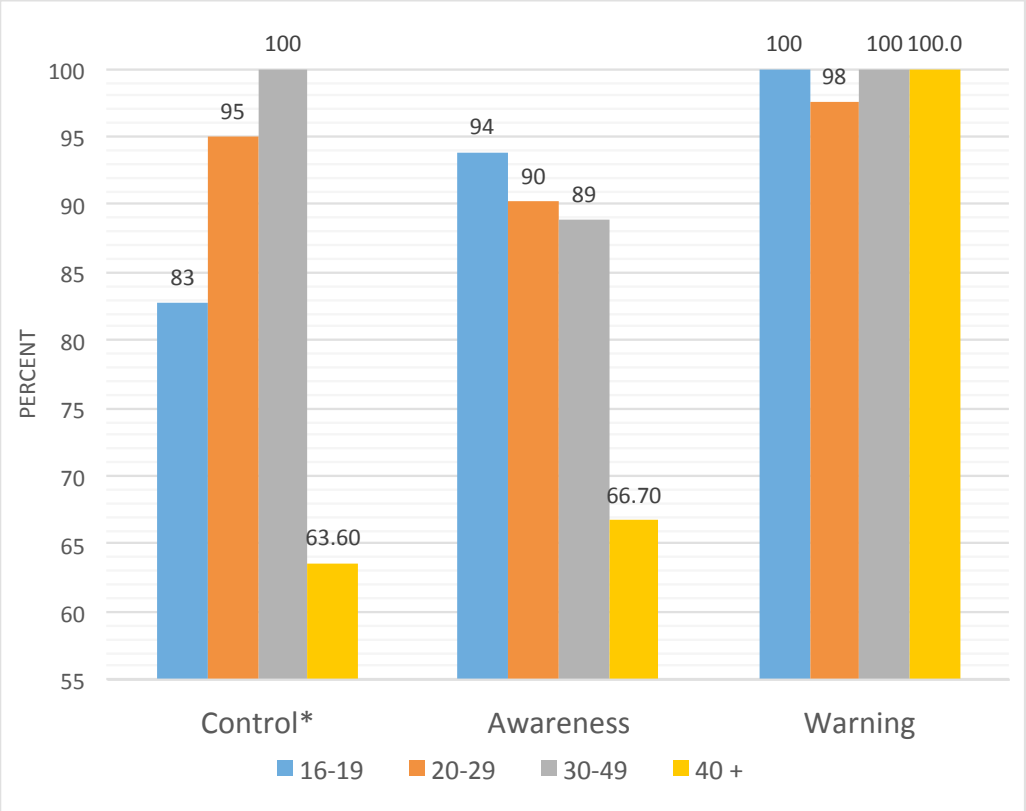

Figure 8.5.: reporting the online web-shop by experimental condition and age, in %, online shoppers only

8.5. Discussion

Humans tend to trust each other and they tend to disclose personal information relatively easily. In general this has beneficial consequences for themselves and society at large. However, it also makes them vulnerable to social engineering attacks. The present study

Table 8.4.: Effect of a warning or priming on disclosure of reporting an email address and a bank account number, logistic regression analysis. All respondents (N=278), selection online shoppers (N=256)

All		**email**				**Bank account number**		
			95% CI				95% CI	
	OR	p	Lower	Upper	OR	p	CI-low	CI-up
Age	0.93	ns	0.83	1.04	1.07	ns	0.96	1.20
Age square	1.00	ns	1.00	1.00	1.00	ns	1.00	1.00
Priming	1.50	ns	0.65	3.47	0.67	ns	0.37	1.20
Warning	0.53	a	0.26	1.07	0.73	ns	0.40	1.33
Constant	26.60	**	0.43					
Online shoppers		**Product**				**Online shop**		
Age	0.82	*	0.68	0.99	1.23	*	1.00	1.51
Age square	1.00	*	1.00	1.01	1.00	*	0.99	1.00
Priming	2.01	ns	0.74	5.44	1.31	ns	0.51	3.40
Warning	0.70	ns	0.20	2.52	16.06	*	1.72	150.17
Constant	1.24	ns			0.45			

a p=.08
* p<.05; ** p<.01

Table 8.5.: Effect of a warning or priming on number of items disclosed, multiple regression analysis, selection: online shoppers (N=256)

	SE	β	p
Age	.03	.54	ns
Age square	.00	-.80	*
Priming	.13	-.07	ns
Warning	.13	-.05	ns
Constant	.41		**

R=.29, R^2=.08
* p<.05, * p<.01

investigated whether an awareness intervention and a warning prevented users to disclose personal information in a social engineering attempt.

To this end 278 respondents filled in a questionnaire in a shopping area in the city centre of Enschede, (the Netherlands). Three conditions were created: 1). an awareness condition, in which respondents had to answer four questions on cybercrime and privacy issues, 2). a warning condition, before handing over the questionnaire, the respondents received a leaflet with a warning not to give away personal information, 3). a control condition, in which respondents filled in the questionnaire, without the awareness questions. Disclosure was measured by asking the respondents, for their email address and 9 digits from their 18 digit bank account number, and, if they shopped online, which product they had purchased and in which web shop. It was coded whether they provided this information. A risk score was the sum of the number of times personal information was reported.

Relatively high disclosure rates were found: 79.1% of the respondent filled in their email address, and 43.5% provided information on their bank-account. Among the online shoppers, 89.8% of the respondents filled in what kind of product(s) they purchased and 91.4% filled in the name of the online shop where they did these purchases.

The level of disclosure is rather high. There are not many studies to compare our results with. However, John et al. (2011) reported in their experiment that 44% of their respondents reported an email address. This is also relatively high but not as high as the present findings. A possible explanation is that there are differences in trust by country. John et al. (2011), study was done in the US and the present study in the Netherlands. This explanation is not supported by cross-cultural research which showed that levels of are broadly similar in the US and the Netherlands (Delhey, Newton, & Welzel, 2011).

The multivariate analysis showed that neither the awareness questions, nor the warning influenced the degree of disclosure. An adverse effect for both interventions was found on disclosing information on the online web shop where they purchased their product(s). In the control condition, elder respondents (40 and older) were more cautious and less often reported the web shop name than younger respondents. However, in the warning condition, age differences in disclosure disappeared. In the warning condition, all age groups had high disclosure rates. The findings of the awareness appear similar to the control group but were statistically non-significant.

The lack of effectiveness of the present interventions does not match with the broader research that noted that priming and nudging people are generally effective ways to influence behaviour (Dolan et al., 2010; Goldstein, Martin, & Cialdini, 2008; Thaler & Sunstein, 2008).

