OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

INF

FAKULTÄT FÜR
INFORMATIK

Context Modelling for IT Security in Selected Application Scenarios

Dissertation zur Erlangung des akademischen Grades
Doktoringenieur (Dr.-Ing.)

angenommen durch die Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg

von: M.Sc. Kun Qian
geboren am 11. Juni 1984 in Anhui, China

Gutachter:

Prof. Dr.-Ing. Jana Dittmann, Otto-von-Guericke-Universität-Magdeburg, Magdeburg, Deutschland

Prof. Dr.-Ing. Matthias Hemmje, FernUniversität Hagen, Hagen, Deutschland

Prof. Dr.-Ing. Rainer Böhme, M.A., Westfälische Wilhelms-Universität Münster, Münster, Deutschland

Magdeburg, Deutschland
14. Januar 2015

ABSTRACT

Despite the fact that "making considerations in context" or "putting things into context" is something that happens regularly in normal people's daily life, it remains a question for the academia and industry that how to escalate such intuitive way of thinking to a scientific methodology for the introduction of context awareness to benefit the security assurance in automatized processes. Aiming to answer this question, the following three research challenges are derived for this dissertation: 1) How to collect and organise useful information that constitutes security related context? 2) What is the general context modelling methodology that can be used to address various security issues in sophisticated data processing systems? 3) How to evaluate the context models, especially from the security point of view?

Addressing these research challenges, the dissertation focuses on using context modelling to help solve security issues in data processing, where the security requirements are clarified on the system level and realised on the data level. For the first time, a theoretical framework is constructed for the security-oriented context modelling, defining a formalised description of security context and specifying it to guide the contextualisation of the security requirements on both levels. The framework is then further applied on two selected application scenarios, namely digital long-term archiving and forensic dactyloscopy, to develop context models either for a data processing system, or a series of data processing procedures, addressing security requirements on the corresponding levels. These two instantiations justify the applicability of the theoretical framework, and the resulting models show the benefit of introducing the context awareness regarding to the security assurance.

The system level instantiation of security framework for digital long-term archive closes the gap of security concerns that have existed for a long time in the corresponding state of the art, whereas the data level separation approach of overlapped latent fingerprints achieves security improvements in multiple aspects, regarding not only the error rate of the separation result but also the non-repudiation of the whole approach.

Additionally, the dissertation also discusses the evaluation metrics for the context models as well as their application on the two instantiations. It derives scenario-unspecific criteria for the quality assessment and identifies several principles that unite existing scenario-specific evaluation methods for the performance assessment. At last, a general methodology is presented, summarising the modelling process and identifying its nature of being an iterative progress: in case that a changing application scenario introduces new context, the modelling process evolves with the evolvement of the context, so does its evaluation.

Despite the above contributions, there are still unclosed gaps regarding both the theory and the practices, which are beyond the possible coverage of one single dissertation. Therefore, several paths for future work are also identified at the end of this dissertation.

DEUTSCHSPRACHIGE VERSION DES ABSTRACT

Trotz der Tatsache, dass im Alltagsleben unablässig „Überlegungen in einem bestimmten Kontext erfolgen" oder „Dinge oder Angelegenheiten in einen Kontext gestellt werden" bleibt es für die Wissenschaft und Industrie eine Herausforderung eine derartige intuitive Herangehensweise zu einer wissenschaftlichen Methodologie aufzuwerten um Kontextbewusstsein als Mittel zur Gewährleistung von Sicherheit in automatisierten Prozessen einzuführen. Um diese Fragestellung zu beantworten werden in dieser Dissertation daher drei Forschungsaufgaben angegangen: 1) Wie können nützliche Informationen, die zu einem sicherheitsbezogenen Kontext beitragen, eingesammelt und organisiert werden? 2) Welches ist eine allgemein nutzbare Methodologie für die Kontextmodellierung mit der unterschiedliche Sicherheitsfragen in komplexen Datenverarbeitungssystemen adressiert werden können? 3) Wie können Kontextmodelle evaluiert werden, insbesondere im Hinblick auf Sicherheit?

Um diese Forschungsaufgaben zu adressieren wird sich diese Dissertation auf den Einsatz von Kontextmodellierung konzentrieren um Sicherheitsfragen in der Datenverarbeitung zu lösen, bei denen die Sicherheitsanforderungen erst auf der Systemebene geklärt und dann auf der Datenebene umgesetzt werden. Zum ersten Mal wird dabei ein theoretisches Rahmenwerk für sicherheitsorientierte Kontextmodellierung geschaffen, das eine formalisierte Beschreibung des Sicherheitskontexts definiert, als auch eine Anleitung wie dies zur Kontextualisierung der Sicherheitsanforderungen auf beiden Ebenen genutzt werden kann. Das Rahmenwerk wird weiterhin auf zwei ausgewählte Anwendungsszenarien angewandt – digitale Langzeitarchivierung und forensische Daktyloskopie – um Kontextmodelle für ein Datenverarbeitungssystem als auch Datenverarbeitungsprozeduren zu entwickeln, die Sicherheitsanforderungen auf den entsprechenden Ebenen adressieren. Beide Anwendungsfälle zeigen die Anwendbarkeit und Vorteile des theoretischen Rahmenwerks zur Gewährleistung von Sicherheit durch Kontextbewusstsein auf.

Die Anwendung des Sicherheitsrahmenwerks auf Systemebene für digitale Langzeitarchive schließt eine Lücke von Sicherheitsbedenken die seit langem im Stand der Technik existiert, während der Ansatz zur Separierung von überlagerten latenten Fingerabdrücken auf der Datenebene Sicherheitsverbesserungen in mehreren Aspekten bringt, so bei den Fehlerraten der Separierungsergebnisse oder bei der Nichtabstreitbarkeit des gesamten Ansatzes.

Zusätzlich diskutiert die Dissertation Evaluationsmetriken für Kontextmodelle als auch deren Anwendbarkeit auf beide Anwendungsfälle. Dabei werden universelle Kriterien zur Qualitätsbestimmung hergeleitet, sowie mehrere Prinzipien identifiziert um existierende anwendungsspezifische Evaluationsmethoden zur Performanzbestimmung zu vereinen. Abschließend wird eine allgemeine Methodologie vorgestellt um den Modellierungsprozess zusammenzufassen und dessen Natur als iterativen Prozess zu identifizieren: sollte ein sich veränderndes Anwendungsszenario einen neuen Kontext einführen, so entwickelt sich der Modellierungsprozess wie auch dessen Evaluierung mit der Entwicklung des Kontexts mit.

Trotz der zuvor genannten Beiträge werden dennoch Lücken in Theorie und Praxis verbleiben, die außerhalb des möglichen Rahmens einer einzelnen Dissertation liegen. Daher werden am Ende dieser Arbeit mehrere Anknüpfungspunkte für zukünftige Arbeiten aufgezeigt.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

5-MTN	5-Methlthininhydrin
AFIS	Automated Fingerprint Identification System
AIP	Archival Information Package
ASP	abstract security policy
BMBF	Federal Ministry of Education and Science (*Bundesministerium für Bildung und Forschung*)
CASPAR	Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval
CCD	Charge-coupled device
CCSDS	Consultative Committee for Space Data Systems
CC/PP	Composite Capabilities/Preference Profiles
CE	context entity
CRL	Center for Research Libraries
CWL	Chromatic White Light
DCC	Digital Curation Centre
DEA	Drug Enforcement Administration
DEUS	Digital Enclosed Ultraviolet imaging System
DFO	1,8-Diazafluoren-9-one
DFT	Discrete Fourier Transformation
DPE	DigitalPreservationEurope
DPPN	Deutsches Pflanzen Phänotypisierungsnetzwerk
DRAMBORA	Digital Repository Audit Method Based on Risk Assessment
ECA	Event-Condition-Action
EER	Equal Error Rate
EPPN	European Plant Phenotyping Network
FAR	False Acceptance Rate
FBI	Federal Bureau of Investigation
FRE	Federal Rules of Evidence
FRR	False Rejection Rate

