Czech Technical University in Prague
Faculty of Nuclear Sciences and Physical Engineering

Department of Mathematics
Branch of Studies: Applied Information Technology

# Design of an Information System for Support of Forensic Audit

Bachelor's Degree Project

Author: Edita Pešková
Adviser: Mgr. Karel Macek, Ph.D.
Language Adviser: Mgr. Hana Čápová
Academic Year: 2015

Před svázáním místo téhle stránky vložíte zadání práce s podpisem děkana (bude to jediný oboustranný list ve Vaší práci) !!!!

**Declaration**

I declare that this Bachelor Project is all my own work and I have cited all sources I have used in the bibliography.

In Prague .....

.....
Edita Pešková

**Acknowledgements**

I wish to thank to my supervisor for his inspiring feedback, my family for their patience and support in my studies and anyone who encouraged me in this project.

Edita Pešková

*Název práce:*
**Návrh informačního systému pro podporu forenzního auditu**

*Autor:* Edita Pešková

*Obor:* Applied Information Technology
*Druh práce:* Bakalářská práce

*Vedoucí práce:* Mgr. Karel Macek, Ph.D.
Oddělení ekonometrie, Ústav teorie informace a automatizace
*Konzultant:* —

*Abstrakt:* Tato bakalářská práce předkládá návrh systému a naznačuje požadavky, které jsou potřebné aby byl popsán smysl a cíle technik digitálního forenzního vyšetřování, vykonávaných forenzními auditory, účetními a inpektory firem. Pomocí různorodých postupů, nástrojů a technik rozpoznáváme v jakých případech mohou nástroje forenzního auditu poskytnout auditorům potřebné informace k provedení forenzního auditu. Bakalářská práce představuje požadavky, které musí splňovat vlastní informační systém použitelný pro podporu vyšetřování a také poskytuje detailní návrh tohoto systému.

*Klíčová slova:* Forenzní audit, Vyšetřování, Návrh systému, Databáze, Sběr dat

*Title:*
**Design of an Information System for Support of Forensic Audit**

*Author:* Edita Pešková

*Abstract:* This bachelor project proposes a system design and suggests requirements that are needed in order to describe the purpose and goals of the digital forensic investigation techniques carried out by the forensic auditors, accountants and examiners of companies. Using various procedures, tools and techniques we identify where the forensic audit tools and the system can provide the auditors necessary information to carry out forensic audit. This bachelor thesis provides requirements that our own information system usable to support forensic audit must follow and also provides a detailed design of this system.

*Key words:* Forensic audit, Investigation, System design, Database, Data collection

# Contents

# Introduction

The topic of this bachelor degree project is forensic audit and its computer support. The field of information technology provides wide potential to support the process of forensic audit. First reason is that due to expanded use of information technology many important pieces of evidence can be found in digital environment. Second is the power that computer support gives in complex analysis of enormous amount of data occurring in investigated cases. The main goal of this project is to describe the process of forensic audit and design an information system helping in the process.

# Chapter 1

# Forensic audit and its computer support

In this chapter we define the term "forensic audit" and other related therms to give insight into the field. Next an overview of frequently investigated types of crimes is given. We demonstrate typical roles and outline the process of forensic audit.

## 1.1 Definition of forensic audit

The term "forensic" can be defined in several ways. According to merriam-webster dictionary, [3] the definition is "relating to the use of scientific knowledge or methods in solving crimes". The term "audit" is explained in the same dictionary as "a complete and careful examination of the financial records of a business or person" [2].

The essence of forensic audit is discovery and investigation of fraudulent intentions and fraudulent behavior.

Forensic audit is often mistakenly interchanged with financial audit. The objective of financial audit is to verify whether financial statements are fairly stated in accordance with accounting standards. Financial auditors search for material errors or other misstatements in the accountancy.

On contrary to financial audit the ultimate goal of forensic audit is to examine existing or gained suspicion and procure evidence concerning possible fraudulent behavior. Deceptive scenarios are discovered in the process of forensic audit and evidence together with a documentation that is usable for a subsequent course of action is gathered. As a matter of principle, forensic auditors are not expected to express their opinion on guilt or innocence of suspects.

## 1.2 Specializations in forensic audit

Forensic audit has already been defined in 1.1. Noteworthy specializations include following terms.

**Forensic Sciences** methodologically correctly apply a broad spectrum of scientific disciplines in order to answer questions relevant to a legal system. [7] An aggregate of these sciences is sometimes also called simply forensics. Forensics is mainly used to deal with previously committed crimes or to prevent similar crimes from happening.

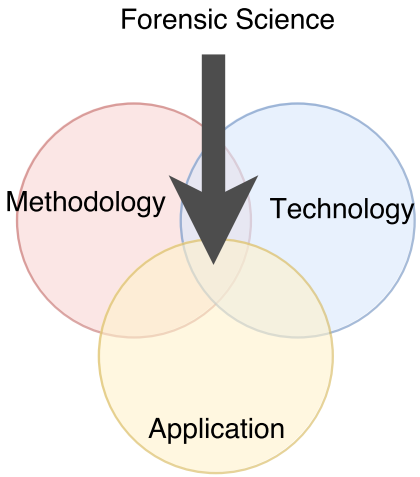



Figure 1.1: Forensic sciences

### 1.2.1 Computer (Digital) Forensics

Computer forensics, alternatively called digital forensics, is a discipline that investigates digital evidence. It is used for recording, safeguarding, extracting and documenting evidence from various kinds of hardware. Examples of digital forensics subbranches are mobile forensics, life-system forensics, file-system forensics, network forensics and multimedia forensics.

Digital evidence can be retrieved from plenty of digital equipment that are able to store usable pieces of information. These include not only computers, mobile devices, (smart) phones, (surveillance) cameras, PDAs etc., but also for instance copy machines, cars, washing machines or houses. This is a correlative of the Internet of Things (IoT) which is currently an emerging trend in technology. Common digital data used as an environment to search for evidence could include, for example, disk image of a device, a memory image of a device, NetFlow data, complete network communication, or even parts of log files gained from the service provider.

Digital data are created mostly when working with digital equipment. When using the RAM (Random Access Memory), programs regularly leave traces such as variables

and input/output data. The user also saves various kinds of data when saving documents and sending emails etc.. This kind of data remains in the memory and can be a source of evidence. For example, it is possible to trace the sending of an email in several instances such as in the browser, on the web page presence of an attachment is recognizable, the email is traveling via the network and the provider running the service has a log about the email. Having adequate authorization, such as the police has, it is possible to request a record about any activity from the provider.

The data in computer memory is divided into several groups. Volatile data are temporary data that is not permanently saved. Data available to users, typically files, are called active data. Meta-data are pieces of information about data. It typically consists of facts concerning the format of data, the time of creation etc.. Using meta-data of a picture as an example, it can store the time that was set in the device when the picture was created, facts concerning the location, the focus distance, the exact name of the camera and possibly other details. Similarly, in case of an document the facts concerning the time and the user who created the document can be found in the meta-data of the data. System logs are another type of data, where, for example, an facts concerning how the system is running, is provided. The purpose of these logs is to make a potential problem easier to find and resolve. Among other things, the system log monitors the activity of the user. Temporary files are files providing auxiliary space. Typically, the .tmp file extension is used and the file is used for saving intermediate results. Another type of data that can be found in computer memory is residual data. It is, in fact, a listing of the memory, the entity of unallocated memory, parts of files, deleted, but not overwritten files and other pieces of memory.

Forensic analysis of digital data searches for evidence in the field of information technology, their interpretation and presentation. The basic requirements are to find the desired evidence in a way that the methods are repeatable and the results are unquestionable. Forensic analysis always adopts an unbiased stance on the result.

First step in preparation of the data is its extraction from the device. These extracted data are also called "best evidence". Immediately after that the data is hashed and another copy of the data is made for processing purposes. This is important in case a change in the working copy is made while processing the data, the hash can always be easily check to verify the correctness of the data. It is important to document the procedure of extraction while extracting. It is possible, for example, to add a new entry to a log file, but it must be well documented to prevent any future mistakes.

### 1.2.2 Computational Forensics

Computational Forensics is a research domain that uses the means of computational methods to study any kind of evidence. Computational forensics involves modeling, computer-based analysis, computer simulations etc.. This specialization uses methods such as machine learning, data mining, image processing, statistical pattern recognition, fuzzy logic or data visualization to discover any hidden forensic knowledge.

The subbranches of digital forensics as well as the specializations of computational forensics play their own part in the investigation of complex cases. They are used to provide intermediate results for the investigation by answering closely specified questions. Some of these methods can also be used indirectly as functionalities provided by a software that is used in the investigation.

## 1.3 Existing software analysis

The existing software in the market is either inefficient in terms of results or does not full fill the requirements for conducting proper evaluation. For this reason, Computer Forensic Tool Catalog has been created as an effective way of connecting practitioners to the tools they need. This catalog is available on the "National Institute of Standards and Technology" web pages [8]. The information about forensic functionalities and associated technical parameters and their values can be provided by any vendor, in order to be integrated to the database. All tools are being reviewed before posting. This catalog can be considered rich in the database and it is also systematic and searchable. The only problem is it is not possible to select free tools only. This catalog can be recommended to any practitioner as a useful instrument for choosing suitable software for their situation. Although other lists of software can be found browsing the Internet, this one is reliable and provides the possibility of search in the tools according to technological needs.

The functionalities of the tools contained in the catalog are mostly usable in one of the first stages of forensic audit. This stage could also be called preparatory stage because the auditor is preparing background for the search for evidence. Documents and all possibly available data related to the studied case are being collected. Depending on the character of the case, data from various pieces of hardware can be collected. In some situations, if the tool is sophisticated enough, the following analysis of the data can be performed immediately within the same tool.

The work of the auditor usually starts by getting to know the situation because each case has its specific character. After that they need to collect all the hardware that holds important data and extract all the data stored in these devices. According to the type of device they may use proper software that is good for acquisition of the data stored.

**Quantative software** Specialized fraud identifying software is making a passage into the business, with extortion module discharges from both ACL and IDEA®, extortion parts in SAS and SPSS, and a devoted extortion discovery program in Picalo. Programming organizations like ACL and IDEA® give exercise manuals that can be utilized as a part of courses, however, these exercise manuals have couple of illustrations of direct misrepresentation recognition particularly with cutting edge procedures. Approaches like the theory testing methodology are an initial phase in giving procedure research, immense extra research—both observational and field—are

expected to approve and augment the current ideas.