It is important to note that the literature on effectiveness has yielded contrasting results. Several unsuccessful interventions have been described. For instance, a recent review concluded that interventions to raise awareness are generally not very successful (Bada & Sasse, 2014). A number of experimental studies also showed that, weaker, interventions were not effective while stronger interventions were. (Jenkins, Durcikova, & Reeves, 2013) reported that a training to prevent the disclosure sensitive information was not enough to improve secure behaviour, but the combination of a training with a reminder was successful in achieving this goal. Similar findings were presented by (Dodge et al., 2012). In the field of accidents, (Stave, Törner, & Eklöf, 2007) found that just talking about safety during a series of workshops did not impact safety behaviour, but a more structured approach that also discussed incident reports did.

An adverse effect was found for the warning condition on disclosing the name of the online shop. This is a counter intuitive finding. A similar counter intuitive finding was reported by (John et al., 2011) who found that their respondents disclosed personal information more often on unprofessional looking website, which were also evaluated as less secure, while disclosing less information on a professional looking website - as was mentioned above. It has been noted in other fields that some interventions can have adverse effects. For instance, in a review of the literature on non-response in surveys showed that, for sensitive subjects, such as sexual behaviour or drug use, stronger assurances of confidentiality elicit higher response rates or better response quality. However, when the topic of the

research is innocuous, stronger assurances of confidentiality appear to backfire, leading to less willingness participation, and greater expressions of suspicion and concern about what will happen to the information requested, (Singer, 2004, p. 44). Together, these findings suggest that adverse effects of interventions occur sometimes and the present study suggests that this is also possible in the field of information security. Perhaps, just as there is a privacy paradox (H. J. Smith et al., 2011), there is an 'intervention paradox'.

Why did the interventions not succeed in reducing disclosure? A first possibility is that our interventions were not noticed. However, this is very unlikely. In the awareness condition, respondents had to answer questions on phishing and privacy. Accordingly, they had to read and to fill out and there were no missing values on these questions. The warning was handed over and was read before handing over the questionnaire. This could be checked by the researcher, as he was in direct contact with the respondents. There are no reasons to think that anyone skipped this (in this condition) and that this explains the lack of effectiveness.

8.5.1. Missing the link

A possibility, also based on our impression during study collection, is that respondents did not make the connection between the information that they provided and the more general issue of cybercrime or phishing. They believed that what they were filling in was neutral information and not something that could be abused by someone else. We saw that some respondents took pleasure in filling in their bank account number and considered this as a proof of their good memory. This lack of understanding of how security breaches occurs has been noted by several authors. The vast majority of computer users has little computer security knowledge or training. Nevertheless, they still make security-related decisions on a regular basis. Wash found that home computer users have a variety of different mental models of security threats (Wash, 2010). These models describe how users think about security, risks, prevention and suggest appropriate lines of action. These models are not necessarily accurate. Several studies described models that users have about their PC/Laptop, on the security threats that they face and, accordingly, what they could and should do about them (Blythe & Camp, 2012; Jones et al., 2011; Kauer, Günther, Storck, & Volkamer, 2013; Wash, 2010; Wash & Rader, 2011).

8.5.2. Distraction

It is possible that filling in the questionnaire distracted them from their shopping or other activities. Many studies showed that humans are not very good at executing more than one task at the time (Lavie, 2010; Pashler, 1998), and this might be true for security behaviour (Jenkins et al., 2013). While they were busy with other tasks, they filled in the questionnaire and were insufficiently able to focus on the task at hand. Accordingly, when having in mind their shopping activities, and the filling in of the questionnaire they hardly noticed the intervention. Fraudsters know this, distraction is a principle often used by

fraudsters. According to Stajano and Wilson (2009), 'Distraction is at the heart of innumerable fraud scenarios; it is also a fundamental ingredient of most magic performances' (Stajano & Wilson, 2009, p. 71).

8.5.3. Liking and reciprocity

Several principles guide our behaviour. Among them (Goldstein et al., 2008) noted, liking, and reciprocity: People are easily persuaded by other people that they like and they tend to return a favour. Dolan, in a review of the literature mentioned a 'messenger', and 'affect' (Dolan et al., 2010): we are heavily influenced by who communicates information and our emotional associations can powerfully shape our actions. It is possible that these principles have played a role in neutralising the effectiveness of the interventions. The researcher who collected the data was a friendly student of the University of Twente. The University of Twente has a good reputation and is well known in the region. It is possible that both the messenger and the university were 'liked', by the respondents. In addition, it is possible that handing over a leaflet leads to reciprocity: in exchange for the leaflet, the respondents filled out the questionnaire. This would explain why the warning may have an opposite effect in the case disclosing the name of the online shop.

The present study had some limitations. The sample is not exactly representative of a more general population, although it was broadly representative for population in Enschede younger than 65. The questionnaire had to be brief and future studies might desire to collect more information.

There were some advantages to the present study that are worth noting. Our study resembles a situation in which social engineers act. We did not place respondents in a special room, in a lab or a university classroom, but executed the intervention in the real world during the shopping activities of the respondents. As such, we believe our approach resembles real life interventions as well as real life phishing attempts.

The information requested to disclose pertained to topics that social engineers would be interested and that can be directly used for spear-phishing. The combination of knowing what someone bought in a particular shop, with information on the bank account number and an email address provides a good basis for a 'fine' spear phishing email. In contrast, many measures of disclosure in previous experimental research (John et al., 2011; Joinson et al., 2010) were based on questions on deviant behaviour (i.e., sexual behaviour, illegally downloaded music). Disclosure was measured as the number of times respondents did fill out the questions. It seems plausible that respondents who are not involved in deviant behaviour, might have been less motivated to skip these questions by not filling in or consciously reporting that they did not want to disclose this information. They might have felt that they were not 'disclosing' anything when they answered negatively. In this sense, not being involved in deviant behaviour is confounded with not-disclosing information and this disclosing measure is to some extent a measure of deviant behaviour. Finally, much of the previous research was based on student samples. In the present study we had personal contact with the respondents.

In conclusion, the two interventions that were tested were not effective to prevent people to disclose personal information. On one occasion we found an adverse effect of a warning. Security managers should accordingly be careful in developing interventions to avoid possible adverse effects. Our findings stress the importance of testing interventions.

9. Conclusions

The techniques described in this deliverable contribute to the TRE$_S$PASS model in a number of ways:

- The techniques can be used individually or in combination to help modellers better understand socio-technical contexts.
- The contextual understanding gained from using these techniques help modellers to decide where to focus the analysis and also what aspects of a scenario to model.
- The qualitative outputs obtained from these techniques can be used to focus surveillance and monitoring activities.
- The quantitative outputs obtained from these techniques can be used as input to the likelihood calculations performed by the model.

Throughout this deliverable there are numerous touch-points between the data-gathering methods and their theoretical underpinnings that are described in detail, and the notion of positive and negative security. It is noted that having trust in others has positive outcomes for individuals. We could imagine a possible scenario, or problem space, where several of the research methods described in this Deliverable could coincide to produce an enriched understanding of the positive and negative dimensions of security. Such a scenario might require interrogation on a number of levels, from the technical to the social. Password cracking is one instance where control strengths are often measured along a scale, from weak to strong, and gauging the baseline force the control is capable of resisting (Sect. 7.3.2.2). Positive and negative, black and white, can be used as a way of both measuring and visualising control strength of technical and social controls (Sect. 1.5.2). Password strength also has a social aspect, as utilised by attackers (Sect. 4.5.1). Affect clearly has a heavy bearing on the potential success of attack strategies that use components such as reliance on 'liking', or reciprocity, distraction, and the most commonly used strategy, authority.