FTIR	Fourier transformation infrared spectrometry
GC-MS	Gas Chromatography–Mass Spectrometry
GCE	general context entity
HDF	Hierarchical Data Format
ICA	Independent Component Analysis
ILM	Information Lifecycle Model
IR	Infrared
iRODS	i Rule Oriented Data System
ISO	International Organization for Standardization
IT	information technology
LED	Light Emitting Diode
LRDP	Local Rule Decision Point
MINDTCT	minutiae detector
NASA	National Aeronautics and Space Administration
NBIS	NIST Biometric Image Software
NIST	National Institute of Standard and Technology
nm	nanometre
OAIS	Open Archival Information System
OCLC	Online Computer Library Center
OSI	Open Systems Interconnection
PCM	Path Context Model
ppi	pixel per inch
RDP	Rule Decision Point
REP	Rule Enforcement Point
ROC	Receiver Operating Characteristic
RUVIS	Reflected UltraViolet Imaging System
SCE	security context entity
SERS	Surface-Enhanced Raman Spectroscopy
SHAMAN	Sustaining Heritage Access through Multivalent AchiviNg
SIP	Submission Information Package
SPIE	Society of Photo-Optical Instrumentation Engineers
SR	security rule
SSP	specified security policy
TAR	True Acceptance Rate

TE	target entity
TRAC	Trustworthy Repositories Audit & Certification: Criteria and Checklist
UVC	Ultraviolet C
VMD	vacuum metal deposition
WORM	write-once read-many
XML	Extensible Markup Language
X-RBAC	XML-based Role Based Access Control

ACKNOWLEDGEMENTS

I would like to express my very great appreciation to my supervisor, Prof. Jana Dittmann, for giving me the opportunity to pursue my doctorate in the field of IT security, which I have always been interested in and passionate about. I also thank her for all the joint work with her in the research projects SHAMAN, DigiDak, and DigiDak+, which to a large scale has directly contributed to this dissertation.

Furthermore, I would like to thank Prof. Matthias Hemmje for hosting me to in his PhD students' seminars, where I have engaged in many inspiring discussions. I would like to thank Dr. Chrisitan Krätzer and Maik Schott, for our joint work in project SHAMAN and those discussions which sparkles on the development of my modelling theory. I would also like to thank Dr. Ronny Merkel, Dr. Andrey Makrushin, and Mario Hildebrandt, for our joint work in project DigiDak and DigiDak+, from which this dissertation have also benefited a lot. My grateful thanks are also extended to Prof. Claus Vielhauer, Silke Reifgerste, Tobias Hoppe, Stefan Kiltz, Stefan Kirst, Eric Clausing, Christian Arndt, Robert Fischer, Tobias Scheidat, for all their kind support in various ways for my work in Otto von Guericke University Magdeburg and Brandenburg University of Applied Sciences, and my student, Wenju Zheng, for his contribution regarding the realisation of separation algorithm of overlapped latent fingerprints. Special thanks also go to all my friends who have kept me distracted, in the most positive way. Here is a very probably incomplete list of them: Monique Vorsprach and her lovely family, Yu Sun and his wife Wenyuan Liu, Yikai Wu, Kai Yan, Zhi Jiang, Wenzehan Liu, Xi Li, Shi Du, Qi Chuan, and Chun-ting Hsu.

At last, I wish to thank my family, especially my parents, for their endless support and encouragement ever since I decided to come to Germany seven years ago. I could have never come this far without them backing me up.

1 Motivation and introduction

Just like architects use their blueprints for building skyscrapers, for decades computer scientists and IT specialists have been applying *modelling* for guiding and assisting the construction of sophisticated data processing systems. However, just like error and negligence in the blueprints could be fatal for the skyscrapers, error and negligence in scientific models could also fail the resulting system.

In early 1990s the National Aeronautics and Space Administration (NASA) experienced an awkward and frustrating digital dark age. The scientists there realised that huge amounts of data collected during various space projects in the 1960s could not be accessed, as they not only were poorly archived but also stored in obsolete formats and required obsolete machinery [Blakeslee1990]. Despite the fact that NASA has learned its lesson and adopted HDF5 since 1993 as the standard file format for storing data collected in its projects to avoid similar problems [FC2004], it was still unavoidable for NASA to have to allocate lots of extra manpower and funds to sort out the old data and retrieve useful information.

In August 2012 the charge was dismissed against a fugitive doctor involved in USA's largest prosecution of pharmacies, as the volume of evidence posed a huge burden for the case staying open. The two terabytes electronic evidence confiscated from the doctor's IT systems took up 5% of the worldwide electronic storage available at this time to the Drug Enforcement Administration (DEA), and they were accompanied by several hundred boxes of paper containing 440,000 documents to be digitised, plus dozens of computers and servers [Pfeiffer2012]. As DEA's evidence archive was never designed for such huge amount of data, its capacities and procedures showed severe limitations to adequately process the evidence of this case [Nye2012].

These two exemplary events are both typical examples that large-scale IT systems fail in processing their data, and the reason behind both system failure events can be fundamentally induced as the application of inadequate modelling, i.e. the modelling applied fails to recognise and adapt for the evolved *context*, which can be loosely understood here as any relevant information from the discussed scenario. A further specified and formalised definition of *context* is provided in subsection 2.1.2.

Table 1 describes the inadequate modelling in both events by analysing and comparing the modelled objects, the evolved context that the modelling failed to address, and the consequence that the failure caused. In the first event, NASA's data archive failed to handle the data obsolescence situation, and it caused the security compromise in the aspect of *availability*, i.e. the resources can no longer be used or accessed even by authorised entities. Although NASA took corresponding measures to alleviate future problems, the already caused loss was unfortunately still lost and can only be gradually recovered by slow, inefficient and costly manual work. As for the second event, there is a deeper reasoning than the simply lack of storage space. In fact it is hard to believe that even in the year of 2012 two terabytes of electronic data would still cause so much trouble for law enforcement. As Eric Pfeiffer wrote in his article that covered this dramatic incident [Pfeiffer2012], "an external hard drive with a terabyte of storage can be easily purchased online at outlets like Best Buy for around $100." This narrative might be true, however what it implies is actually not feasible: any law enforcement agency shall not simply purchase external hard drives to store and preserve electronic evidence, due to the rigorous legal and technical requirements on how the evidence shall be properly and securely processed. It is clear that the infrastructure of the DEA evidence archive, including not only its capacities but also specific processing procedures, was not capable of handling the large amount of data with sufficient security assurance. Therefore, the security compromise in this event lies not only in the aspect of availability, but at least also in the aspects of *authenticity* and *integrity*, and as the security compromise in those aspects could be fatal for the processing of evidence, it eventually caused the case dismissal in 2012. The definitions of the security aspects mentioned in this paragraph, e.g. availability, authenticity, and integrity, are provided in subsection 2.3.1.

	Modelled object	***Evolved context failed to be addressed***	***Consequence***
Exemplary event 1	NASA data archive	Substandard data formats and required processing environments which become obsolete over time	Security compromise causing extra manpower and funds
Exemplary event 2	DEA evidence archiving capacities and procedures developed in 80s and 90s	Complications caused by huge amount of electronic evidence	Security compromise leading to case dismissal

Table 1. Analysis of the inadequate modelling in both events

To sum up, as revealed by Table 1, both events would not have happened if the modelling approaches involved had *context* awareness, i.e. if *context modelling* were applied to address evolving context in the first place. The preceding mentioned concepts, namely *modelling*, *context* and *context modelling*, together with various security aspects and goals, are further described in Chapter 2.

Nevertheless, we are indeed living a digital era, which certainly leads to an exponentially growing amount of electronic data. The lesson learned from NASA is that context awareness needs to be considered as early as possible in system modelling, if not from the ever beginning. However, the dismissed prosecution against the fugitive doctor reflects the salient lack of methods of introducing context awareness for IT security in specific application scenarios, e.g. the need of processing large amounts of data in secure manners. The potential solution for the security compromises in these two events is further presented in Chapter 4.

The rest of this chapter is organised as follows: based on the preceding introduction, section 1.1 specifies the problems that this dissertation tackles, and then section 1.2 introduces the approach that this dissertation applies and goals that it addresses, after that section 1.3 describes the scientific contribution of this dissertation, at last section 1.4 outlines the rest of the dissertation.

1.1 Problem statement

Figure 1 shows a generalised data processing system – the system takes in data as input, processes it and then outputs the processed data. Data processing on any scale, from the simplest data processing procedure to the most sophisticated data processing system, including the systems involved in preceding introduced events, can always be generalised to this illustration.