Quantative software organizations give complete preparing on measurable programming and systems. The two driving programming bundles are EnCase by Guidance Software and the Forensic Toolkit (FTK) by Access Data . These suites give shorter expectations to absorb information than past utilities and convey a more prominent number of experts to the field more rapidly. Both bundles give tasks positioned procedures to securing and illustrating forensic information.

Lately, Linux based apparatuses have been prevalent as free distinct options for the conventional suites. Helix, the Penguin Sleuth, and Security Tools Distribution are Linux dispersions that run specifically from CD, giving clean situations to seeking a PC without the requirement for cloning (Causey, 2005). These instruments boot a suspect PC straightforwardly to Linux and mount the client hard drives in read just mode, basically bypassing most passwords and security insurances. While Linux based instruments are more hard to use and do not have the same point of reference in court as EnCase and FTK, they have to be well known with a few evaluators.

**The R Project for Statistical Computing** R is mathematical software specialized in statistics. It concerns an open-source implementation of the S language, which is used by other professional statistical applications. It is a good choice in case that great amount of data needs to be processed because R is suited for large data sets.

**Using SAS to evaluate risks** Measurable packages like SAS and SPSS give full drifting modules to the intrigued inspector. Statement on Auditing Standards (SAS) and consideration of fraud in a financial statement audit requires inspectors to evaluate the risk that misrepresentation might physically misquote budgetary articulations. Regardless of SAS 99 being an imperative necessity towards expanded misrepresentation, a study by "Marczewski and Akers"[6] uncovered that Certified Public Accountants (CPAs) do not envision that SAS 99 will generously expand review viability. Another study found that while SAS 99 expanded examiner obligation, most reviewers experienced issues recognizing the work and its risks.

**SPSS for statistical and analytical services** SPSS is one of the leading analytical instruments. It provides statistical and analytical services the abbreviation originally meant Statistical Package for the Social Sciences. This software supports areas of social survey and market research, predictive and advanced analytics, decision management and deployment or predictive solutions.

**COFEE** Useful tools for basic forensics are Microsoft's COFEE tools. They are a set of PC forensic and evaluating tools that Microsoft puts on a USB key and provides for law requirement. It is used in attempting to concentrate information from a PC. There was some apprehension that it was an "indirect access", however, individuals demanded it was no such thing, yet only a gathering of essential tools. Still, the this framework was advanced as being valuable for decoding passwords

and examining a PC's information and web action that appeared to be alarming. If Microsoft was giving it out to law requirement, it appeared to be likely that others would have entry to it also.

**IDEA®** IDEA® is an effective and easy to understand information investigation tool intended to help reviewers; bookkeepers and other account experts perform information examination rapidly to help enhance reviews and distinguish control breakdowns. It permits to dissect 100% of the information to ensure information trustworthiness and gives simple investigation more that 100 major studied activities.

At the point when the information originates from distinctive sources and in a mixed bag of configurations, information investigation can infrequently be a test. The capacity to import different information sets and represents it as if they were one that lets us to see the master plan. We have the capacity to recognize connections, examples and inconsistencies and additionally direct a far reaching examination of value-based information.

Import Data from Practically Any Source IDEA® permits us to rapidly import a boundless measure of records from for all intents and purposes of any source, including spreadsheet and **database programming**, mid-extent bookkeeping projects, ERP frameworks, legacy centralized servers, telecommunication switches, travel and costs applications, level and printed documents, for example, PDFs, plain content (.txt), and print-report (.prn) records. Rearranged analysis instead of programming macros can use more than 100 review particular undertakings that easily search for copies, identify dumps in numeric groupings, group information by classes and channel various lines and sections of data in seconds. IDEA® permits us to record each investigative step and reuse for future utilization, through a graphical interchange and customized interface.

**Microsoft Analytics Services** Procedures utilized for misrepresentation discovery fall into two essential classes: factual systems and computerized reasoning. Examples of measurable information examination methods are:

- Data preprocessing strategies for discovery, acceptance, slip adjustment, and topping of missing or off base information.
- Calculation of different factual parameters, for example, midpoints, quintiles, execution measurements, lihelihood circulations. For instance the midpoints may incorporate normal length of call, normal number of calls every month and normal defers in bill installment.
- Models and likelihood dispersions of different business exercises either regarding different parameters or likelihood circulations.
- Computing client profiles.
- Time-arrangement investigation of time-ward information.

- Clustering and grouping to discover examples and relationship among gatherings of information.
- Matching calculations to recognize abnormalities in the conduct of exchanges or clients when contrasted with already known models and profiles. Methods are likewise expected to dispose of false alerts, assessment chances, and foresee eventual fate of current exchanges or clients.

## 1.4 Types of investigation

Forensic auditor can be asked to investigate extensive range of situation types. Any kind of suspicion of potentially unlawful behavior can be an impulse for investigation. However, not all kinds of crimes are within the scope of this bachelor's degree project. For example, violent crimes, cash theft, assault, robbery, rape, actual bodily harm and similar crimes fall into the detective work. Forensic auditors typically do not deal with such crimes.

On the contrary, fraudulent crimes are those ideally fitting for forensic investigation. There are several different kinds of fraud suitable for forensic auditors. To provide an overview of the wide range, these frauds can be classified into following three groups: corruption, asset misappropriation and financial statement frauds. These frauds are the most frequently investigated scenarios.

**Corruption** Corruption can be characterized as a misuse of status. The motivation behind corruption is a desire to gain money, material possession or other advantages. By extension an involvement in a corruption relationship is a sign of a moral and ethical failure of individuals. Specific kinds of corruption include bribery, acceptance of a bribe, extortion, misuse of power, misuse of information or conflicts of interest. Research shows that corruption is involved in around one third of all frauds. [9]

**Bribery** is an unauthorized advantage in the form of recipient's behavioral alternation in exchange for money, goods or other forms of recompense. On contrary to bribery, extortion occurs when money or some action is demanded in order to secure a particular outcome. Conflict of interest and misuse of power or information can be categorized as belonging to one group. In such situations fraudsters use their power to achieve personal profit without regard to their vocational responsibility. Consequently, their behavior negatively influences the company which entrusts them with power. For example, a manager of a local division of a given company would choose his acquintance to be the paper supplier, even though his price is significantly much higher than the price of other paper suppliers.

**Asset misappropriation** is the theft or embezzlement of company assets by directors, other fiduciaries or employees. It also is by far the most common economic crime experienced by organizations reporting any fraud, with 69% of respondents

suffering from it. [1] Asset misappropriation takes place when the assets of an organization are stolen by those who are entrusted with their management. The theft usually starts at small scale and increases the size when perpetrators gain confidence. In consequence asset misappropriation can be a serious problem and can cause immense losses.

Concrete examples include cash theft, inventory theft or the use of company assets for personal purpose. A special case of asset misappropriation is fraudulent disbursement, i.e., fictional billing or raising an invoice that is not in compliance with the field of the supplier's enterprise. An interesting particularity is that asset misappropriation is far more likely to be committed by individuals than a group of perpetrators[4].

**Financial statement fraud** This kind of fraudulent activity is based on deliberate falsification of accounting records. It normally includes manipulation with expenses, manipulation of revenue figures, overstatement of assets or improper disclosure. This is regularly done to give the budgetary proclamations a specific inclination, for instance covering liabilities to enhance any examination of liquidity.

Aside from the investigations mentioned above, forensics can also be used to help solving partial questions found in the process of investigation. Similarly, when partial issues such as file recovering have to be dealt with, a specialist in forensics might be asked for help.

## 1.5 Preparation of forensic audit

On the basis of what had been found the process of forensic audit works as follows: When it is decided that a certain situation is to be investigated in forensic audit, it is important to prevent all individuals that are being investigated from access to all related documents and electronic evidence. It is also recommended to limit their access to corporate information systems.

The next step is to properly formulate the assignment and to define the extent and expectations of the outcome of forensic audit. In order to prevent a misunderstanding, the assignment should be as specific and detailed as possible. It is best to choose the right audit company based on references and its experience with similar cases as the one that has previously beepecified.

When a client contacts an audit company with an assignment a mutual meeting is scheduled and the assignment is formulated and a contract signed. The oclient should be prepared to provide access to corresponding electronic and paper documentation as well as accept the fact that auditors are going to question employees and case-related persons. On the other hand the audit company undertakes to refrain from sharing all the confidential information with third parties. A team of specialists that are convenient to the assignment is formed and the inspection is launched.

In some cases the extent of the order is quite complex such as investigation of

processes in one separate division of a company or investigation of complex corruption crimes. Nevertheless there may also be much easier cases where the audit firm is asked only to recover certain documents, or to examine particular piece of hardware.

The following steps of precise methods of forensic audit are not definite. The ability to adapt in new situations is one of many essential capabilities for the team of forensic auditors. The variety of investigated cases is so vast that there is no universally valid and precise course of action in the same time. Therefore on this place we present only general methodology of forensic audit. Several selected methods of forensic audit will be described later in this document.

## 1.6 General methodology of forensic audit

In this section we present basic phases that are used while performing forensic audit. This process is most commonly divided into four stages: Accumulation, Examination, Analysis and Reporting. This basic methodology starts after the selection of the audit company and after the specification of the assignment; at the same time when the real work of forensic auditors begins.

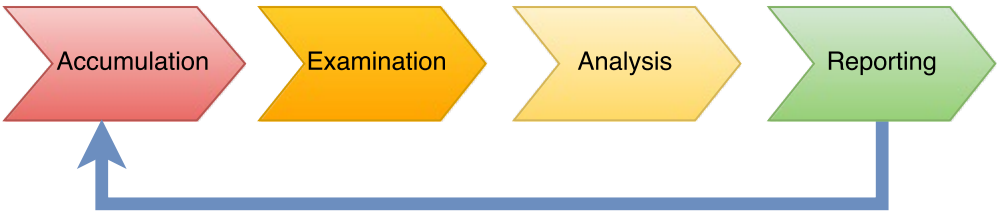



Figure 1.2: General methodology of forensic audit

**Accumulation:** The main purpose of this stage is to acquire as much usable data as possible. It means recognize possible sources of data and provide backup record. All the sources of data and information, including necessary cross examination and other sources of evidence, should be utilized in this phase. All the information from the conceivable sources of pertinent information should be gained.