Keywords are also a means of coding the social data provided by participatory physical modelling sessions, and help to qualify the different locations and practices that are represented as a rich picture or landscape view of the problem space under consideration. A number of these keywords can be described as ambivalent, being either positive or negative in character, depending on their context. Others are clearly in one camp or the other, and it is noticeable that, linguistically, there is a rich negative vocabulary, more so than for positive terms (Schrauf & Sanchez, 2004):

> Negative emotions signal problems or threat in the environment and are accompanied by detailed and systematic cognitive processing, while positive emotions signal a safe or benign environment and are accompanied by heuristic, schema-based cognitive processing (p.266).

Thus, the contextually situated negative and positive aspects of social practices contribute to the comprehensiveness or completeness of the modelling of these relationships, as can be clearly seen in the outputs of our participatory research methods (Fig. 3.7). In the international telecommunications field, 'Call Detail Records' (CDRs) and 'Customer Relationship Management' (CRM) systems are also theoretically able to be a source of negative and positive constructions upon social data that has a primarily technical basis, especially if analytical techniques are able to extract significant patterns that can be fed directly into the model as annotations, and hence into libraries that feed the Attack Map Navigator (ANM).

The discussion and investigation of persuasion principles used in attacks shows how the coding of narratives of social engineering attacks results in a tree of attack steps used; these are represented graphically as a tonally graduated tree with weighted lines reflecting the frequency of use of each of the typical strategies (Chapt. 4, Fig.4.8). In these narratives keywords, such as 'posing' for example, refer to classes of these techniques such as authority, impersonation, and others.

Survey responses concerned with disclosure of personal information, to take another example, were rated positively or negatively by being scored from 0 to 4, where the maximum must reflect the highest degree of risk that is possible (Sect. 8.3.2). While not described in binary terms, there are clear implications here for positive or negative security.

The same scenario could also be interrogated using participatory methods, as a way of getting closer to the mental models of that participants have of security threats. One important conclusion of this Deliverable is the importance of testing interventions before implementation, in order to avoid the potential adverse effects of interventions (Sect. 8.5.3). The playful, tangible, and collaborative nature of these engagements is precisely aimed at revealing the kind of threat that may not be immediately discovered in any other way, and may allow for these threats to be addressed by reinforcing service design vulnerabilities, such as interventions that may compromise the data or interests of clients. These were discussed as unforeseen consequences of alerts emanating from client accounts (Heath et al., 2014).

Negative and positive relational entanglements can only be said to make sense in relation to a model of these relationships. Such a model explains or describes how the relational dimension of social practices interact with the supporting technical infrastructures that enable information-sharing practices.

Unifying the methods of TREsPASS.

In order to imagine how information security might be better achieved requires temporarily, at least, moving away from the fixation on networks and network traffic and focusing on the security of people by looking at the social practices that surround information exchange, by going back to the physical environments in which trust and resilience are built.

It is imperative to study the physical places and the social situations where security and security risks typically occur, as well as those where 'everyday' routines prevent such events from occurring. This is in order to understand not just how, why and as part of what social practices human error creates a 'weak link', but where and how organisations have successfully avoided being made into the targets of attacks and where and how strong, resilient social networks are formed as a form of natural protection of assets.

This deliverable has presented a number of tools and methods for analysing social data, which when taken together and when acting *in concert* as they will within the TREsPASS interface, will enable an analyst to gain a well-rounded overview of multiple and layered risk assessments of an ever-increasing number of potential scenarios, and to arrive at decisions that are accordingly based upon this balanced overview. In this way the aim is to provide a *rich* or *replete* representation, well-supplied (filled) with appropriate information. This should also include the facility to establish linkages between disparate types of social and technical data, and deploy a unified conceptual expression of this at both finer and coarser levels of detail, and based upon the readily available grammar of positive and negative security.

The tools and methods described in this Deliverable help to better understand given socio-technical scenarios, and therefore the tools help the analyst to select the parts of a scenario that are most important, and that may require special attention to the social dimension. The outcome will also be that certain circumstances can be identified as being of particular interest, and that these may be added to an operational watch-list of situations to be on guard against. Some of the tools will also be of help to find numeric values with which to assess the probabilities of success for social engineering attack steps.

> A surveyable representation produces precisely that kind of understanding which consists in *'seeing connections'*. Hence the importance of finding and inventing intermediate links.
>
> *Ludwig Wittgenstein, Philosophical Investigations, Section 122.* (Wittgenstein, 1967)

References

Abraham, S., & Chengalur-Smith, I. (2010). An overview of social engineering malware: Trends, tactics, and implications [Journal Article]. *Technology in Society*, *32*(3), 183-196. Retrieved from http://www.sciencedirect.com/science/article/pii/S0160791X10000497 doi: 10.1016/j.techsoc.2010.07.001

Acquisti, A., John, L. K., & Loewenstein, G. (2012). The impact of relative standards on the propensity to disclose [Journal Article]. *Journal of Marketing Research*, *49*(2), 160-174. Retrieved from http://journals.ama.org/doi/abs/10.1509/jmr.09.0215 doi: doi:10.1509/jmr.09.0215

Adams, A., & Sasse, M. A. (1999). Users are not the enemy [Journal Article]. *Communications of the ACM*, *42*(12), 40-46.

Alnajim, A., & Munro, M. (n.d.). An anti-phishing approach that uses training intervention for phishing websites detection [Conference Proceedings]. In *Itng 2009 - 6th international conference on information technology: New generations* (p. 405-410). Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-77649328039&partnerID=40&md5=80b4dcb8ef7b346201811e724de724de (Cited By :5 Export Date: 20 April 2015) doi: 10.1109/ITNG.2009.109

Argo, J. J., & Main, K. J. (2004). Meta-analyses of the effectiveness of warning labels [Journal Article]. *Journal of public policy and marketing*, *23*(2), 193-208.

Arnold, F., Pieters, W., & Stoelinga, M. I. A. (2013, December). Quantitative penetration testing with item response theory. In *9th international conference on information assurance and security, ias 2013, gammarth, tunisia* (pp. 49–54). USA: IEEE. http://eprints.eemcs.utwente.nl/25270/.

Arnold, F., Pieters, W., & Stoelinga, M. I. A. (2014). Quantitative penetration testing with item response theory. *Journal of Information Assurance and Security*, *9*(3), 118–127. http://eprints.eemcs.utwente.nl/25266/.

Bada, M., & Sasse, A. (2014). *Cyber security awareness campaigns: Why do they fail to change behaviour?* (Report). Global Cyber Security Centre.