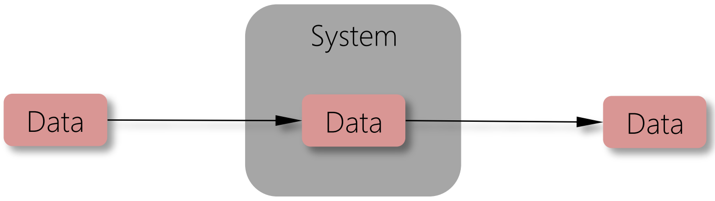


Figure 1. A general data processing system with input and output

In real world application scenarios, it is common that a sophisticated data processing system complicates its security requirements. Take the event with the failing DEA evidence archive introduced earlier as an example: when the processed data is specified as (potential) digital evidence and the system as an evidence archive, not only would various security requirements be raised accordingly for the processed data itself, the system in general should also be capable of handling evolved context with security assurance, e.g. addressing new types of evidence data, upgraded security levels, or simply huge amounts of data like in the introduced event.

Therefore, the general problem that this dissertation addresses is **how to introduce context awareness into the modelling methodology for sophisticated data processing systems to meet various security requirements**. To sufficiently address such general problem, the dissertation focuses on resolving the following general research challenges:

1) **How to collect and organise useful information that constitutes security related context?** In specific, how does the application scenario influence the handling of its information? What information should be considered as relevant hence context? How to formalise and interpret different types of context? How to reveal and meet the security requirements from the application scenario in it?

2) **What is the general context modelling methodology that can be used to address various security issues in sophisticated data processing systems?** In specific, what are the general modelling steps towards security context awareness? How to specify these modelling steps in various application scenarios? What happens to the modelling steps if the application scenario tends to be dynamic, instead of static?

3) **How to evaluate the context models, especially from the security point of view?** In specific, what specific properties that the context models should be evaluated on? How to generate evaluation metrics to assess these properties? What is the relationship between the application scenario and the evaluation metrics? How to apply such evaluation metrics?

1.2 RESEARCH OBJECTIVES

To resolve the research challenges identified in section 1.1, it is necessary for this dissertation to at least achieve the following series of research objectives:

1) **Clarification and formulation of relevant concepts**: As the first step, relevant concepts need to be clarified and formulated. A thorough study is performed on the general state of the art on modelling, context and context modelling. Furthermore, the general security aspects and goals summarised in the state of the art are also identified. Based on these, the concepts of *security context* and *security-oriented context modelling* are derived, formulated, and specified, to form a theoretical framework. This objective directly addresses research challenges 1) and 2) identified in section 1.1.

2) **Induction towards the constitution of security context and eventually the methodology**: As general methodology is usually derived from specific applications, the state of the art on the two selected application scenarios (digital long-term preservation and digital dactyloscopy) needs to be studied and introduced. On this basis, in both application scenarios, it is demonstrated that how context models can be generated and applied to tackle various security issues. Combined with the derived concepts, the constitution of security context and the general methodology of security-oriented context modelling are further induced. This objective also addresses research challenges 1) and 2).

3) **Proposal and utilisation of evaluation metrics**: To assess security-oriented context models, evaluation metrics need to be proposed and utilised, covering the assessments from following two angles: a) the general quality of a derived model itself, and b) the performance of the derived model regarding its application in the specific application scenario. The metrics on both a) and b) should be compatible with existing evaluation criteria and methods, for the sake of universality and reproducibility. This objective directly addresses research challenge 3).

1.3 Scientific contributions

As the general solution to the problem described in section 1.1 and at the same time the core scientific contribution, this dissertation presents the concept of **security-oriented context modelling**. This concept is introduced, formulated, specified, further applied and eventually assessed; on both system and data levels, which are the two levels of granularity where this dissertation employs to analyse the security concerns of a sophisticated data processing system: On the system level, the infrastructure of the system must be security compliant, i.e. the system components must be designed to function in conformity with the security requirements imposed by security policies. On the data level, the data is processed either within a system component or among multiple ones, so there are also various security requirements for the corresponding processing procedures. Take the example of the digital evidence archive from section 1.1 again: for such system the security requirements can be defined either on the system level, e.g. the evidence must be securely managed, or on the data level, e.g. the integrity of a particular piece of evidence must be verified. Therefore, as illustrated by Figure 2, in this dissertation context modelling is applied respectively on the system and data levels, addressing security issues on both levels.

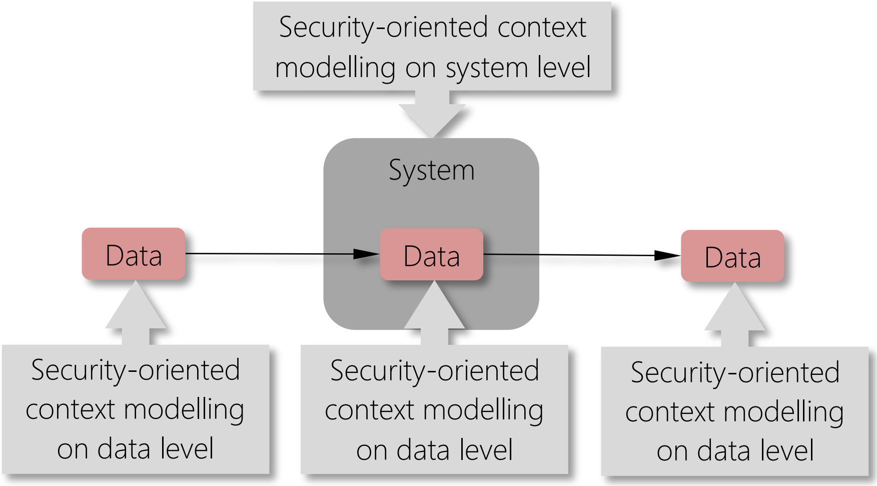

Security-oriented context modelling on system level
System
Data
Data
Data
Security-oriented context modelling on data level
Security-oriented context modelling on data level
Security-oriented context modelling on data level

Figure 2. Context modelling applied on the general data processing system to address system and data level security

Specifically, surrounding the concept of security-oriented context modelling, the scientific contributions that this dissertation makes can be further categorised as primary and secondary ones. The primary contributions include:

1) A **theoretical framework** is developed for the conception of security-oriented context modelling. Within the range of this framework, a series of concepts are derived and formulated, including *scientific modelling*, *general context modelling*, and eventually *security-oriented context modelling*. A series of meta-models are developed for them, acknowledging their key constitutions, especially with emphasis of the concept of *security context*. The

dissertation also investigates how the nature of security context differs on the system and data level, so the framework is further extended correspondingly, forming specific meta-models for security-oriented context modelling on both levels. The framework also comprises a descriptive scheme, which serves not only for the application of the context model, but also together with the other parts of the framework as the basis of the generalised modelling methodology described later.

2) Two **exemplary instantiations** are presented respectively on system and data levels, where digital long-term preservation and digital dactyloscopy are selected respectively as the demonstrative application scenarios. The concepts on both levels within the theoretical framework are projected thus further specified in the application scenarios, showing how the context is collected, interpreted, and organised in both cases. Furthermore, the resulting context models are evaluated using existing methods, reflecting the advantage brought by the introduction of the context awareness. As such, not only these two instantiations serve as exemplary cases for future applications, the applicability of the proposed theoretical framework is also verified.

3) As it is essential to verify the security assurance of the generated context models within their lifetime, this dissertation presents the development of their **evaluation metrics** on two domains: First, a set of general criteria are derived on a general scenario-unspecific domain based on the state of the art, so the quality assessment metrics can be generated and applied. Second, the basic principles of developing scenario-specific evaluation metrics are also summarised, so the performance assessment metrics can be correspondingly derived, encompassing the existing evaluation methods. The dissertation summarises following properties, that the metrics on both domains shall possess: a) the metrics are able to reflect the security level that the models achieve on both system and data levels, e.g. security aspects to cover, security goals to meet, etc.; b) the metrics differ on system and data level, due to the different nature of context on these two levels; c) the metrics are applied regularly through the life time of the context models to handle the evolution of context, i.e. they enable routine evaluation on the context model thus are able to determine when it needs to evolve or even be replaced; d) the metrics can also evolve if needed, to absorb emerging new criteria, including those with no particular emphasis regarding security, so the metrics and even be extended to evaluate context models in general. Furthermore, the dissertation also demonstrates the proposed scenario-unspecific evaluation metrics by applying them on the generated context models from the two instantiation examples, so their evaluation can be completed.