**Examination:** Examinations include forensically preparing all the gathered data. This can be done using a blend of computerized and manual systems to survey and concentrate specifically compelling information.

**Analysis:** The following period of the procedure is to investigate the consequences of the examination, using legitimately reasonable routines and systems. The aim is to infer helpful data that addresses the inquiries that were the impulse for performing the accumulation and examination.

**Reporting:** The last stage is reporting the consequences of the investigation, which may include depicting the activities utilized, clarifying how devices and methods were chosen, figuring out what different activities should be performed and giving proposals to change to approaches, rules, techniques, apparatuses, and different parts of the measurable procedure.

## 1.7 Use case diagram of forensic audit in general

As we see it there are three basic roles in forensic audit. A customer who wants a particular case to be investigated. The action of the customer is to assign the order to a audit company. A representative of the company i.e., for example, forensic auditor takes over the order and starts the investigation. The operations requiring specialized skills can be delegated to an expert from the proper field. Forensic auditors together with specialists then investigate the case and search for the evidence. In the end of the investigation the auditor reports the conclusion to the customer.

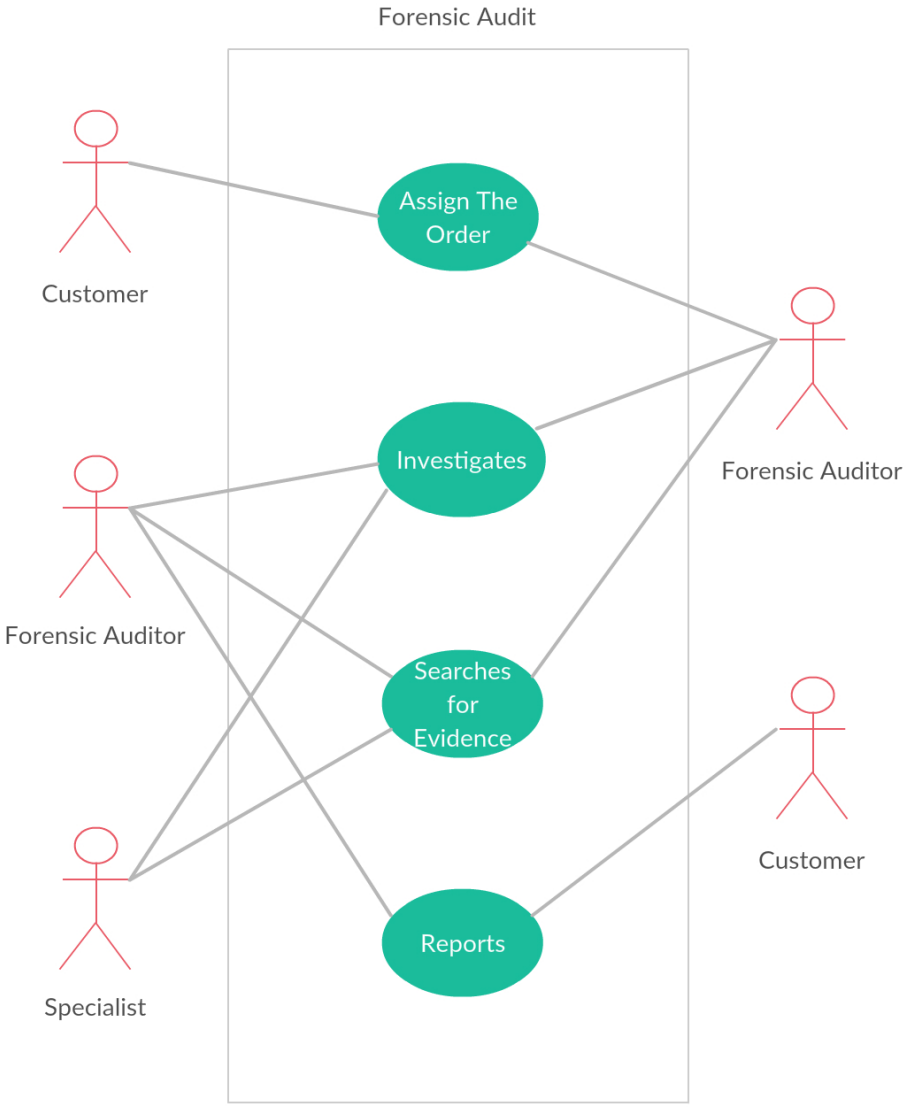



Figure 1.3: Use case diagram of forensic audit in general

# Chapter 2

# Original methodology for a computer aided forensic audit

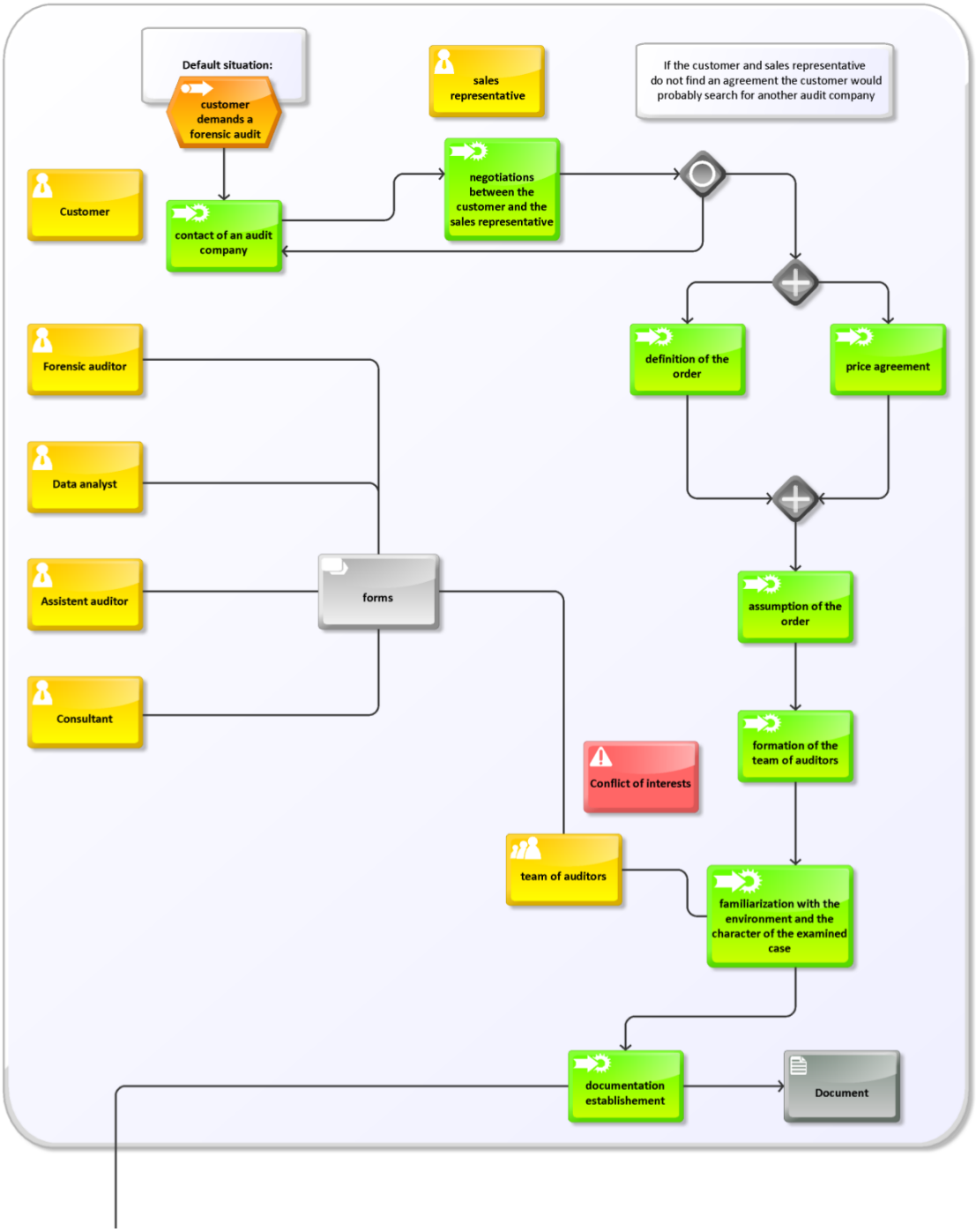
We are aware to have only limited knowledge of forensic audit and no practical experience. This chapter should not be considered as a manual explaining how forensic audit should be done. However, this is rather a serious attempt to describe the process of forensic audit.

## 2.1 Preparation of the process of forensic audit

Before the process of forensic audit starts there is a block of organizational or administrative affairs that needs to be done. It all starts in an institution where there is a suspicion of an act which is not following required rules. The institution can be of various kinds. For example it can be an international corporation suspecting one of its local branches of a money leakage.

The rules might be either internal regulation, corporate policies or even law. Basically suspicion of any kind of misbehavior can be investigated by forensic audit. Given the background of the case, the ordering party might also be from various branches of economy. They usually suspect possible internal or external risks that can be represented by fraudsters or people committing some other type of crime. Generally the ordering party is the head of a given company. Some specific cases are: CEO of certain company, new holder of an existing company, authorized representative of a supervisory board etc.. However, forensic audit can be also assigned by a court or the police.