Beunder, K., Kerkers, M., & Orij, J. (2014). *The effects of awareness on the disclosure of personally identifiable information*. Thesis.

Blythe, J., & Camp, L. J. (2012). Implementing mental models [Conference Proceedings]. In *Security and privacy workshops (spw), 2012 ieee symposium on* (p. 86-90). doi: 10.1109/SPW.2012.31

Boeije, H. (2009). *Analysis in qualitative research*. SAGE Publications.

Bullée, J. H., Montoya, L., Pieters, W., Junger, M., & Hartel, P. H. (2015). The persuasion and security awareness experiment: reducing the success of social engineering attacks [Journal Article]. *Journal of Experimental Criminology*, *11*(1), 97-115. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0

-84925504369&partnerID=40&md5=9972abe661173891f7e9e2c69e7f9637 (Export Date: 20 April 2015) doi: 10.1007/s11292-014-9222-7

Bullée, J.-W., Montoya, L., Pieters, W., Junger, M., & Hartel, P. (2015). The persuasion and security awareness experiment: reducing the success of social engineering attacks. *Journal of Experimental Criminology*, *11*(1), 97-115. Retrieved from http://dx.doi.org/10.1007/s11292-014-9222-7 doi: 10.1007/s11292-014-9222-7

Bürgi, P. T., Jacobs, C. D., & Roos, J. (2005). From metaphor to practice in the crafting of strategy. *Journal of Management Inquiry*, *14*(1), 78–94.

Bursztein, E., Benko, B., Margolis, D., Pietraszek, T., Archer, A., Aquino, A., ... Savage, S. (2014). Handcrafted fraud and extortion: Manual account hijacking in the wild [Conference Proceedings]. In *Proceedings of the acm sigcomm internet measurement conference, imc* (p. 347-358). Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84910145784&partnerID=40&md5=d3f061dbba2174d93f35757317cece78 (Export Date: 20 April 2015) doi: 10.1145/2663716.2663749

Caputo, D. D., Pfleeger, S. L., Freeman, J. D., & Johnson, M. E. (2014). Going spear phishing: Exploring embedded training and awareness [Journal Article]. *IEEE Security and Privacy*, *12*(1), 28-38. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84896486103&partnerID=40&md5=eef0b991beb41dc5d4e3dd0ccf3a3177 (Cited By :1 Export Date: 20 April 2015) doi: 10.1109/MSP.2013.106

Charmaz, K. (2011). Grounded theory methods in social justice research. *The Sage handbook of qualitative research*, *4*, 359–380.

Cialdini, R. (2009). *Influence*. HarperCollins.

Cozby, P. C. (1973). Self-disclosure: a literature review [Journal Article]. *Psychological bulletin*, *79*(2), 73.

Davinson, N., & Sillence, E. (2010). It won't happen to me: Promoting secure behaviour among internet users [Journal Article]. *Computers in Human Behavior*, *26*(6), 1739-1747. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-77956187913&partnerID=40&md5=d9fc8227386bce45bea1adf7c0e269e0 (Cited By :22 Export Date: 20 April 2015) doi: 10.1016/j.chb.2010.06.023

Dechesne, F., Hadziosmanovic, D., & Pieters, W. (2014, Nov). Experimenting with incentives: Security in pilots for future grids. *Security Privacy, IEEE*, *12*(6), 59-66. doi: 10.1109/MSP.2014.115

Delhey, J., Newton, K., & Welzel, C. (2011). How general is trust in "most people"? solving the radius of trust problem [Journal Article]. *American Sociological Review*, *76*(5), 786-807.

Denzin, N. K., & Lincoln, Y. S. (2009). Qualitative research. *Yogyakarta: PustakaPelajar*.

Dindia, K., Allen, M., Preiss, R., Gayle, B., & Burrell, N. (2002). Self-disclosure research: Knowledge through meta-analysis [Journal Article]. *Interpersonal communication research: Advances through meta-analysis*, 169-185.

Dodge, R., Coronges, K., & Rovira, E. (2012). *Empirical benefits of training to phishing susceptibility* (Vol. 376 AICT) [Generic]. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84863964012&partnerID=40&md5=06c8f1b335036f595f1734a8ed2e26f8 (Cited By :1 Export Date: 20 April 2015) doi: 10.1007/978-3-642-30436-1_37

Dodge, R., & Ferguson, A. J. (2006). *Using phishing for user email security awareness* (Vol. 201) [Generic]. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-33845523685&partnerID=40&md5=e6144982fa9b6e63a5251832c94b2110 (Cited By :3 Export Date: 20 April 2015) doi: 10.1007/0-387-33406-8_41

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2012). The intergenerational transmission of risk and trust attitudes [Journal Article]. *The Review of Economic Studies*, *79*(2), 645-677. Retrieved from http://restud.oxfordjournals.org/content/79/2/645.abstract doi: 10.1093/restud/rdr027

Dolan, P., Hallsworth, M., Halpern, D., King, D., & Vlaev, D. (2010). *Mindspace: Influencing behaviour through public policy* (Report). Cabinet Office and Institute for Government.

Dourish, P. (2004). What we talk about when we talk about context. *Personal and Ubiquitous Computing*, *8*(1), 19–30.

Ferguson, A. J. (2005). Fostering e-mail security awareness: The west point carronade [Journal Article]. *EDUCASE Quarterly*, *28*(1), 54-57.

Fetchenhauer, D., & Dunning, D. (2009). Do people trust too much or too little? [Journal Article]. *Journal of Economic Psychology*, *30*(3), 263-276.

Frattaroli, J. (2006). Experimental disclosure and its moderators: A meta-analysis [Journal Article]. *Psychological Bulletin*, *132*(6), 823-865. doi: 10.1037/0033-2909.132.6.823

Giddens, A. (1984). The constitution of society: Outline of the theory of structuration. *Cambridge: Polity*.

Glanville, J. L., & Paxton, P. (2007). How do we learn to trust? a confirmatory tetrad analysis of the sources of generalized trust [Journal Article]. *Social Psychology Quarterly*, *70*(3), 230-242.

Goldstein, N. J., Martin, S. J., & Cialdini, R. B. (2008). *Yes!: 50 scientifically proven ways to be persuasive* [Book]. New York: Simon and Schuster.

Gollmann, D., Herley, C., Koenig, V., Pieters, W., & Sasse, M. A. (2015). Socio-Technical Security Metrics (Dagstuhl Seminar 14491). *Dagstuhl Reports*, *4*(12), 1–28. Retrieved from http://drops.dagstuhl.de/opus/volltexte/2015/4974 doi: http://dx.doi.org/10.4230/DagRep.4.12.1

Goodman, N. (1976). *Languages of art: An approach to a theory of symbols*. Hackett Publishing.