4) Based on the above three primary contributions, a first **generalised methodology** of security-oriented context modelling is developed, summarising the preceding instantiations on system and data levels. The methodology is designed, applied and extended for further application scenarios: on the system level, it can be the guidance for either system modelling towards a future system with context awareness, or introducing context awareness to an already existing system; whereas on the data level it aims at a thorough description on how the relevant data objects is processed with context awareness, serving for particular security requirements. The methodology summarises how the modelling process proceeds from raw information from the application scenario to the final constructed context model together with other contents in its descriptive scheme. More importantly, it points out that under the circumstance of evolving scenario, how the modelling process should also correspondingly evolve, with regard to a simultaneously evolving evaluation process.

Besides the above presented primary contributions, the dissertation also makes the following secondary contributions:

1) As the system level instantiation of security-oriented context modelling, this dissertation develops a **security framework for digital long-term preservation**. Taking into consideration the security-related issues revealed in the OAIS standards [OAIS2002] [OAIS2009], security context is extracted and a "top-down" context modelling approach is applied, yielding a context model as a prototype of such security framework. As the context in the framework is eventually interpreted by policies, a four-level policy hierarchy is proposed for policy management, together with an enhanced solution for security policy generation, implementation and enforcement. The dissertation also illustrates with a specific application example, demonstrating that the proposed framework is appropriate for its application on a security-oriented archival system, which a) can be constructed with reasonable storage solution for huge amounts of data, b) ensures specific security requirements using specific policies, c) manages large amounts of security policies reflecting its high complexity using policy hierarchy, and d) bears context awareness so the system evolves itself with evolving context (e.g. obsolescence issue) by introducing and adopting new policies while abolishing old policies if necessary and at the same time invoking a complete audit trail, therefore can pose as a solution for the salient need of electronic evidence archive. Furthermore, selected part of state-of-the-art assessment framework is also applied on the developed context model, so its security coverage is analysed.

2) As the data level instantiation of security-oriented context modelling, this dissertation develops a **separation approach of overlapped latent fingerprints** for forensic dactyloscopy. The state-of-the-art approaches serving the similar purpose show apparent drawbacks in various aspects, especially regarding the need from the forensic scenario. Therefore, in this dissertation, the development of the approach starts with presenting a series of security rules to summarise the requirements to regulate the approach in the forensic scenario. With the objective to meet the requirements, a "bottom-up" context modelling approach is utilised, with the gradual elimination of redundancy in the context, resulting the context-based separation approach. Following the theoretical framework, the approach identifies and specifies the different types of context in both acquisition and processing environments, yielding the implementation of an enhanced separation algorithm with optimised parameters. The approach is evaluated both subjectively (based on the derived security rules) and objectively (with an investigation of the error rate on generated test sets), showing the advantage introduced by the context awareness.

The following figure illustrates the nexus between the general problem raised from the motivation and tackled by this dissertation, the specific research challenges that this dissertation addresses, the research objectives that it aims at, and the primary/secondary scientific contributions that it achieves.

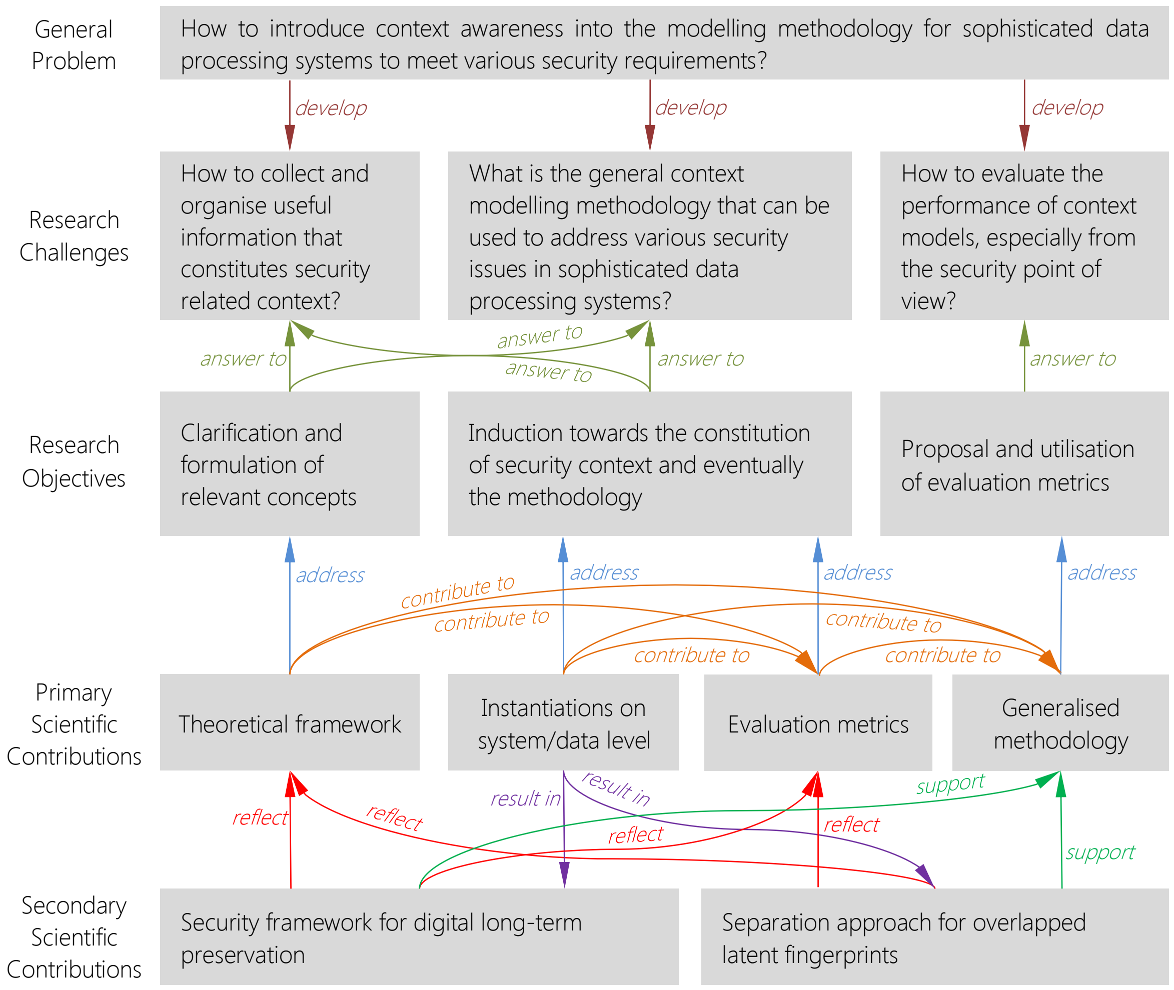

General Problem
How to introduce context awareness into the modelling methodology for sophisticated data processing systems to meet various security requirements?
develop
develop
develop
Research Challenges
How to collect and organise useful information that constitutes security related context?
What is the general context modelling methodology that can be used to address various security issues in sophisticated data processing systems?
How to evaluate the performance of context models, especially from the security point of view?
answer to
answer to
answer to
answer to
answer to
Research Objectives
Clarification and formulation of relevant concepts
Induction towards the constitution of security context and eventually the methodology
Proposal and utilisation of evaluation metrics
address
address
address
address
contribute to
contribute to
contribute to
contribute to
contribute to
Primary Scientific Contributions
Theoretical framework
Instantiations on system/data level
Evaluation metrics
Generalised methodology
reflect
reflect
result in
result in
reflect
reflect
support
support
Secondary Scientific Contributions
Security framework for digital long-term preservation
Separation approach for overlapped latent fingerprints

Figure 3. A summary and overview of the nexus between the general problem, research challenges, research objectives and primary/secondary scientific contributions of this dissertation

1.4 OUTLINE OF THE DISSERTATION

The rest of this dissertation is outlined in the following way:

Chapter 2 introduces the fundamental concepts as well as relevant state of the art. Section 2.1 summarises existing definitions of modelling and context, followed by Section 2.2 the state of the art of context modelling in general. These two sections together serve as the fundamentals of the theoretical framework derived in Chapter 3. Section 2.3 covers the basic concepts of IT security, together with the state of the art relevant to the application of context modelling for IT security. After that, section 2.4 describes the state of the art of the application of context modelling in digital long-term preservation, serving as the fundamentals of the system level instantiation in Chapter 4. Section 2.5 introduces the state of the art of the application of context modelling in digital dactyloscopy, serving as the fundamentals of the data level instantiation in Chapter 5. At last, section 2.6 summarises the whole chapter and analyses the research gaps revealed by the state of the art it introduces.