After the decision to perform forensic audit the ordering party contacts the audit company and they negotiate with a business representative about the sphere of the contract, price and deadlines. If there is no agreement it is probable that the ordering party contacts another audit company. It is important to familiarize the ordering party with the fact that while conducting forensic audit the auditors might need access to various confidential information concerning the ordering company. On the other hand the audit company promises to manipulate with this data in a protective

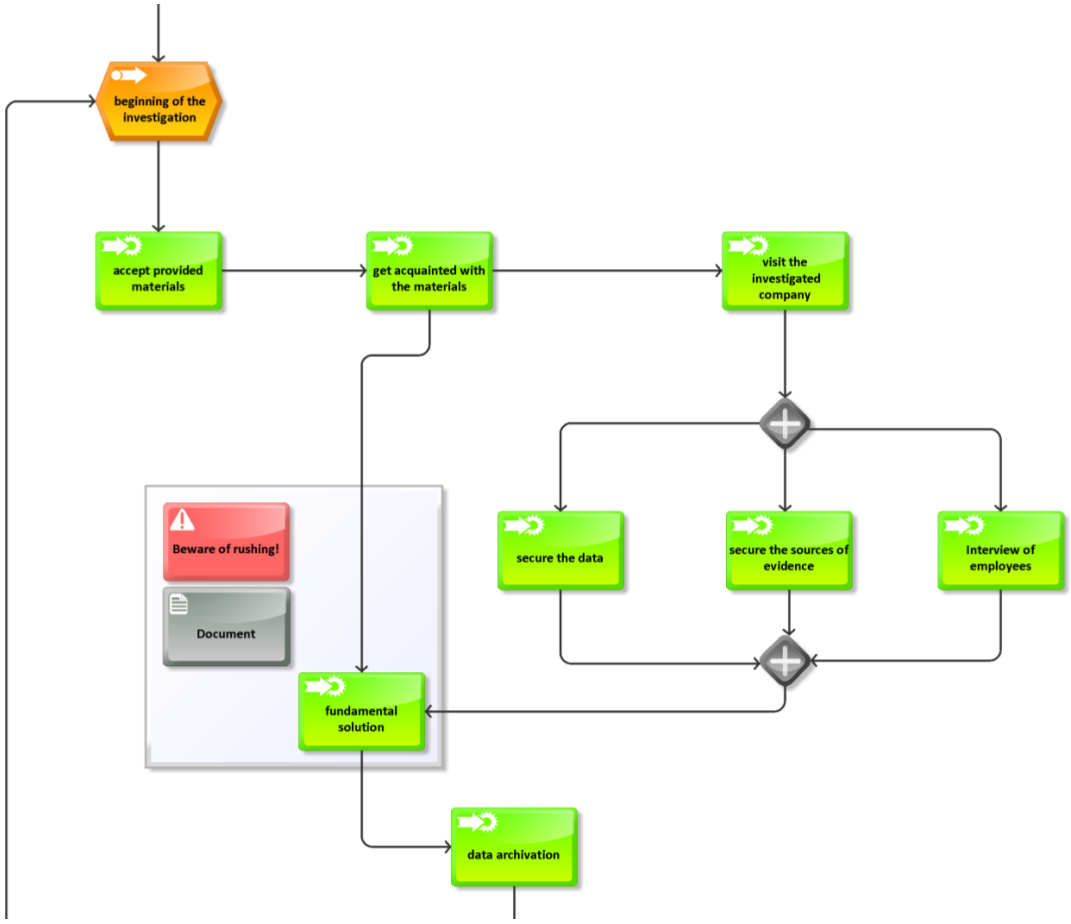



way and not to expose it to the risk of misuse. If all parties agree with the conditions the contract is made and the case is accepted.

The audit company has a team of specialists established according to the character of the case. It is important to be aware of conflict of interest in this phase. Members of the team must not be related in any way to the original case. Members of the team can be forensic auditors, data analytics, specialists in any particular field (biometry, data recovery, criminology, law, etc.), consultants and assistants.

When the team of auditors is established the next step is to create a document for record of the investigation. It also serves as documentation. It is also common to create a nickname for important subjects and for the case to provide security of the confidential information. This is the end of all the administration that needs to be done before the actual forensic audit.

## 2.2 Accumulation of data




At the beginning of the investigation the team needs to gain an insight into the background situation of the case. They can learn it from the materials they have been provided with the client. An appropriate strategy for investigation is created. It means that all the relevant methods are taken into account of and the best ones are chosen.

In all the cases it is important to secure all the documents, data repositories and all other possible sources of (digital) evidence. No unauthorized person is allowed to manipulate with any possible evidence. For this reason a backup of all the digital data is made. According to a FBI statistic the average investigated case size is approximately 500 GB. [5] In fact there are usually two copies of the data. The first is only for backup and the second can be used for analysis and other work with the

data.

The investigating team needs to evaluate all possible available sources of important information, access and gain the data. If the case is somehow special an expert might be needed to be authorized to collect specific pieces of evidence. An example could be a case in which a fingerprint recording is needed.

It might be useful to visit the workplace of the investigated company and check if there are some other possible sources of evidence. After having all the data secured it can also be beneficial to perform a cross examination in some cases . It can even be done by mere conversation with involved persons, but the dialogue should be recorded.

It is possible that already in the end of collection of data the forensic audit team has a idea or even hypothesis about what happened in the case. Having this idea might be useful in further investigation. On the other hand auditors should beware of jumping into conclusions. All the actions and also any potential hypothesis should be documented in the case documentation.

## 2.3 Examination

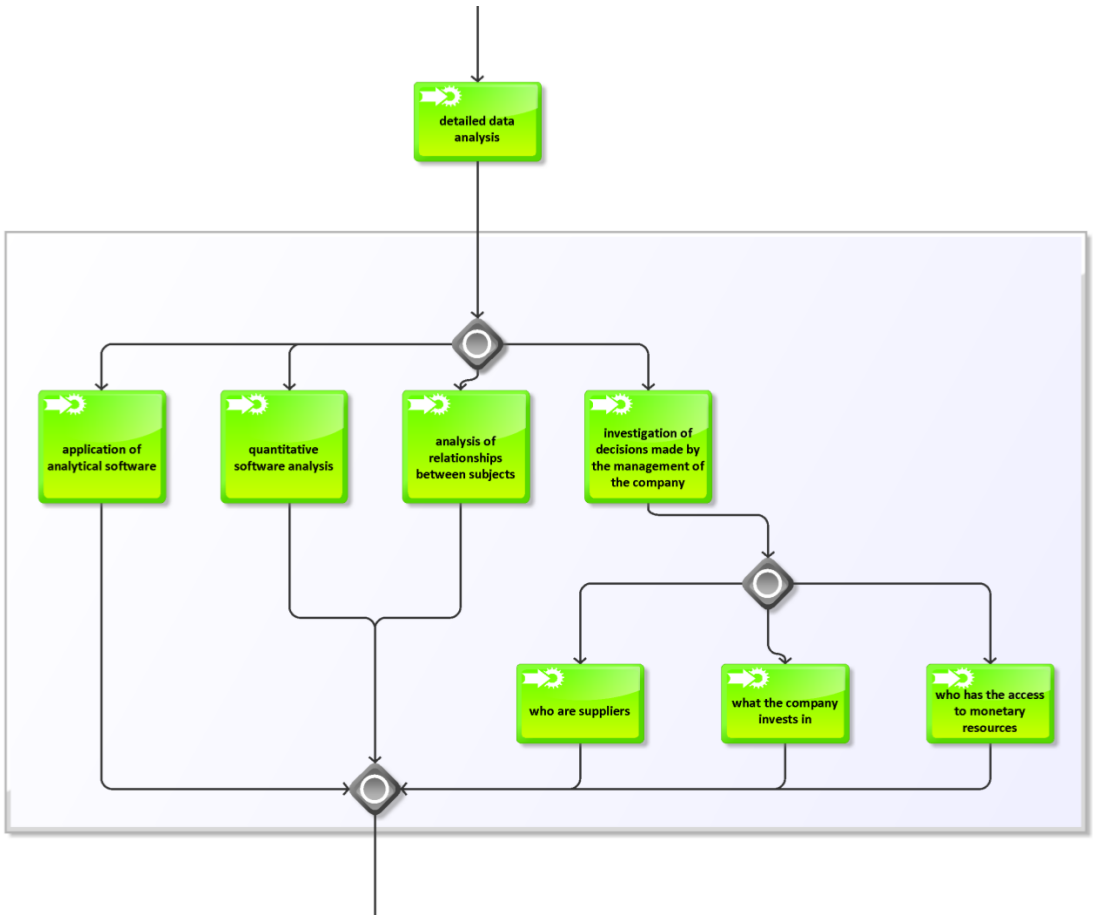
The examination phase comes when all the data is collected. Methods of examination are very different according to the character of the case, the sources of data and also the field that is investigated.

In the examination phase it is necessary at first to assess the data and mine the relevant pieces of information from all the collected data. It starts by identification of the data files that contain information of interest. Forensic auditors must not be discouraged by the size of data. After those files are identified it is often demanded to filter the extraneous information and leave only the coarsely filtered data.

## 2.4 Analysis

The most extensive part of the investigation comes after the data is pre-filtered. The aim of analysis is to study and analyze the data to draw conclusions from it or to determine that no conclusion can be drawn. By the end of analysis most important subjects, events, people, and people and relationships between them should be recognized. In order to find the conclusion it is required to unite information gained from multiple sources of data.

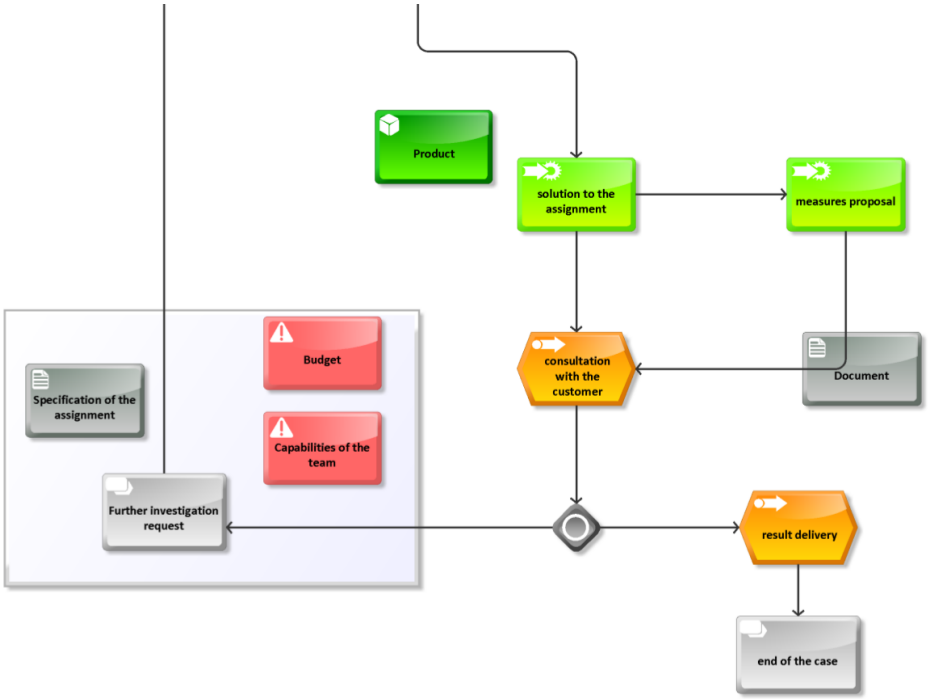
However first of all specific sources of data need to be analyzed by a competent specialist to obtain the information hidden in the data. For example, provided that a file contains billing data, the task is delegated to accountancy specialists. They can use appropriate accounting software or extract only data of particular interest and use some quantitative software. If the data contain log files or deleted files it is a job for information technology specialists to recover or examine the data.