Greis, N. P., Nogueira, M. L., & Kellogg, S. (2012). *The millennial cybersecurity project. improving awareness of and modifying risky behavior in cyberspace* (Report). Kenan-Flagler Business School. The University of North Carolina at Chapel Hill. Retrieved from http://sites.duke.edu/ihss/files/2011/12/IHSS_FinalReport_MillenialCybersecurity_Greis.pdf

Hadnagy, C., & Wilson, P. (2010). *Social engineering: The art of human hacking*. Wiley.

Harrison, S., & Dourish, P. (1996). Re-place-ing space: the roles of place and space in collaborative systems. In *Proceedings of the 1996 acm conference on computer supported cooperative work* (pp. 67–76).

Harvey, A. S. (1999). Time use research: The roots to the future. In J. Merz & M. Ehling (Eds.), *Time use – research, data and policy* (p. 123-150). Baden-Baden Germany: Nomos Verlagsgesellschaft.

Heath, C. P., Coles-Kemp, L., & Hall, P. A. (2014). Logical lego?: Co-constructed perspectives on service design. *NordDesign 2014, Proceedings*.

Herley, C., & Pieters, W. (2015). "if you were attacked, you'd be sorry": Counterfactuals as security arguments. In *Proceedings of the 2015 new security paradigms workshop (nspw)*.

Hoeben, E. M., Bernasco, W., Weerman, F. M., Pauwels, L., & van Halem, S. (2014). The space-time budget method in criminological research. *Crime Science*, *3*(1), 1–15.

Hong, J. (2012). The state of phishing attacks [Journal Article]. *Communications of the ACM*, *55*(1), 74-81. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84856033479&partnerID=40&md5=74527c60014b6d6030bc96474073ef57 (Cited By :23 Export Date: 20 April 2015) doi: 10.1145/2063176.2063197

Hoogensen, G., & Rottem, S. V. (2004). Gender identity and the subject of security. *Security Dialogue*, *35*(2), 155–171.

Huang, Z., Shen, C.-C., Doshi, S., Thomas, N., & Duong, H. (2015). Cognitive task analysis based training for cyber situation awareness [Book Section]. In *Information security education across the curriculum* (p. 27-40). Springer.

Huysmans, J. (2002). Defining social constructivism in security studies: The normative dilemma of writing security. *Alternatives: Global, Local, Political*, *27*(1), S41.

Hägerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association*, *24*(1), 6-21.

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing [Journal Article]. *Commun. ACM*, *50*(10), 94-100. doi: 10.1145/1290958.1290968

Jansson, K., & von Solms, R. (2011). Phishing for phishing awareness [Journal Article]. *Behaviour and Information Technology*, *32*(6), 584-593. Retrieved from http://dx.doi.org/10.1080/0144929X.2011.632650 doi: 10.1080/0144929X.2011.632650

Jenkins, J., Durcikova, A., & Reeves, K. S. (2013). Exploring how dual-task interference influences end-user secure behavior [Conference Proceedings]. In *The dewald roode workshop on information systems security research, ifip wg8.11/wg11.13.*

John, L. K., Acquisti, A., & Loewenstein, G. (2011). Strangers on a plane: Context-dependent willingness to divulge sensitive information [Journal Article]. *Journal of consumer research*, *37*(5), 858-873.

Joinson, A. N., Reips, U.-D., Buchanan, T., & Schofield, C. B. P. (2010). Privacy, trust, and self-disclosure online [Journal Article]. *Human–Computer Interaction*, *25*(1), 1-24.

Jones, N. A., Ross, H., Lynam, T., Perez, P., & Leitch, A. (2011). Mental models: An interdisciplinary synthesis of theory and methods [Journal Article]. *Ecology and Society*, *16*(1), 1-13.

Kauer, M., Günther, S., Storck, D., & Volkamer, M. (2013). A comparison of american and german folk models of home computer security [Book Section]. In L. Marinos & I. Askoxylakis (Eds.), *Human aspects of information security, privacy, and trust* (Vol. 8030, p. 100-109). Springer Berlin Heidelberg. Retrieved from http://dx.doi.org/10.1007/978-3-642-39345-7_11 doi: 10.1007/978-3-642-39345-7_11

Kenrick, D. T., Neuberg, S. L., & Cialdini, R. B. (2005). *Social psychology: Unraveling the mystery* [Book]. Pearson Education New Zealand.

Kirlappos, I., & Sasse, M. A. (2012). Security education against phishing: A modest proposal for a major rethink [Journal Article]. *Security and Privacy, IEEE*, *10*(2),

24-32.
Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., & Pham, T. (2009). *School of phish: a real-world evaluation of anti-phishing training* [Conference Paper]. ACM. doi: 10.1145/1572532.1572536
Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (n.d.). Lessons from a real world evaluation of anti-phishing training [Conference Proceedings]. In *ecrime researchers summit, 2008* (p. 1-12). doi: 10.1109/ECRIME.2008.4696970
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159–174.
Lankhorst, M. M., Proper, H. A., & Jonkers, H. (2009). The architecture of the archimate language. In *Enterprise, business-process and information systems modeling* (pp. 367–380). Springer.
Lavie, N. (2010). Attention, distraction, and cognitive control under load [Journal Article]. *Current Directions in Psychological Science*, *19*(3), 143-148. Retrieved from http://cdp.sagepub.com/content/19/3/143.abstract doi: 10.1177/0963721410370295
Maggi, F., Sisto, A., & Zanero, S. (2011). *A social-engineering-centric data collection initiative to study phishing* [Conference Paper]. ACM. doi: 10.1145/1978672.1978687
Mann, I. (2008). *Hacking the human: Social engineering techniques and security countermeasures*. Gower.
McSweeney, B. (1999). *Security, identity and interests: a sociology of international relations* (Vol. 69). Cambridge University Press.
Mitnick, K., Simon, W., & Wozniak, S. (2011). *Ghost in the wires: My adventures as the world's most wanted hacker*. Little, Brown.
Mitnick, K. D., & Simon, W. L. (2002). *The art of deception. controlling the human element of security* [Book]. Indianapolis, Indiana: Wiley.
Mitnick, K. D., & Simon, W. L. (2005). *The art of intrusion. the real stories behind the exploits of hackers, intruders and deceivers* [Book]. Indianapolis, Indiana: Wiley.
Monk, A., & Howard, S. (1998). Methods & tools: the rich picture: a tool for reasoning about work context. *interactions*, *5*(2), 21–30.
Netherlands, S. (2015). *Bevolking; geslacht, leeftijd, burgerlijke staat en regio, 1 januari 2015 (population; sex, age, marital status and region, 1st january 2015)* (Web Page Nos. June 28, 2015). Centraal bureau voor de statistiek. Retrieved from http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=03759ned&D1=0,3,6,9,12&D2=129-132&D3=60,82,341&D4=25-26&VW=T
Omarzu, J. (2000). A disclosure decision model: Determining how and when individuals will self-disclose [Journal Article]. *Personality and Social Psychology Review*, *4*(2), 174-185.
Openshaw, S. (1984). *The modifiable areal unit problem*. Norwich, U.K.: 0-86094-134-5.
Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action: Presidential address, american political science association, 1997 [Journal Article]. *American political science review*, *92*(01), 1-22.
Pacione, M. (2009). *Urban geography : a global perspective / michael pacione*. Milton Park, Abingdon, Oxon ; New York: Routledge. (3rd ed. Includes bibliographical references and index.)