Chapter 3 describes the theoretical framework that this dissertation proposes. Based on the fundamentals introduced in section 2.1, section 3.1 formulates a general meta-model of context models. Integrated with the security-relevant concepts introduced in section 2.2, section 3.2 takes the derived meta-model and projects it further on both system and data levels. On the system level, a hierarchical structure for security policies is proposed to identify and clarify the security requirements, while on the data level the concept of primary and secondary context in both acquisition and processing environments are defined, yielding the effective execution of security mechanisms. After that, Section 3.3 proposes a descriptive scheme to further approach the application of security-oriented context model. Section 3.4 summarises the whole chapter in the end.

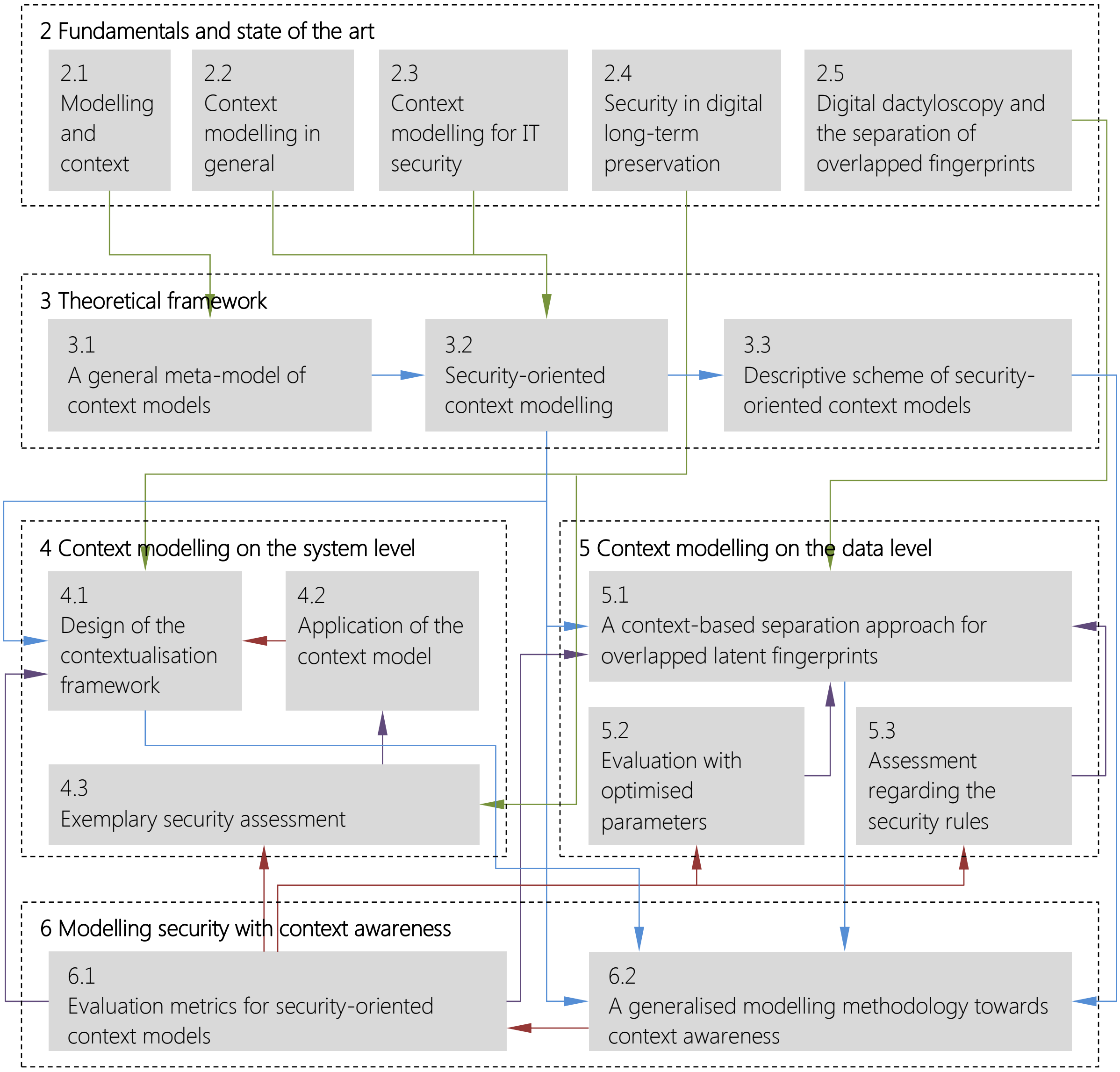
Chapter 4 applies the theoretical framework of security-oriented context modelling proposed in Chapter 3 in the scenario of digital long-term preservation. At the same time, it also serves as a system level instantiation, which contributes to the general methodology summarised in Chapter 6. Section 4.1 designs a contextualisation framework, which applies the theoretical conception to collect, interpret, and organise security context, which eventually functions in the form of security policies on various levels. Section 4.2 provides a specific application of the designed framework to show how it works. Section 4.3 demonstrates an exemplary assessment on the application described in section 4.2. At last, section 4.4 summarises the whole chapter.

Chapter 5 also applies the theoretical framework of security-oriented context modelling introduced in Chapter 3, but in the scenario of digital dactyloscopy. Similar to Chapter 4, it serves as a data level instantiation, which also contributes to the general methodology summarised in Chapter 6. Section 5.1 proposes a context-based separation approach for overlapped latent fingerprints. It contextualises the forensic scenario that the separation is supposed to be conducted, specifies the various types of context defined by the conception, and derive a context-aware algorithm as well as its optimised parameters for the particular non-invasively acquired high-resolution samples to be processed in the scenario. Section 5.2 generates corresponding test sets and applies the evaluation on them for the derived algorithm with the optimised parameters. Section 5.3 further subjectively assesses the security coverage of the derived approach. At last, section 5.4 summarises the chapter.

Chapter 6 discusses the evaluation metrics of the security-oriented context models and summarises the general methodology of security-oriented context modelling. Based on the theoretical framework proposed in Chapter 3, as well as the two instantiations described in Chapter 4 and 5, section 6.1 tackles the evaluation issue from two angles: First, it proposes a series of criteria that constitutes the scenario-unspecific evaluation metrics to assess the quality of context models, so the context models derived in Chapter 4 and 5 can be evaluated accordingly; Second, it also summarises the basic principles for the development of scenario-specific evaluation metrics for the assessment of the performance of context models, with which the evaluations conducted in Chapter 4 and 5 conform. After that, section 6.2 generalises the presented models and modelling approaches and induces a general methodology, following which in further application scenarios proper context models can be generated with security assurance.

Chapter 7 concludes the whole dissertation and discusses the future work. Section 7.1 summarise the dissertation, yielding the conclusions. Section 7.2 analyses the gaps still to be closes within the range of this dissertation and suggests some feasible directions for future works.

The following Figure 4 illustrates the nexus within the main body (i.e. Chapter 2, 3, 4, 5 and Chapter 6, excluding the summary sections, which trivially summaries their own chapters, respectively) of this dissertation for better clarification.


2 Fundamentals and state of the art
2.1 Modelling and context
2.2 Context modelling in general
2.3 Context modelling for IT security
2.4 Security in digital long-term preservation
2.5 Digital dactyloscopy and the separation of overlapped fingerprints
3 Theoretical framework
3.1 A general meta-model of context models
3.2 Security-oriented context modelling
3.3 Descriptive scheme of security-oriented context models
4 Context modelling on the system level
4.1 Design of the contextualisation framework
4.2 Application of the context model
4.3 Exemplary security assessment
5 Context modelling on the data level
5.1 A context-based separation approach for overlapped latent fingerprints
5.2 Evaluation with optimised parameters
5.3 Assessment regarding the security rules
6 Modelling security with context awareness
6.1 Evaluation metrics for security-oriented context models
6.2 A generalised modelling methodology towards context awareness

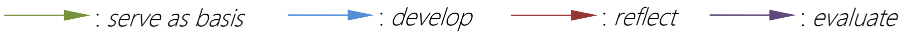


Figure 4. The nexus within the main body of the dissertation

2 Fundamentals and state of the art

This chapter introduces the fundamentals which directly contribute to the theoretical framework derived in Chapter 3 and are reflected by the general methodology and evaluation metrics in Chapter 6. It also introduces related state of the art, based on which the work in Chapter 4 and 5 is developed.