Some of the methods the team of investigators might use are: quantitative analysis, analysis of relationships between persons, making use of analytical software. In case of suspicion from corruption an investigation of decisions made by the management of the company might be needed. Investigators might pose questions similar to following: Who are the suppliers of the investigated company? In what business the company invests and who makes the decisions about it? Who does have an access to (financial) resources?

There might be cases when specialists in various sub-branches of forensics might be needed. Even an external company might be hired to solve a particular problem. Examples of auxiliary specialists can be experts in computer forensics, forensic dactylography, digital forensics, forensic biometry, forensic psychiatry, forensic linguistic etc. According to opinions of all available experts the situation in the case is gradually revealed.

All the steps taken in the analysis result in understanding of the situation in the case. A verdict is pronounced according to the information investigators gained and it is retrospectively supported by the data collected in the beginning of the process.




## 2.5 Reporting

After having investigated the assigned case the team of forensic auditors has some results to show. Either they have found what undesirable actions happened or there might be cases when they have not discovered any signs of such behavior. In both cases the results are presented to the ordering party.

There are multiple possible results of previous investigations. Basically either a suspicion has been confirmed or it has not. If the investigation discovered a crime the exact methods that lead to the discovery must be mentioned in the report together with the evidence supporting the result. The report should clarify if there is enough evidence. If there is not enough evidence the type of data needed to prove the result should be described and reasoned. The explanation why there is no evidence yet should be also given.

If no crime has been discovered in the reporting stage, it still must be explained in the report what data has been investigated and how the investigation was conducted. Everything that has been found out about the case should be presented in an appropriate form. Special effort should be made in presenting possible alternative explanations. One of them might be that there has not been enough time to process all available data. In that case it leads to a customers decision. They decide whether they want to continue the investigation in more details or whether they are content with the result and they already believe that no crime has been committed. This situation is very case-specific.

According to the result of the investigation several additional information might be suggested. This might be for example an advice about subsequent actions preventing

repetition of the undesirable behavior. While investigating a case forensic auditors might sometimes discover a completely new problem in the investigated company. This additional information should be also included in the report.

The actual reporting is affected by many factors. When preparing the final presentation these factors should be considered. First, the presentation should be prepared with regard to the audience the speaker is going to have. The extent of professional details regarding the methods used in the process should be adjusted to the anticipated knowledge of the audience. However, if asked, the speaker should be prepared to explain the details. The presentation should be easily comprehensible and explain everything discovered in the investigation.

One part of the report can be focused on alternative explanations of what happened. This may happen if the information that has been found about an event is incomplete. Alternatively it may happen if there are more plausible explanations about what happened. Each of these explanations should be analyzed in the previous phase and an attempt to prove or disprove them should be made.

Besides an oral presentation the output of the assignment should be also a written document describing all the important facts about the case. This document should be handed over to the ordering party and should contain:

- information about the agreement between the two sides,
- list of sources of evidence in the investigation,
- description of methods used in the investigation,
- the sequence of important evidence found in the data,
- explanation of the evidence,
- summarization of what the evidence results in,
- recommendation on how to prevent similar future situations,
- optionally suggestions what steps should be taken to improve the current state,
- conclusion of the case.

In the end of the reporting stage the forensic audit is either concluded or a decision on further investigation is made and precise conditions specified. Forensic audit is concluded if the particular crime has been revealed and evidence supporting this is found. Alternatively it might be concluded if all the information about the case suggests that no crime has been committed. Another possibility is when the ordering party for some reason does not want to continue the investigation.

On the contrary if there is still a suspicion that a crime has been committed the investigation might continue after reporting stage. This can happen if a suspicion on a new crime has emerged. The investigation also continue if it is expected that the evidence supporting the crime might be found soon if given more effort. If decided so, the investigation might continue from any of the previous steps forwards.

## 2.6 Work following after a forensic audit

When the process of forensic audit ends both the ordering party and the audit company do a follow-up work. The ordering party usually uses the results of forensic audit for their own benefit. They can use the evidence gathered during the forensic audit in court of justice. They can also use it to prevent similar scenarios from happening or eliminate the source of this behavior. This can mean to fire an untrustworthy employee or change the rules or internal policies.

Follow-up work in the audit company applies mainly to extension of experience and expert's opinions in the audit company. Such company usually has more than one team and works on multiple cases at the same time. Improving their knowledge is an important part of their development. On the other hand the audit company must respect confidentiality of the information about the case and guarantee its unconditional anonymity.

# Chapter 3

# System Requirements

## 3.1 Motivation

In the end of most kinds of investigation there needs to be an examiner, an auditor or generally an intelligent person, who would understand what happened in the case. Their aim is to ascertain who is responsible for it. In this situation the examiner usually already has some outputs from some software. However, the situation may be quite complex and the examiner may be using more than one piece of software in this stage. The inspected period of time can be also relatively long. Altogether it can cause difficulties, make the task harder and as a result prolong the time of investigation. Our application should help forensic auditors, to see what was happening in the case during the examined period of time. This application should provide the possibility to see the big picture and so help with the investigation.

The end of the investigation is also specific in the aspect that there are not so many subjects that play the main role in the case. Usually the most important participants are already known. This application is ideal for a situation where there are approximately five to fifteen subjects. It is possible to process the database where there is much higher number of subjects. Examines, e.g. users can adjust which subjects they want to track. This application can be useful when a fraud is being uncovered, when some of the subjects pursue a corruption practice and also in other similar scenarios of a similar kind.

The system should allow the auditor to replay and visually show what was happening during the investigated period.

## 3.2 Graphical User Interface

We have decided that the scheme of graphical user interface of our program will be as shown in figure 3.1. The main and most important part of the area of the application should be the middle frame. This frame should contain one dot for each subject of our case. These dots will be in a circle by default, however, the user should

be able to move them freely in the frame area. The visualization should also be later replayed in this place. This visualization is the bigger picture that the application should provide to the user.

As shown in figure 3.1 there should be a time axis under this panel with a slider and play/stop button. After clicking the play/stop button, the main action of this application should start. In the middle frame the dots representing the subjects should each display an activity every time the subject is involved in some event. The application should process activities which are mentioned and explained later on.

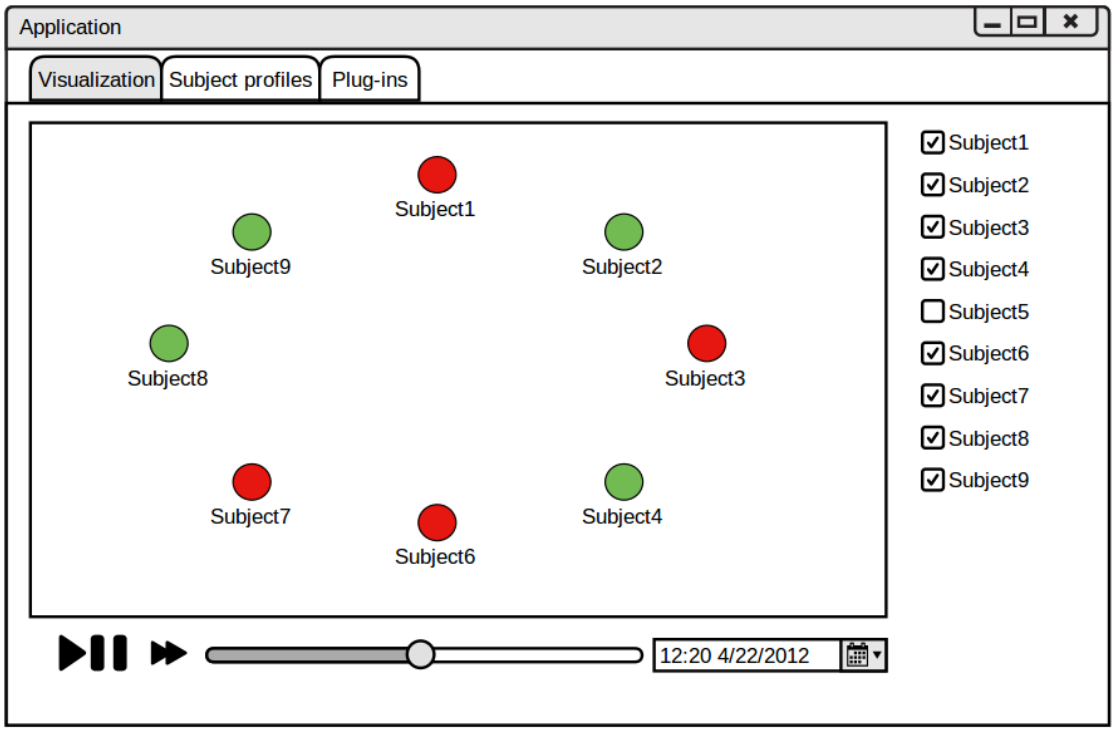



Figure 3.1: GUI Scheme

## 3.3 Activities of the application

**Selecting subjects** Users should be able to select whatever subset of actions and subjects they want to replay in the animation of the case. After uploading the data, the application should generate names of all subjects and next to each one of them a check-box widget. Users should have the possibility to choose from them. Generally it would also be good to provide a possibility to choose a subset of all events directly from our database using a SQL query.