Pashler, H. E. (1998). Attentional limitations in dual task performance [Book Section]. In H. E. Pashler & J. C. Johnston (Eds.), *Attention* (p. 155-190). Hove, UK: Psychology Press.

Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives [Journal Article]. *Annu. Rev. Psychol.*, *56*, 365-392.

Pentland, B. T., & Feldman, M. S. (2007). Narrative networks: Patterns of technology and organization. *Organization Science*, *18*(5), 781–795.

Pieters, W., Hadžiosmanović, D., & Dechesne, F. (2014). Cyber security as social experiment. In *Proceedings of the 2014 workshop on new security paradigms workshop* (pp. 15–24). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2683467.2683469 doi: 10.1145/2683467.2683469

Pieters, W., Hadžiosmanović, D., & Dechesne, F. (2015). Security-by-experiment: Lessons from responsible deployment in cyberspace. *Science and Engineering Ethics*, 1-20. Retrieved from http://dx.doi.org/10.1007/s11948-015-9648-y doi: 10.1007/s11948-015-9648-y

Ratnasingham, P. (1997). The importance of trust in electronic commerce [Journal Article]. *Internet Research: Electronic Networking Applications and Policy*, *8*(4), 313-321. Retrieved from http://www.ingentaconnect.com/content/mcb/172/1998/00000008/00000004/art00003http://dx.doi.org/10.1108/10662249810231050 doi: 10.1108/10662249810231050

Rittel, H. W., & Webber, M. M. (1973). Planning problems are wicked. *Polity*, *4*, 155–69.

Roe, P. (2008). The 'value' of positive security. *Review of International Studies*, *34*(04), 777–794.

Roos, J., Victor, B., & Statler, M. (2004). Playing seriously with strategy. *Long Range Planning*, *37*(6), 549–568.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust [Journal Article]. *Academy of management review*, *23*(3), 393-404.

Salem, O., Hossain, A., & Kamala, M. (2010). Awareness program and ai based tool to reduce risk of phishing attacks [Conference Proceedings]. In *Computer and information technology (cit), 2010 ieee 10th international conference on* (p. 1418-1423). doi: 10.1109/CIT.2010.254

Schatzki, T. R. (1996). *Social practices: A wittgensteinian approach to human activity and the social*. Cambridge Univ Press.

Schrauf, R. W., & Sanchez, J. (2004). The preponderance of negative emotion words in the emotion lexicon: A cross-generational and cross-linguistic study. *Journal of Multilingual and Multicultural Development*, *25*(2-3), 266–284.

Schulz, K.-P., & Geithner, S. (2013). Creative tools for collective creativity the serious play method using lego bricks. *Learning and Collective Creativity: Activity-Theoretical and Sociocultural Studies*, 179–197.

Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Anti-phishing phil: The design and evaluation of a game that teaches people not to fall for phish [Conference Proceedings]. In *Acm international conference proceeding series* (Vol. 229, p. 88-99). Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0

-36849073159&partnerID=40&md5=64133f8b5f4e6fb108d9506e4c30d1f3 (Cited By :8 Export Date: 20 April 2015) doi: 10.1145/1280680.1280692

Sherman, L. W. (1995). Hot spots of crime and criminal careers of places. In *Crime and place: Crime prevention studies 4.* Willow Tree Press.

Shove, E. (2003). *Comfort, cleanliness and convenience: The social organisation of normality*. Berg Oxford.

Singer, E. (2004). Confidentiality, risk perception, and survey participation [Journal Article]. *Chance*, *17*(3), 30-34.

Smith, G. M. (2005). Into cerberus' lair: Bringing the idea of security to light. *The British Journal of Politics & International Relations*, *7*(4), 485–507.

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: an interdisciplinary review [Journal Article]. *MIS Q.*, *35*(4), 989-1016.

Sprecher, S., Treger, S., & Wondra, J. D. (2013). Effects of self-disclosure role on liking, closeness, and other impressions in get-acquainted interactions [Journal Article]. *Journal of Social and Personal Relationships*, *30*(4), 497-514.

Stajano, F., & Wilson, P. (2009). *Understanding scam victims: seven principles for systems security* (Report No. 754). Technical Report nr 754. Retrieved from Availableat: http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-754.pdf

Statistics Netherlands. (2014). *Dutch census 2011: Analysis and methodology*. The Hague/Heerlen: Statistics Netherlands.

Stave, C., Törner, M., & Eklöf, M. (2007). An intervention method for occupational safety in farming — evaluation of the effect and process [Journal Article]. *Applied Ergonomics*, *38*(3), 357-368. Retrieved from http://www.sciencedirect.com/science/article/pii/S0003687006000640 doi: http://dx.doi.org/10.1016/j.apergo.2006.04.025

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge. improving decisions about health, wealth, and happiness* [Book]. London, UK: Puinguin Books.

The TRE$_S$PASS Project, D1.3.3. (2015). *Dynamic features of socio-technical security models.* (Deliverable D1.3.3)

Tyler, T. R. (2003). Trust within organisations [Journal Article]. *Personnel review*, *32*(5), 556-568.

Von Solms, R., & Van Niekerk, J. (2013). From information security to cyber security. *Computers and Security*, *38*, 97–102.

Wash, R. (2010). Folk models of home computer security [Book Section]. In *Soups '10 proceedings of the sixth symposium on usable privacy and security* (Vol. Article No. 11). New York, NY, USA: ACM. doi: 10.1145/1837110.1837125

Wash, R., & Rader, E. (2011). Influencing mental models of security: A research agenda [Conference Proceedings]. In *Proceedings of the 2011 workshop on new security paradigms workshop* (p. 57-66). ACM.

Wenger, E. (1998). Communities of practice: Learning as a social system. *Systems thinker*, *9*(5), 2–3.

West, R., Mayhorn, C., Hardee, J., & Mendel, J. (2009). The weakest link: A psychological perspective on why users make poor security decisions. *Social and Human elements of information security: Emerging Trends and countermeasures*, 43–60.

Wheeless, L. R. (1978). A follow-up study of the relationships among trust, disclosure, and interpersonal solidarity [Journal Article]. *Human Communication Research*,

4(2), 143-157. Retrieved from http://dx.doi.org/10.1111/j.1468-2958.1978.tb00604.x doi: 10.1111/j.1468-2958.1978.tb00604.x

Wheeless, L. R., & Grotz, J. (1977). The measurement of trust and its relationship to self-disclosure [Journal Article]. *Human Communication Research*, *3*(3), 250-257. Retrieved from http://dx.doi.org/10.1111/j.1468-2958.1977.tb00523.x doi: 10.1111/j.1468-2958.1977.tb00523.x

Wittgenstein, L. (1967). *Philosophische untersuchungen-philosophical investigations*. B. Blackwell.