The chapter starts in section 2.1 with a brief introduction on the basic concepts of modelling and context. Then it is followed in section 2.2 by the state of the art on context modelling in general, covering its original motivation, theories as well as existing applications. After that section 2.3 gives the basic concepts regarding IT security and summarises the existing examples of applying context modelling for IT security. As two instantiations will be demonstrated, first in Chapter 4 for digital long-term preservation and then in Chapter 5 for digital dactyloscopy, section 2.4 and 2.5 respectively introduces the state of the art in these two application scenarios. At last, section 2.6 summarises the previous sections to clarify the connection between this chapter and following chapters and identify the gaps revealed in the current state of the art and to be closed in this dissertation.

2.1 Modelling and context

Before digging into the concept of context modelling, it is necessary to briefly clarify the fundamental concepts of *modelling* and *context* in this section.

2.1.1 What is modelling?

The word *model* is originally referred to as three-dimensional representation of a person or thing or a proposed structure and later extended to an abstract description of a system or process to assist calculations and predictions [Oxford2013a]. For hundreds of years models have been pervading in various aspects in our society to represent physical objects and phenomena as well as abstract concepts and theories. As either the activity of making three-dimensional models or devising/use of abstract models, *modelling* approaches usually vary depending on the specific type of the model.

In many scientific disciplines, models are used to explain and predict the behaviour of real objects or systems, and *scientific modelling* is referred to as the generation of *scientific models*, which are physical, conceptual or mathematical representation of real phenomena that usually are difficult to observe directly [Britannica2013a]. Scientists benefit from scientific models by applying them to better understand or operate with those being modelled, which are also known as *modelling targets*, as Frigg and Hartman introduces in [FH2012]. Depending on the nature of the modelling targets, scientific models can represent a selected part of the world (i.e. model of phenomena or data) and/or a theory (i.e. interpretation of the laws and axioms of the theory) [FH2012].

The model of phenomena has been extensively studied in the past decades, addressing mainly two problems. The first problem is the fundamental philosophical explanation of what a model is, scientifically [FH2012], and it is addressed in the recent literature including [Bailer-Jones2003], [Suárez2003], [Giere2004], [Suárez2004], [vanFrassen2004], [Frigg2006], [SS2006], [Contessa2007], [Morrison2009], [Knuuttila2009], [Suárez2009], [Elgin2010], [Frigg2010], [Thomson-Jones2010], [Toon2010], [Toon2011], and [Toon2012]. The second problem is how to general or select the proper representation styles according to specific circumstances [FH2012], and it is discussed in the literature such as [Peirce1931-1958], [Black1962], [Hesse1963], [Achinstein1968], [McMullin1968] [Ackerlof1970], [Hesse1974], [GV1978], [Redhead1980], [vanFrassen1980], [Musgrave1981], [McMullin1985], [Mundy1986], [Giere1988], [Cartwright1989], [Kroes1989], [Laymon1991], [Swoyer1991], [HT1995], [Psillos1995], [Teller2001], [BB2002], [Suppes2002], [DF2003], and [Giere2004]. While the specific content of those studies are not directly connected to the focus of this dissertation, it should be pointed out here that, despite the broadness of the studies in this field, no unanimous systematic account has ever been reached regarding all the proposed different ways in which models can relate to reality and of how these ways compare to each other [FH2012]. Therefore, there is little endeavour towards any potential formalised way of representing such models, and it is similar situation in the studies of the model of data. Despite the concept was proposed a few decades ago (see [Suppes1962]) and there have been enough follow-up studies (see [Laymon1982], [FS1994], [Mayo1996], [Galison1997], [Harris2003], and [Staley2004]), there is no broadly agreed formalisation of such models.

However, in modern logic, it is relatively easier to develop a formalised description of model of theory. As Hodges introduces in [Hodges1997], in this case a theory is taken as a set (which is usually deductively closed) of sentences in a formal language, and the model of such theory is formalised as a logic structure (in the form of 3-tuple) that makes all sentences of it true. Therefore, such logic structure ***S*** is a model in the sense that it is what the theory represents and it is defined as [FH2012]:

$$\boldsymbol{S} = (U, O, R), U \neq \emptyset, R \neq \emptyset \tag{2.1}$$

where 1) a non-empty set U denotes the domain (or universe) of ***S***, 2) an indexed set O, which can be empty, denotes the operations on U, and 3) a non-empty indexed set R denotes the relations on U. If all sentences of a theory are true when its symbols are interpreted as referring to the elements of a structure ***S***, then ***S*** is a model of this theory. The definition given here comes on a highly abstract level, thus nothing matters about what the elements actually are – they are to be extensionally specified.

2.1.2 What is context?

The word *context* is derived from the Latin *con* (with or together) and *texere* (to weave) [Oxford2013b], describing an active process dealing with the way humans weave their experience within their whole environment, to give it meaning [BCQ+2007]. In general, it is referred to as the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood [Oxford2013b]. For decades various definitions of context have been derived from various angles (see e.g. [ST1994], [WJH1997], [RCD+1998], [Pascoe1998], and [SDA1999]). Aiming at applying the concept of context in computing, based on previous definitions, Dey derives the following definition in [Dey2001]:

> *"Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and the applications themselves."*

Based on this, a *context-aware* computing system is defined as a system that "uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task" [Dey2001].

Better clarifying the boundary between an entity, its context and non-relevant information, Lee derives a definition of context in [Lee2011]:

> *"Context is a set of things, factors or attributes that are related to a target entity (TE) in important ways (e.g. operationally, semantically, conceptually, pragmatically) but are not*

closely related to the target entity that they are considered to be exclusively part of the target entity itself."

Based on this definition, Lee further proposes a contextual information framework, in which he classified context into nine general classes [Lee2011]:

- *Object* is a bounded discrete entity that can be characterised as having one or more properties or states, persisting across multiple points in time and place, being uniquely identified, interacting with other objects, and being acted by an agent.
- *Agent* is an entity that can carry out actions.
- *Concept* (or *Abstraction*) refers to ideas or other individually/socially recognised properties or qualities as distinguished from any particular embodiment of the properties/qualities in a physical medium.
- *Time* is a limited stretch or space of continued existence, as the interval between two successive events or acts, or the period through which an action, condition, or state continues.
- *Place* is a designated point or region in space.
- *Relationship* is an association between two or more entities (or classes of entities), which cannot be reduced to or adequately expressed as a property of the entities (or classes of entities) themselves.
- *Occurrence* is a characterisation, for a given span of times and places, of either the state of a set of entities or their interactions. It can be either a simple event or a process, which can be considered as a series of events. Furthermore, it can take the form of either general phenomena, where there is no specific acting entity, or actions, which are carried out by identifiable entities.
- *Form of expression* is a particular way of expressing ideas or information.
- *Purpose* is mandate, norms, values, intention, rules, standards, virtues, or functions to which agents can advance or with which they can conform, attempt to advance or conform, hope to advance or conform, or perceive/expect entities (or set of entities) to advance or conform.

Besides Lee's way of categorisation from a semantic perspective of view, when the concept of context rests in a technical environment in computing systems, there exist other various categorisation schemes. Among those, representative ones are listed here.

Schilit et al. categorise context into three classes in [SAW1994]:

- *Where you are* includes all location related information, e.g. GPS coordinates, common/specific names, specific addresses, user preferences, etc.
- *Who you are with* is information about the people present around the user.
- *What resources are nearby* includes information about resources available in the area where the user is located, e.g. machinery, smart objects, utilities, etc.

Henricksen categorises context into four classes in [Henricksen2003]:

- *Sensed* data is directly sensed from the sensors.
- *Static* information does not change over time.

- *Profiled* information changes over time with a low frequency.
- *Derived* information is computed from primary context.

Perera et al. categorise context into two classes in [PZC+2014]:

- *Primary context* is any information retrieved without using existing context and without performing any kind of sensor data fusion operation.
- *Secondary context* is any information that can be computed using primary context.

A very similar categorisation scheme is also mentioned by Bettini et al. in [BBH+2010], where raw information from physical sensors is referred to as *low-level context* and the information derived from low-level context is called *high-level context*.

Furthermore, instead of context, van Bunningen et al. classify the context categorisation schemes into two classes [BFA2005]:

- *Conceptual categorisation* categorises context based on the meaning and conceptual relationships between the context.
- *Operational categorisation* categorises context based on how they are acquired, modelled and treated.