**Adjusting the time period** Under the animation frame there should be a scroll bar and a field indicating the corresponding time and date. Users should be able to

move the scroll the bar below the animation frame manually and this way to travel in investigated period of time. At each point the corresponding scene should appear in the animation frame. The time should also be adjustable manually in the field that indicates corresponding time and date. After clicking the Play/Stop button replaying the visualization from this time further should start.

**Visualization** After clicking on the button play/stop the animation of the scene should start replaying. The program will go through the database of events that the subjects did and for each event a spot representing the subject will change its color for a constant time period. If user stops (pauses) this playback and then starts it again the animation should continue from where it stopped. There should be a possibility to adjust the speed of replaying the case.

**Modes of replaying** The application should provide two modes of replaying. First mode should be replaying according to the real time. Between every two events there should be the same proportion of time with regard to the time in reality. The purpose of this mode is to give the forensic auditor the idea of the time flow in the case. Second mode would serve as a quick replay of events. The period of time that has passed between every two events should be represented by a constant period of time in the visualization. The speed of time flow while replaying should be adjustable in both modes.

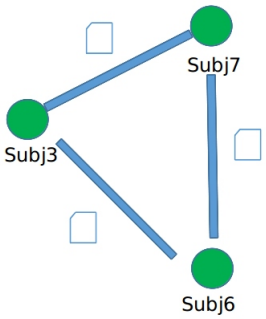
**Subject activities** The application should distinguish between several types of subject activities as shown in following overview.

Activity: There is currently no Event involving this Subject.
Action: The color of the dot representing the subject is red (Fig.3.3)


Activity: Subject is doing something e.g. subject is causing an event
Example: Subject4 is buying a new yacht.
Action: The color of the dot representing the subject changes from red (no action) to green (indicating activity). After a globally adjustable time period the event the color of the dot changes back to red. If we have details about this event a small icon appears above the dot revealing details of the event.

Activity: Subject is causing an event related to another subject.
Example: Subject1 transfers some amount of money to Subject2's bank account.
Action: The dot of the active subject changes color to green and a arrow from the active subject to the related one appears. After a globally adjustable time period the arrow disappears and the color of the dot changes back to red. If we have details about this event a small icon appears above the arrow. After clicking on it the icon reveals details of the event.


Activity: Several subjects are part of a event together.
Example: Subject3, Subject6 and Subject7 have a business meeting together.
Action: All dots representing corresponding subjects change color and a line connecting them together appears. After a globally adjustable time period all the dots are in default state (red color, no linking). If we have details about this event a small icon appears above the arrow. After clicking on it the icon reveals details of the event.




## 3.4 Data

The application should process data from very different sources. Since the application shall be used in last stages of forensic audit we expect that the data might be pre-selected by some quantitative or analytical software. Even though the application should not be limited by the amount of data, average usage is expected to work with the order of hundreds of events. Visualization of more events may not have the effect to help understand the situation. However, the application should be able to hold much more data in the database so that the forensic auditor could select form them.

Since there are many sources of data of different structure in forensic audit, our application is expected work as a superstructure with the purpose of the data projection only. The application should use data inserted to the application by various plug-ins. These plug-ins should transform the structure of data stored in the source (for example, instant messenger, email, accounting system, data from

social networks etc.) to the data format used in our application. This two-layer division is chosen as it is easily expandable and universal. This project describes a standard interface for communication with these plug-ins.

Even though the implementation is not part of this project and the aim of this project is only to design the application, the design should be precise enough to make the implementation easy.

## 3.5 Example case

The action of this application may be represented in following example case. Let us presume there is a company named "XYZ" where some money is missing. However, we are aware of those specific employees, of that certain enterprise, who have direct access to the company accounts. We also know the certain managers that are responsible for the company investments. Moreover, there are business partners of the company and also some regular employees such as workers. This case of money being disappeared has occurred during the period of 6 months.

Forensic auditors receive the data from internal email, IM, telephone communication, from internal accounting system and is aware of the business operations of the company. They have also investigated information on the Internet on social sites. After accumulating all the data they use some analytical or quantitative software 1.3 to discover the abnormalities in the accounting evidence. The rec SW recognize various connections between some of the subjects (common after work activities, common interests etc.). They have enough input, but still they need to find out what exactly happened.

Forensic auditors decide to make use of our application for the investigative procedure. Firstly, they load all the outputs to the application and then they replay the events that happened during the last 6 months. Our system gives them an opportunity to deselect some subjects that seem to be irrelevant. They can also select the set of information for replaying via SQL query, i.e. the forensic auditor may arbitrarily replay those specific events that they demand.

# Chapter 4

# Design of an information system for support of forensic audit

This chapter deals with the design of the information system. Approaches to the implementation of requirements on the system stated in 3 are described here. We also provide a discussion about technologies useful for implementation and the description is accompanied with diagrams demonstrating the design.

## 4.1 Application

The architecture of our application will be based on two fundamental parts. One of them will be collecting the data we will use and the other will display them to the user. From now on we will call this the *Collector* and the *Displayer*. This is a useful division because it simplifies the design and also keeps it easier to maintain. The aim of the *Collector* will be to transform the data from various sources to some generic format for easier manipulation.

There are numerous types of sources of data that can be used as evidence in forensic audit. The data can differ in inner structure and also in the content. Examples of this diversity are IM messages, pictures, GPS coordinates, audio-video files,emails logs of network activities, public information from social sites and internal company management or accounting system.

Ideally our system is able to work with all of them, however we exclude pictures and the audio-video material from these sources because this material does not need visualization. Nevertheless, our system needs to be able to process the information this type of material contains, in case that it could be converted to the basic structure of our application, which will be described later. The task to provide this functionality is assigned to the *Collector*.

Each of the previously mentioned type of source of data is somehow specific. Thus we need a specific solution for each one of them. This is an ideal job for a plug-in architecture.

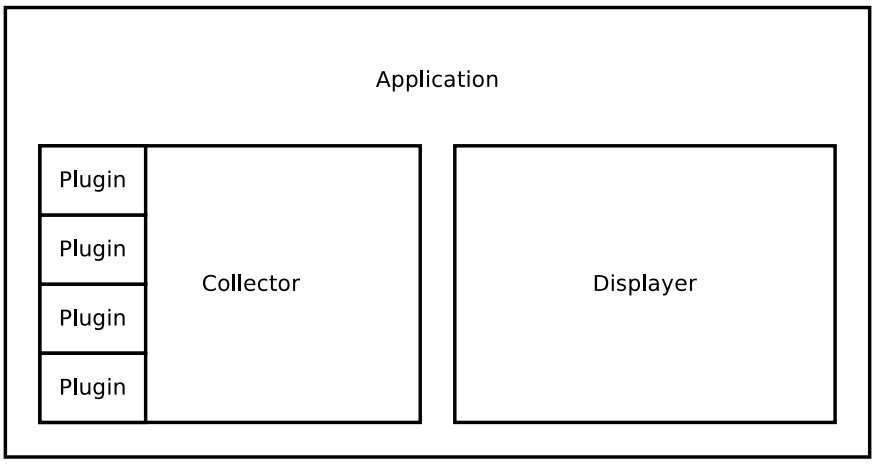



Figure 4.1: Architecture of the application

Plug-in is a piece of software that does not work separately, it only extends the functionality of the application. The application usually provides an Application Programming Interface (API) so that the plug-in extension could communicate with the application properly. In case of this application it means that for each source of data there will be a specific plug-in. each plug-in will be made specifically for the format of the source file. The purpose of the plug-in will be to read the source files and prepare the data for the *Collector*. The collector than saves it to the database. Finally the *Displayer* will access the database and provide the visualization of the case.

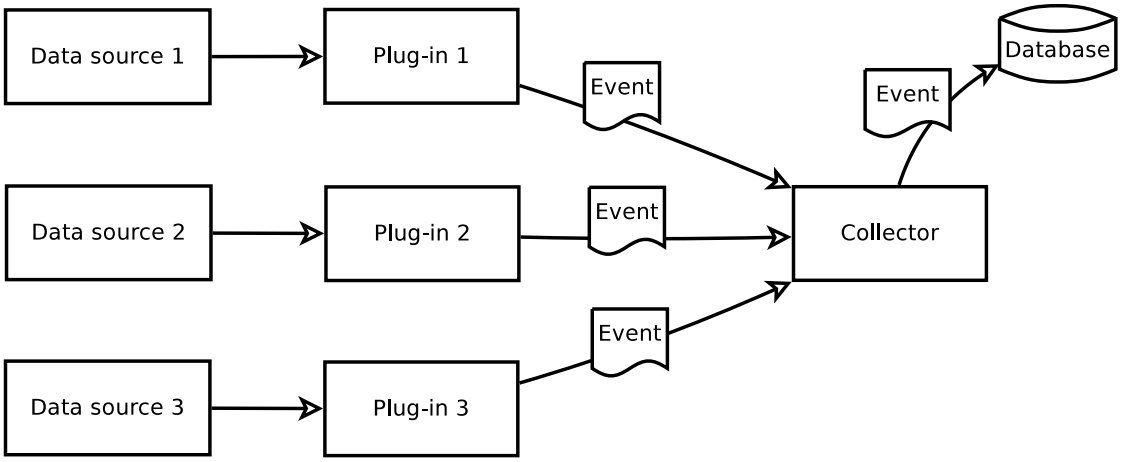



Figure 4.2: Collector

### 4.1.1 Database

**Entity *Event*** The database model of our application is based on a entity called *Event* and several other entities. The Primary unique identification number (UID) of *Event* is Integer ID. It also consists of one mandatory time-stamp that indicates the time this event occurred or the time of the start if the event was long-lasting. We

need to distinguish between one-time events and longer lasting ones so in there is a mandatory boolean attribute *LongLasting* playing the role of a flag. If this attribute is true, another time-stamp attribute indicating the end of the *Event* is to be filled in.