Wogalter, M. S., Laughery, K. R., & Mayhorn, C. B. (2012). Warnings and hazard communications [Book Section]. In *Handbook of human factors and ergonomics* (4th ed., p. 868-894). Hoboken, NJ: John Wiley and Sons, Inc. Retrieved from http://dx.doi.org/10.1002/9781118131350.ch29 doi: 10.1002/9781118131350.ch29

Worthy, M., Gary, A. L., & Kahn, G. M. (1969). Self-disclosure as an exchange process [Journal Article]. *Journal of Personality and Social Psychology*, *13*(1), 59-63. doi: 10.1037/h0027990

Wright, R. T., Jensen, M. L., Thatcher, J. B., Dinger, M., & Marett, K. (2014). Research note—influence techniques in phishing attacks: An examination of vulnerability and resistance [Journal Article]. *Information Systems Research*, *25*(2), 385-400. Retrieved from http://dx.doi.org/10.1287/isre.2014.0522 doi: 10.1287/isre.2014.0522

Wueest, C. (2014). Targeted attacks against the energy sector [Journal Article]. *Symantec Security Response, Mountain View, CA*.

Zak, P. J., & Knack, S. (2001). Trust and growth [Journal Article]. *The economic journal*, *111*(470), 295-321.

Zhang, B., Wu, M., Kang, H., Go, E., & Sundar, S. S. (2014). Effects of security warnings and instant gratification cues on attitudes toward mobile websites [Conference Proceedings]. In *Proceedings of the 32nd annual acm conference on human factors in computing systems* (p. 111-114). ACM.

A. Project Summary

This chapter gives an overview of the TRE$_S$PASS project and its use cases. The section is shared by the public deliverables to provide the necessary background and to put the current deliverable in context.

Information security threats to organisations have changed completely over the last decade, due to the complexity and dynamic nature of infrastructures and attacks. Attacks like StuxNet involve technical and human factors, and they damage physical infrastructure. The recent attack on a German steel mill [1] was a combination of both targeted phishing emails and social engineering attacks. The phishing helped the hackers extract information they used to gain access to the plant's office network and then its production systems. As a result, the technical infrastructure of the mill suffered severe damage.

The attack on the German steel mill illustrates that we need to integrate the social and technical aspects of systems in assessing their security - and we need to do so today. Socio-technical systems pose new challenges by combining parts for which we often understand the security issues; the combined system is however much more complex due to interactions between these parts.

The main innovation of the TRE$_S$PASS project is the attack navigator, a tool and metaphor that enables defenders to predict and preventing attacks on socio-technical systems. The attack navigator supports current risk-assessment techniques with the TRE$_S$PASS process (developed in Work Package WP5), an analytical approach to identifying attacks and evaluating their impact.

The four main stages in the TRE$_S$PASS process are *data collection*, *modelling*, *analysis*, and *visualisation*. Data collection (WP2) is vital to understanding the nature of a scenario and providing input to subsequent tasks of modelling, analysis and visualisation. Within the project, the focus has been on collection and analysis of social, technical and physical data and the ways in which these relate to one another. Within each of these domains, different approaches have been taken to provide different viewpoints on the nature of the organisation being investigated.

The models (WP1) developed in TRE$_S$PASS can be adapted to the application scenario. We have developed physical modelling techniques in order to understand where further investigation may usefully be targeted. The TRE$_S$PASS model describes relevant aspects of the organisation and their connections. To explore contractual and commercial relationships, the e3value method has been adopted.

[1] BBC News, *Hack attack causes 'massive damage" at steel works*, http://www.bbc.com/news/technology-30575104, last visited October 31, 2015.

The analysis methods (WP3) developed in TRE$_S$PASS identify attacks in models and identify the most effective controls to prohibit these attacks. The analyses are supported by tools and together they provide the defender with a comprehensive understanding of properties attacks, *e.g.*, cost for the attacker, required skills, or required time.

The innovative visualisations (WP4) developed in TRE$_S$PASS focus focus particularly on visualising elements of the analysis, as this is key to the overall project goal of providing "decision support" to practitioners. However, visualisations contribute also to model development and data gathering.

Practitioners can access the TREsPASS toolkit via the attack navigator map interface, which provides an intuitive means of selecting appropriate tools (WP6) for data gathering, modelling, analysis and visualisation. These can be used, individually or in combination, to strengthen operational and strategic decision-making.

A.1. Case Studies

The TREsPASS process and tools are validated by means of case studies (WP7) in the area of cloud infrastructure, telecommunications infrastructure, ATM infrastructure, and an organisation processing privacy sensitive data.

A cloud infrastructure shares infrastructure within or across organisations, giving the cloud services provider and its employees full physical and logical access to all resources across the different consumers. In TRE$_S$PASS we formalise typical components in cloud infrastructures as well as human actors and their interrelationships, to identify their contribution to attacks on the organisation.

In telco infrastructure new products need to be launched under significant time pressure, often opening up loopholes for so-called knowledge insiders who know the market very well, trying to make as much monetary gain from the new products as possible. In TRE$_S$PASS we model both the infrastructure and contractual relationships to identify physical and monetary attacks.

The ATM infrastructure connects machines that are composed of a money safe and a computer that controls the ATM's devices. There are well protected ATMs installed inside bank branches, while others are deployed in the street and some are not even embedded in a wall. ATM attacks are common and include classic physical attacks and emerging digital attacks. In TRE$_S$PASS we model ATM installations, and identify attack likelihoods using geospatial data.

The organisation processing privacy sensitive data develops a system supporting primarily elderly and disabled people in performing online payments and managing their own money from their home. This case study involves from strictly technical security aspects, such as how information is protected while stored or transmitted, to socio-technical security aspects covering security issues arising from the use of and interaction with the technology. In TRE$_S$PASS we identify social-engineering and trust-based attacks on such systems.

A.2. Overview of TRESPASS Integration

The TRESPASS workflow involves several stages with various activities, some of which are optional. Figure A.2 shows the architecture diagram and Figure A.1 shows a visual description of the notation used. In practice, stages may not follow a linear order. For example, depending on the goal of the risk assessment, new data requests may be issued later in the process, or automatic updates of data may be supported.

The **Data collection stage** prepares for analysis and modelling steps, and may require the gathering of one or more of the following kinds of data.

Physical data collection provides knowledge about the physical layout of the organization including locations, buildings, rooms, doors, windows, etc.

Digital data collection gathers information about the organization's IT infrastructure.

Social data collection focuses on organisational and individual data, and results in actor profiles containing, *e.g.*, attributes of employees, stakeholders, or potential attackers.

Commercial data collection gathers information required for *e3fraud* analyses, which focus on potential fraud.