Therefore, according to the standard from van Bunningen et al., Lee's categorisation scheme is conceptual, while the others' can all be regarded all operational. Nevertheless, while Lee's conceptual semantics based scheme is considered as quite thorough, the operational ones vary in their emphases hence are hard to always accommodate the demands from the computing systems. As shown in Chapter 3, the proposed theoretical framework uses Lee's categorisation as fundaments and formulates it for IT security on system and data levels. On both levels the context is specified on various granularities, from the coarsest conceptual level to the finest operational level, closing the gap between two classes of categorisations.

2.2 Context modelling in general

Serving as the fundamentals of the theoretical framework of this dissertation, this section provides a brief state of the art of context modelling in general.

2.2.1 Why context modelling?

When humans have conversations, they are able to use context (in this case implicit situational information) to increase the understanding of each other. However, this ability does not transform well to the interactions between human and computers, not to say communications between computers, so the flexibility of human language is severely restricted in computer applications. Therefore, the concept of context modelling is proposed for computing, aiming at improving the computer representation and understanding of context to increase the richness of communication in the field [Serrano2012].

Context modelling is introduced to pervasive computing, due to the requirement for its application to be flexible, adaptable, and capable of acting autonomously on behalf of users [BBH+2010], as a context-aware system is able to extract, interpret and use context information and adapt its functionality to the current context of use [BC2004]. Specifically speaking, general context models are developed for pervasive computing applications to [BCQ+2007]:

- adapt interfaces [VTH2006],
- tailor the set of application-relevant data [BCS+2006],

- increase the precision of information retrieval [STZ2005],
- discover services [RRC+2006],
- make the user interaction implicit [PNS+2000],
- build smart environment [DSS+2006].

2.2.2 How to perform context modelling?

Currently pervasive computing is the only field where context modelling is widely and systematically applied. The context modelling approach has evolved with the gradual enrichment understanding of context. Based on how the context is understood as well as how the information is organised and represented, there are several most popular context modelling approaches list as follows:

- *Key-value modelling* uses simple key-value pairs in different formats to define the list of attributes, whose values describe context information used by context-aware applications/systems [BBH+2010] [PZC+2014]. This is for instance applied in [SAW1994], where the key-value pairs act as environment variables.
- *Markup scheme modelling* is an improvement over the key-value modelling approach and utilises a variety of markup languages to organise and represent context information [SL2004] [BBH+2010]. A typical example is the Composite Capabilities/Preference Profiles (CC/PP) developed in [KRW+2004].
- *Graphical modelling* uses graphical components to describe relationships [SL2004] [PZC+2014]. Typical examples include the applications of Unified Modelling Language (UML) in [Bauer2003] and Object-Role Modelling (ORM) in [HIR2003].
- *Object based modelling* applies object-oriented concepts to model data using class hierarchies and relationship [PZC+2014], bringing encapsulation and reusability to pervasive computing environment [SL2004]. A representative of such approach is the Active Object Model developed in [CMD1999].
- *Logic based modelling* consequently defines context as facts, expression and rules [SL2004], therefore provides more expressive richness compared to previous approaches and makes reasoning possible up to a certain level [PZC+2014]. An example is the Sensed Context Model proposed in [GS2001], where first-order predicate logic is used as a formal representation of contextual propositions and relations.
- *Ontology based modelling* use semantic technologies to organise context with ontologies [PZC+2014], which are essentially descriptions of concepts and their relationships. Therefore the approach can apply various standards as description logic and achieve automatic reasoning, e.g. Resource Description Framework Schema (RDFS) [AH2011] as well as Web Ontology Language (OWL) [AH2011] and its subsets [GWP+2004] [HWD2013].

The pros and cons as well as the applicability of the above approaches are analysed in [SL2004] and further summarised in [PZC+2014] as shown in Table 2:

Approaches	*Pros*	*Cons*	*Applicability*
Key-value	▪ Simple ▪ Flexible ▪ Easy to manage with small size	▪ Strongly coupled with applications ▪ Not scalable ▪ No structure or schema	▪ For limited amount of data ▪ For independent and non-related information

		▪ Hard to retrieve information ▪ No way to represent relationships ▪ No validation support ▪ No available standard processing tools	▪ For less complex temporary modelling requirements
Markup scheme	▪ Flexible ▪ More structured ▪ Validation possible through schemas ▪ Processing tools available	▪ Application depended as no standards for structures ▪ Can be complex in case of multiple levels of information ▪ Moderately difficult to retrieve information	▪ For intermediate data organisation or data transfer over network ▪ For decoupling data structures used by multiple components in a system
Graphical	▪ Allows relationships modelling ▪ Information retrieval is moderately easier ▪ Different standards and implementation are available ▪ Validation possible through constraints	▪ Querying can be complex ▪ Configuration may be required ▪ Interoperability among different implementation is difficult ▪ No standards but governed by design principles	▪ For long term and large volume of permanent data archival ▪ Historic context can be stored in database
Object based	▪ Allows relationships modelling ▪ Can be well integrated using programming languages ▪ Processing tools are available	▪ Hard to retrieve information ▪ No standards but governed by design principles ▪ Lack of validation	▪ For representation of context in program-ming codes level ▪ Allows context runtime manipulation ▪ Short term and temporary ▪ Supports data transfer over network
Logic based	▪ Allows to generate high-level context using low-level context ▪ Simple to model and use ▪ Supports logical reasoning ▪ Processing tools available	▪ No standards ▪ Lack of validation ▪ Strongly coupled with applications	▪ For generating new knowledge ▪ For modelling events and actions ▪ For definition of constrains and restrictions
Ontology based	▪ Supports semantic reasoning ▪ Allows more expressive representation of context ▪ Strong validation ▪ Application independent and	▪ Representation can be complex ▪ Information retrieval can be complex and resource intensive	▪ For modelling domain knowledge ▪ For structuring context based on relationships defined by the ontology ▪ Data can be stored in appropriate data sources rather than on ontologies,

	allows sharing ▪ Strong support by standardisations ▪ Fairly sophisticated tools available		while structure is provided by ontologies

Table 2. Comparison of state of the art context modelling approaches, adapted from [PZC+2014]

The above context modelling approaches are categorised based on their different emphases regarding the understanding, organisation and representation of context information, therefore they are not mutually exclusive. As a matter of fact, to make best use of their advantages and compensate for their disadvantages, it is not only reasonable but also feasible to develop hybrid approaches, which integrate various modelling and reasoning techniques with each other. A representative is the CARE framework for context awareness developed in [ABR2009], which integrates a markup model with an ontological model to realise the semantic reasoning supported by representation formalism. As described in Chapter 3, the proposed theoretical framework also embraces the idea of hybrid modelling and integrates ontology model with logical reasoning, so it is capable of processing the sophistication that is usually expected in large-scale systems.

2.2.3 How to evaluate context models?

Despite the huge amounts of literature on developing various context modelling techniques, some of them even comparing them hence the results shown in Table 2, there are few publications on how the resulting context models should be evaluated.

Nevertheless, Bettini et al. summarise several requirements for context models and their context management systems in pervasive computing in [BBH+2010]:

- *Heterogeneity and mobility*: Context models have to be able to handle a large variety of context information, which differs in source, update rate, and semantic level. The models should also be able to express the different types of context information, and the context management systems should be able to manage the information depending on its type. Furthermore, the models and the systems should also be able to support either mobile context-aware application or mobile context information sources.

- *Relationships and dependencies*: Context models and their context management systems should be able to correctly capture and handle the relationships, especially dependencies between different types of context information.

- *Timeliness*: Context models and their context management systems should be able to capture and manage context histories, in case the context-aware applications need access.

- *Imperfection*: Context models and their context management systems should be able to handle the situation that the context information comes with variable qualities or even incorrect, or it is incomplete or conflicting with other context information.

- *Reasoning*: Context models and their context management systems should be able to support both consistency verification and context reasoning techniques, to decide whether any adaptation is necessary or derive new context fact from existing ones.

- *Usability of modelling formalisms*: The modelling formalism should be easy for model designers to translate real world concepts to the modelling constructs and for the context-aware applications to at runtime use and process context information.

- *Efficient context provisioning*: The context models should provide clear and efficient access to context information for the context-aware applications.

As shown in Chapter 6, evaluation metrics are derived for the assessment of the security-oriented context model, closing the gap between model generation and evaluation.