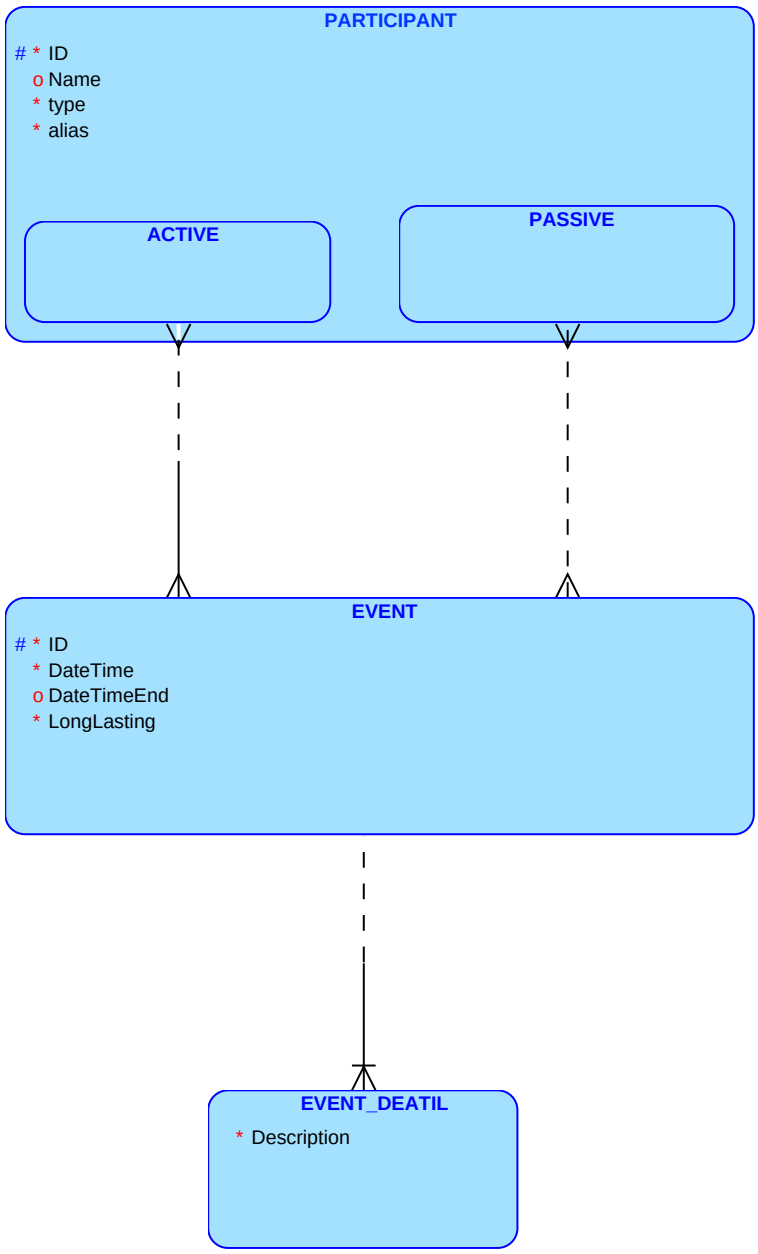



Figure 4.3: Logical model of the database

**Entity *Event_Detail*** We have prepared a entity *Event_Detail* in case there would be some other details concerning the *Event*. This entity is in a 1:n relationship with *Event*. It means, that if we will need, there can be more details concerning

*Event*. *Event_ Detail* has only one mandatory attribute called *Description*, that is prepared for 128 characters of text. For *Event* the relationship with *Event_ Detail* is optional, however, for *Event_ Detail* it is compulsory and the ID of *Event* is its foreign key.

The data we expect to use as a source should also contain information about the subject causing the event. For the subject we have a special model. Before describing it, let us explain why we need it. We cannot be sure what sources and what information concerning the subjects we will get. Our source files should primarily contain logs of various kinds of activities, not only logs of one subject. Because of this we cannot easily connect actual person with events. The identifier of the subject can be of different kind in different sources. For example while email clients are identified by email address, e.g. character stings, IM users are identified by integer ID, nickname or even e-mail address. However this identifier is definitely not the same as the bank account number. Still ne subject may be identified by any number of any of these ways.

For simplicity we decided not to deal with merging these various accounts of one subject in the database as it would increase complexity greatly. This will be done later by the *Displayer*.

***Participant*** In our database model we now establish the *Participant* entity. The primary key of this entity is an integer ID and next attribute is optional a 36-char-long *Name*. This attribute is prepared for the case that the name of the participant would be stated in the source. For actual information about the source we prepare two extra mandatory attributes. The first, called *type*, is designed to inform about the source of data the particular information came from. The second one is called *alias* and it is an array of 32 characters prepared to store the original identifier of the participant.

The *Participant* entity contains two other entities called *Active* and *Passive*. Each one of them has the entity *Participant* as its superstructure. The relationship between *Participant* and *Event* is arranged via these two entities. Both these relationships are of m:n type. The only difference between them is that for the *Event* entity the relationship with *Active Participant* is compulsory whereas with *Passive Participant* is optional.

This model enables *Events* to have more *Participants*, but forces the *Event* to have at least one *Active Participant*.

### 4.1.2 *Collector*

The next section focuses on the *Collector* part of the application. Before we start let us mention that the main purpose of the *Collector* is to provide an interface for inserting data to the database and managing the plug-ins that translate the source format to our format (based on the event structure). The *Collector* has its own GUI that enables the user to choose from existing plug-ins or add new ones. The

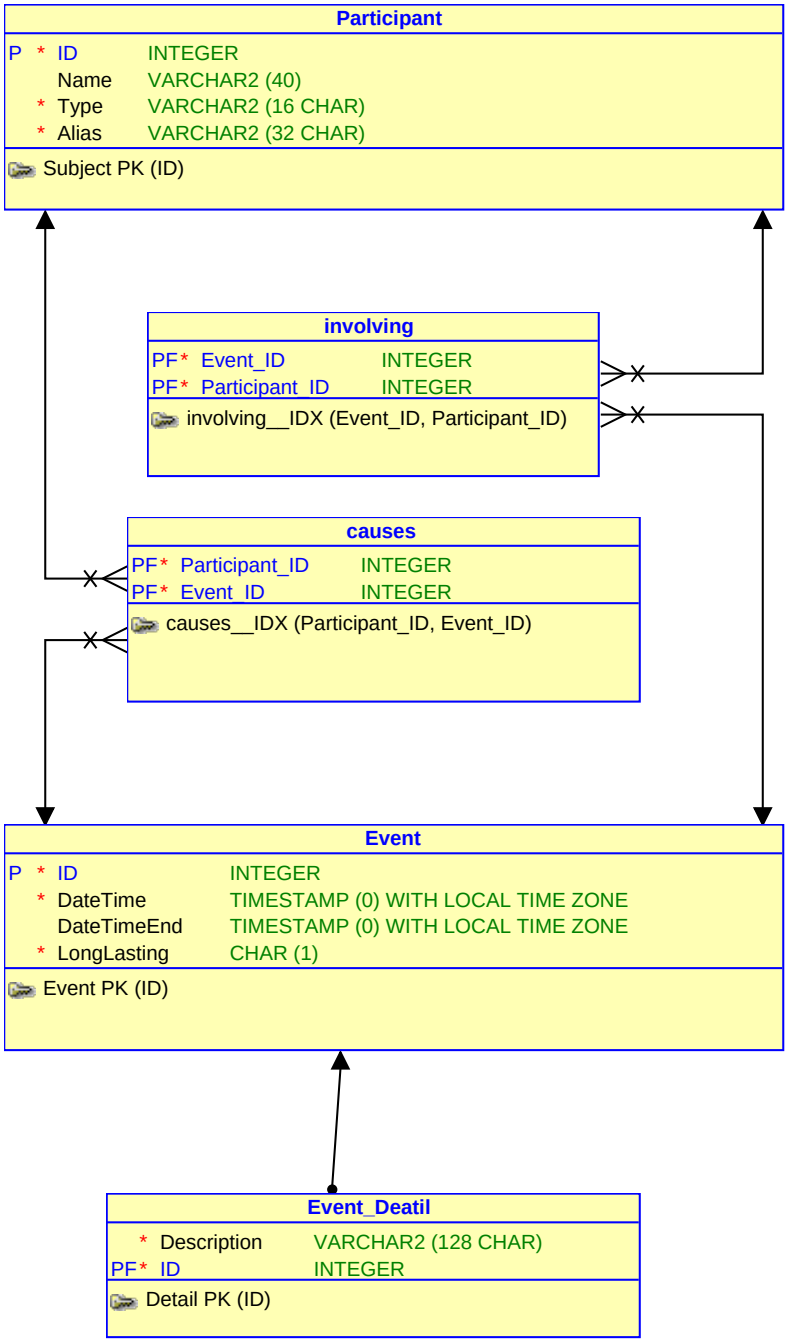



Figure 4.4: Relational model

GUI saves the configuration to a configuration file and then hands it over to the *Collector* back end. The *Collector* provides the user the possibility of selecting and running the plug-ins. Each plug-in has its own GUI so that the user can specify any information needed for them to run of the plug-in (such as the path to a file with the source material).

The *Collector* also manages the plug-ins added to the application. It means that it has a configuration file where all paths to the plug-ins are saved together with short clear description of the source.

The key API that the *Collector* provides to the plug-in is basically the access to the database. Plug-ins load data into the tables and the *Collector* pushes the data into the database. Each plug-in only processes the source file to the data model, create a SQL query and fill in the database. Note, that this application is expected to process data of a text character only. The application is not prepared for mining data from files of multimedia character unless the content is described in a structured text containing the date and some information identifying the related subject.