Stakeholder goal collection identifies assets and policies the protection of which is critical to one or more stakeholders.

The **model creation stage** handles the creation of the TRESPASS model and associated actor profiles. The e3value model creation process is complementary to the main TRESPASS model, for cases requiring a more specific financial focus:

TRESPASS model creation is a key activity result in a system model that can be further extended and analysed.

- **Components customization (optional)** takes place before or during the TRESPASS model creation to create specialized custom model components.
- **Attacker profile creation** creates the attacker profile that the TRESPASS model analysis should consider, based on ready-made attacker profiles.
- **Defender/target profile creation** creates similar profiles for the other actors in the model based on the social data gathered in the social data collection activity.

e3value model creation This interactive activity involves using the *e3value toolkit* [2] to create business value models. These models structure the commercial information gathered in the data collection stage in a formal way.

In the **analysis stage** different analyses are possible depending on the model chosen. The analysis of the TRESPASS model involves these steps:

1. In the **attacker profile selection**, the user selects the attacker profile to use in the analysis.

[2] http://e3value.few.vu.nl/tools/

2. The **attacker goals creation** provides the attack generation with the attacker goals. These can be derived by hand from the stakeholder goals or deduced automatically from the selected attacker profiles.
3. The **scenario selection** selects a scenario, consisting of a single pair of attacker and attacker goal, to run the TRESPASS analysis on.
4. To extend attack trees, **attack pattern creation and sharing** provides libraries with known attack steps. The attack tree generation can only reach a certain level of abstraction, which may not be sufficient for quantitative analyses.
5. **Attack generation** transforms the TRESPASS model to an attack tree.
6. **Attack tree annotation & augmentation** then extends the attack tree with attack patterns and decorates leaf nodes with parameter values from the data collection stage for quantitative analysis.
7. The **attack tree analyses** compute quantitative properties of attacks, *e.g.*, utility for the attacker or success probability of the attack.

The analysis of the **e3value model** is complementary to the core TRESPASS analysis and has only one step:

1. For the **fraud model generation**, the user needs select an attacker and an interval of expected occurrence rates of the commercial transactions specified by the e3value model. The e3fraud tool then identifies all possible violations of contracts, the loss for actors, and the delta in profit for the other actors.

The **visualisation stage** can be used continuously to provide practitioners with feedback regarding the results of their activities:

1. **Fraud model visualisation** shows the generated attacks as a ranked list of textual descriptions of the attack steps and displays charts showing the profitability for each actor.
2. **Attack tree visualisation** shows the intermediary attack trees.
3. **Attack tree analysis visualisation** visualises analysis results.

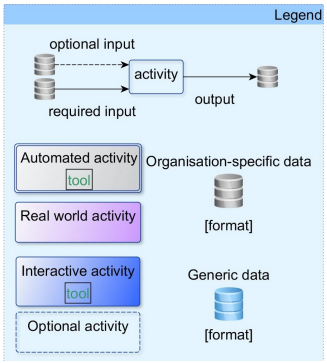

Figure A.1.: Legend for the Integration diagram in Figure A.2.

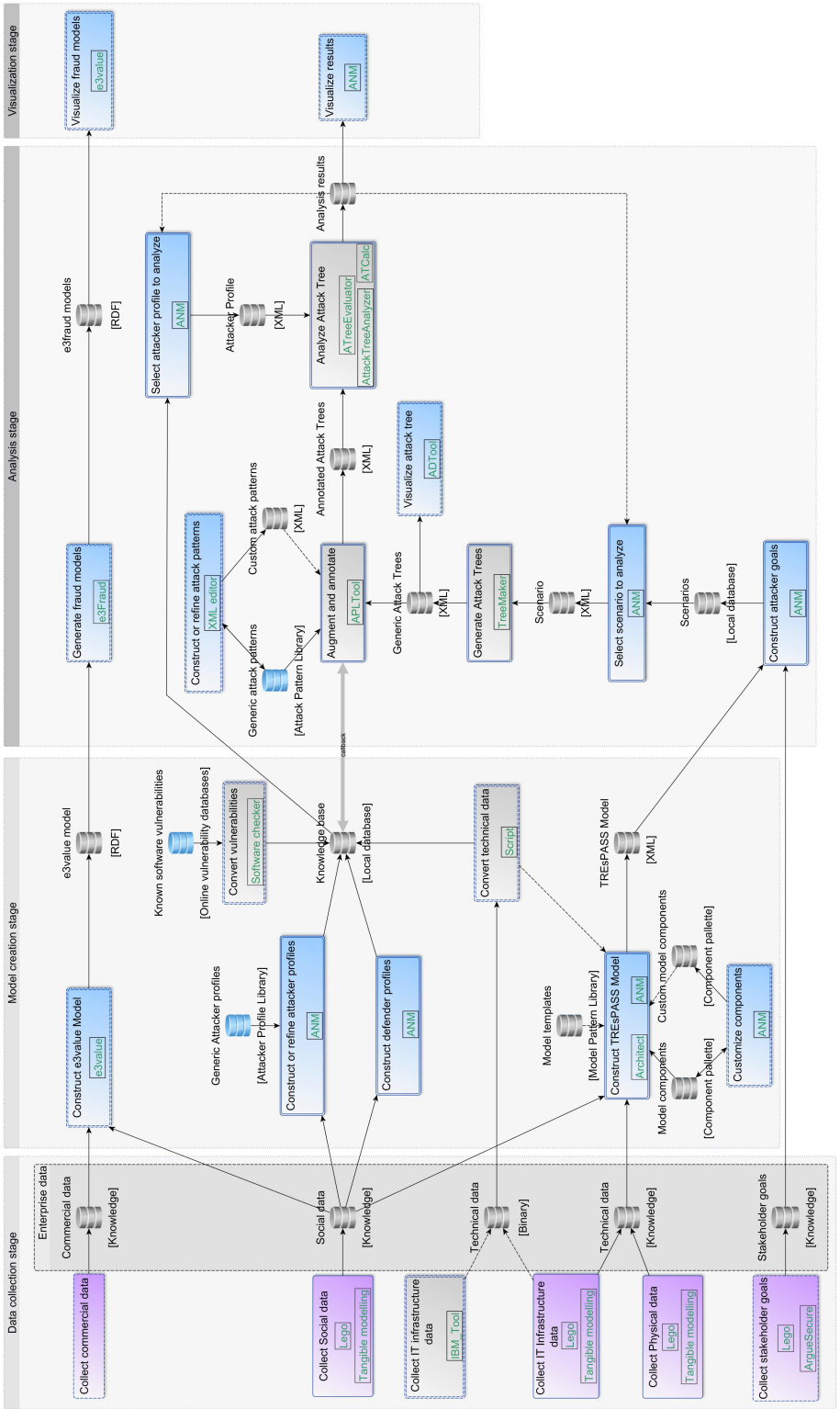

Figure A.2.: Integration diagram for the TRESPASS project.