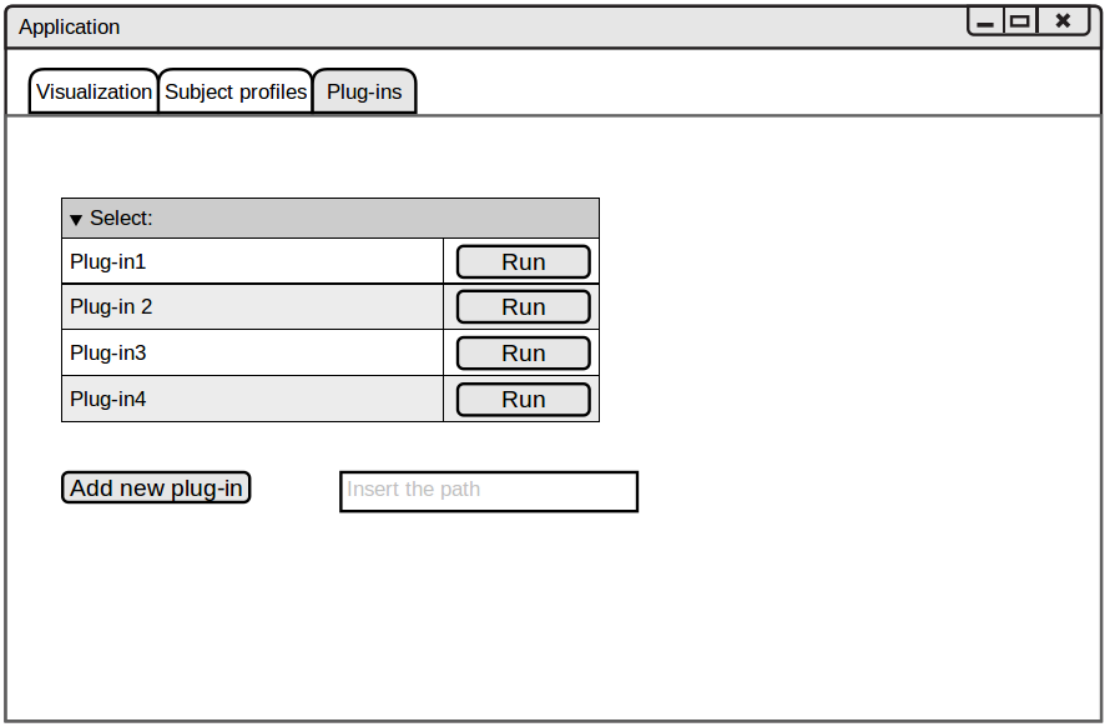



Figure 4.5: Graphical user interface of the plug-in manager

### 4.1.3 *Displayer*

The main purpose of the displayer is to visualize the data saved in the database. To be able to do this, it is necessary at first to deal with the problem of multiple *Participants* in the database that all represent the same *Subject*. We can overcame this obstacle easily by creating a table for *Participant* unification to *Subjects*. However, because of the essence of the data we are working with and the general purpose of our application, we cannot try to guess which *Participants* belong together. Therefore there is a special form helping the auditor unify them to subjects.

Sample form is shown in the figure 4.6. In the left table there is a list of existing subjects. The user can add new subjects easily by filling in the name and surname and by clicking on the "Create subject" button. Auditor can also remove a subject by clicking on "Remove subject" button or edit name or surname of subject by clicking into the table. The *Displayer* creates a table containing the information

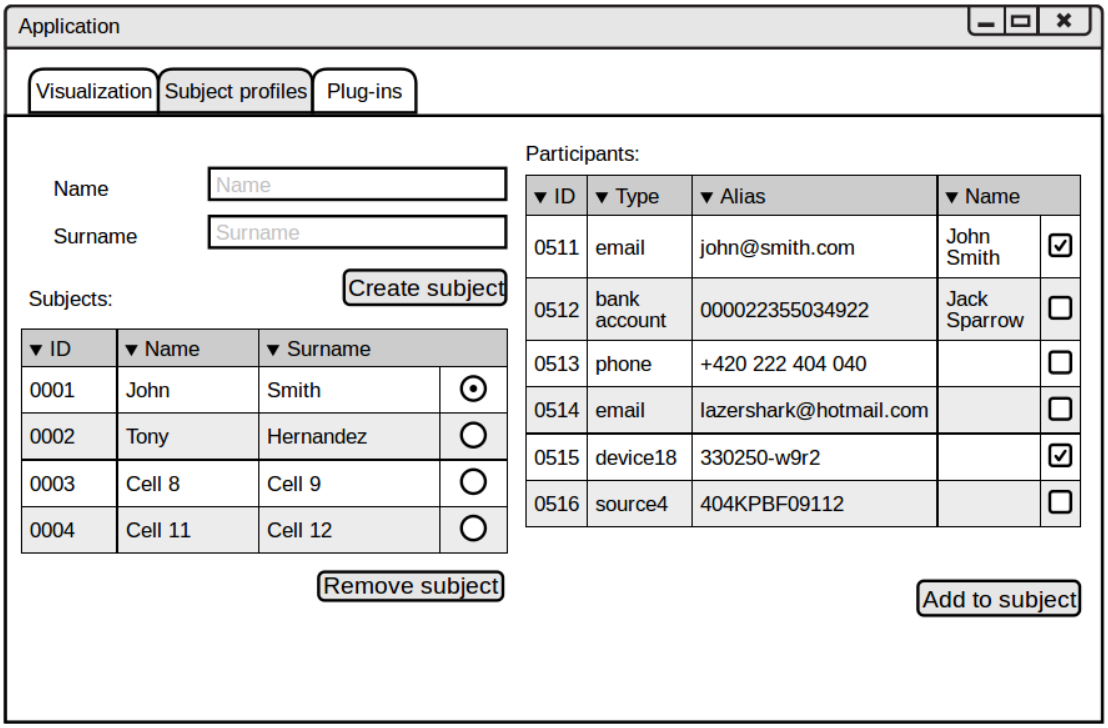



Figure 4.6: Form for unifying *Participants* to Subjects

about this mapping. When replaying the displayer uses this information to identify the correct events for each subject. The data model shown in picture 4.7 indicates that each *Subject* can have many *Participants*, but each *Participant* can have only one *Subject*.

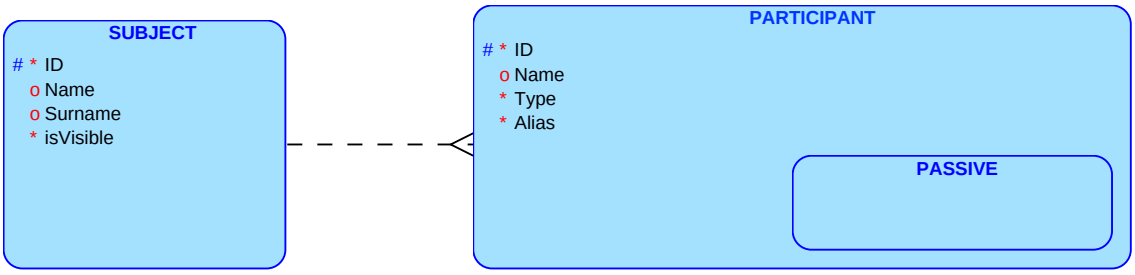



Figure 4.7: The relation between *Subject* and *Participant*

Now we can focus on the most important function of the application. We have a background for the visualization in the above modeled database. In the initialization phase the *Displayer* generates names of all the known *Subjects* and places a checkbox next to each one of them. The state of the checkbox decides whether we want to include the related *Subject* in the visualization.

In the next step the *Displayer* enters proper SQL query. IDs of the subjects are replaced by the entire group of *Participants*. We then search in the database for all *Events* whose starting time is later than the date selected in the visualization tab

and at the same time the some of the *Subjects* takes part in it. We also search for *LongLasting Events* that end after the selected time frame. The result of the selected *Events* is then replayed by time.

The process of replaying has a small obstacle concerning various animations in different cases. These differences are:

1. If *Subject* does not take part in current *Event* it is displayed as a red dot.
2. If *Subject* takes part in a *Event* alone it is displayed as a green dot.
3. If *Subject* causes an *Event* the color of the active *Subject* is green and a arrow aiming at the passive *Subject* is displayed.
4. If more *Subjects* actively participate in one *Event* they are all displayed in green color and a line connecting them together is displayed between them.

The solution for the first and second case is trivial. The color of all *Subjects* is red by default so only if we run into a *Event* where a subject takes a part we change the color for a constant period of displayed time. If the *Event* has a passive *Participant*, as described in third case, we draw an arrow from the active to the passive *Subject*. Similarly, if the *Event* has more *Active Participants* the line between them is drawn.

All these situations are easily recognizable using SQL queries.

# Chapter 5

# Discussion

**Database** We recommend to use the Oracle database. The main reason is that it is one of the fastest and safest relational databases. If it was necessary since it is a relational database it could be replaced for any other relational database and the Oracle Data Modeler is also very easy to use, we have already some good experience with it.

**Open Database Connectivity** Open Database Connectivity (ODBC) is a standardized software API that enables the access to database servers. The aim of ODBC is to provide an access independent on programming language, operating system or the database system. We recommend to use this technique while implementing this application.

We would also recommend to use the C/C++ programming language. The Qt library can be used for implementation of the GUI of this program.

**Further improvement** Follow-up work would definitely include the implementation of the designed application, testing and hands-on experience with real data. Mainly real-world use will show deficiencies, if any, and provide necessary insight for modifications and extensions.

# Bibliography


[1] Asset misappropriation; the perennial leader among economic crimes. http://www.pwc.com/gx/en/services/advisory/consulting/forensics/economic-crime-survey/economic-crimes/asset-misappropriation.html, 2015.

[2] Merriam webster dictionary - audit. http://www.merriam-webster.com/dictionary/audit, 2015.

[3] Merriam webster dictionary - forensic. http://www.merriam-webster.com/dictionary/forensic, 2015.

[4] Chad Albrecht, Mary-Jo Kranacher, and Steve Albrecht. Asset misappropriation research white paper for the institute for fraud prevention. http://www.theifp.org/research-grants/IFP-Whitepaper-5.pdf, 07-Oct-2010.

[5] D. Ariu, G. Giacinto, and F. Roli. Machine learning in computer forensics (and the lessons learned from machine learning in computer security). 2008.

[6] Michael Akers Donald C. Marczewski. Cp as perceptions of the impact of sas. 2005.

[7] Katrin Franke. Digital & computational forensics. Norwegian Information Security Laboratory (NISlab), Gjøvik University College, 2012.

[8] CISA CFE CrFA MBA Mustapha Mugisa. Today's digital forensics & tools. https://www.isaca.org/chapters2/kampala/newsandannouncements/Documents/Introduction%20to%20forensic%20investigation;%20approach%20and%20tools%20available-January%202012.pdf.

[9] Lisa Weaver. Forensic auditing relevant to acca qualification paper p7. *Student Accountant*, Sep 2008.

# Attachment A

## Contents of the CD

### The text of this bachelor project in pdf

The text of this bachelor project is saved as `BP_Peskova.pdf` in the root folder.