2.3 Context modelling for IT security

As this dissertation focuses on adopting context modelling to ensure security, this section starts with some basics on IT security, then summarise the state of the art of context-aware applications regarding security.

2.3.1 What is IT security?

When the concept of security is introduced to the scope of computer science, it is usually referred to as the protection of systems and information from harm, theft, and unauthorised use [Britannica2013c].

Regarding how security can be specifically provided, the ISO mentions several security services within the framework of OSI (Open Systems Interconnection) Reference Model in the standard ISO 7498-2 (recommendation ITU-T X.800) [ISO1989]: *authentication* (*peer entity authentication* and *data origin authentication*), *access control*, *data confidentiality*, *data integrity*, and *non-repudiation*. Additionally, it also considers *availability* as a general requirement of OSI security management.

Based on these, three main components are generally identified, known as the *CIA triad* or *CIA model*, as summarised by Bishop in [Bishop2002]: *confidentiality*, *integrity*, and *availability*. Confidentiality is the concealment of information or resources, similar as stated in the OSI model, but also covering the access control service [ISO1989]. Integrity refers to the trustworthiness of data or resources and usually phrased in terms of preventing improper or unauthorised change, and therefore also covers the scope of the data origin authentication service in [ISO1989]. Availability refers to the ability to use the information or resources as desired.

Based on the above identified concepts, the following definitions of security aspects are selected for their usage in this dissertation (based on [ISO2012]):

- *Authenticity* is an ambiguous security aspect that can refer to two related concepts for authentication: *Data origin authenticity* refers to the proof of the origin of the data and ultimately together with its genuineness, truth, or realness. *Entity authenticity* is the proof that an entity (e.g. a person or other agent) has been correctly identified as the originator, sender or receiver, i.e. the ability that ensures an entity to be the one that it claims to be. It is often that authenticity is referred to as entity authenticity solely, as data origin authenticity is considered as a distinct aspect called *provenance*.

- *Integrity* is a security aspect that describes the accuracy and completeness of objects. It refers to the ability of preventing improper and unauthorised change of data.

- *Confidentiality* is a security aspect that refers to the non-disclosure of resources with respect to unauthorised entities. Depending on the security level this can also include the concealing of traces of communication and transmission of resources.

- *Availability* is a security aspect denoting that resources can be accessed and used by authorised entities.

- *Non-repudiation* is a security aspect to prove the actual occurrence or non-occurrence of an event and its participating entities with respect to third parties.

Additionally, *privacy* is also sometimes mentioned as the sixth security aspect. Although in a strict sense privacy is restrained to refer to a person whereas confidentiality refers more generally to data, they are often used interchangeable.

To further interpret the above security aspects, the concept of *security policy* is defined as statement of what is allowed and what is not, and the concept of *security mechanism* is derived as a method, tool, or procedure to enforce a security policy [Bishop2002]. Following a security policy's specification of "secure" and "non-secure" actions, the corresponding mechanisms can be used to achieve various security goals, mainly in three aspects as summarised also in [Bishop2002]:

- *Prevention* means that an attack will fail. Prevention mechanisms can prevent security compromise of parts of the system, and at least in theory the resource protected by such mechanism need not be monitored for security issues.
- *Detection* means, despite that an attack cannot be prevented there are still mechanisms to either determine the attack is underway or report after its occurrence. Such mechanisms monitor the attack to provide data about its nature, severity and results. The resource protected by these mechanisms is continuously or periodically monitored for security issues.
- *Recovery* has two forms. The first is to stop an attack and to assess and repair any damage caused by the attack. Moreover, it can also involve identification and fixing of the vulnerabilities revealed by the attack. The second form is that the system continues to function correctly while an attack is underway, or it detects incorrect functioning automatically and then corrects the error.

In the context of a complex system, security policies regulate what in the system should be protected [BS2002] to meet various aspects of security requirement depending on the application scenario. Most complex systems face the problem that how to organise large quantity of security policies properly and efficiently. As one solution, Baskerville et al. proposed in [BS2002] a functional hierarchy of policies uses a three level division consisting meta-policies, high-level policies, and low-level policies. This hierarchy increases in granularity from the abstract meta-policies to specific detailed policies, which may be so concrete that they directly demand or prohibit certain implementations or mechanisms. However, as the systems become more complex and the number of low-level policy increases, it may occur that a large quantity of low-level child policies originates from a single parent high-level policy. Thus policy management would become rather complicated in this case.

2.3.2 Use context models to address security issues

With the development of context-aware pervasive computing systems, the concern about security has been raised and gradually increased [HSK2009]. As context modelling is in the first place commonly used in the field, it has also been applied to generate context models which lead to systems with security-related requirements.

Jiang and Landay described in [JL2002] a theoretical model privacy control in context-aware systems based on a core abstraction of information spaces, which provide a way to organise context, namely relevant information, resources and services. Bhatti et al. pointed out in [BBG2004] the lack of context-aware models for access control in web services, designed an access control scheme to address the issue, and proposed an extended, trust-enhanced version of their XML-based Role Based Access Control (X-RBAC) framework which incorporates context-based access control. Hill et al. in [HAC+2004] developed a programmable middleware architecture named HESTIA as the solution to secure the cyber infrastructure for large-scale pervasive computing environments, providing services including monitoring, intrusion detection, replication, authentication, etc. Minami and Kotz presented in [MK2005] a secure context-sensitive authorisation system which protects confidential information in facts or rules and allows multiple hosts in a distributed environment to perform the evaluation of an authorisation query in a collaborative way. Falkovych and Nack in [FN2006] proposed an extended authoring support approach by integrating processes of topic identification, context collection and discourse structure building in a single environment, allowing identification of the context of the authoring process. Zhang et al. designed in [ZQZ+2007] a context-aware privacy protection framework for context aware services and privacy control methods about accessing personal information in pervasive environment, addressing uncertainty issues using a fuzzy privacy decision information system. Jürjen et al. in [JSB2008] conducted a security analysis of the context information of mobile system architecture by applying a model-based approach with a UML extension. Filho and Martin proposed in [FM2008] an owner-centric QoC (Quality of Context)-aware context-based access control model, namely QACBAC, to take into account both context information and its QoC indicators to grant and adapt access permission to resources in pervasive computing environments. Seifert et al. presented in [SLC2009] a context-sensitive security model for privacy protection on mobile phones, and then implemented the model in a system named TresurePhone, which not only handles the user's context specific need for privacy but also integrates supporting context information based on locations and actions. For achieving effective security in mobile computing environment, Johnson et al. defined the term of security-relevant context and proposed the notion of shrink-wrapped security which couples a user's situation with security in [Johnson2009], and further presented in [JSG+2011] an approach to practically incorporate the security-relevant context into security services with a focus on access control. Khan and Sakamura in [KS2012] explored the relationship of access control and context awareness in pervasive computing and proposed a comprehensive context-aware access control model for pervasive healthcare services.

Besides pervasive computing, there are also other scenarios where context is involved in the modelling process for security. As early as in 1993, Woo and Lam in [WL1993] specified authentication protocols as formal objects with precise syntax and semantics and defined a semantic model that characterises protocol execution. Despite the actual term of "context" was never mentioned, by the standards summarised in earlier sections, their work was indeed all about how to extract and organise security-related context, which was referred to as "correctness properties" in their paper. Another early example would be the Path Context Model (PCM) of security proposed by Boshoff and von Solms in [BS1989], where security demands were analysed in real world commercial computer environments and the model was originated by using context-sensitive grammars to accommodate secure environment concepts and form a basis for automatic security evaluation and profile generation. There are also more recent examples that rest outside the field of pervasive computing. Dunkerley and Tejay developed in [DT2009] the information system security success model and used it for e-Government context for the dimensions of integrity, confidentiality, and authenticity. Simpson et al. in [SSE+2010] established a context for secure information systems development as well as a set of models used to develop and apply a secure software production pedagogy, which aimed to enable even entry-level university students to learn how to acquire new knowledge and adapt their standard software security approaches as a direct result of the fast flux of new technology development and services built on these new technologies.

Furthermore, there are also some security-related scenarios, where context modelling exists in a de facto way, i.e. modelling in those scenarios is actually context-aware, despite viewed from such perspective. A perfect example of such is the development and maintenance of antivirus software: typical antivirus software is able to run signature based detection, i.e. the software analyses its environment and compares the file contents to a dictionary of virus signatures to identify viruses and other malware [Landesman2009]. This comparison can be based on simple key-phrase matching, or more effectively semantic-based approach (see [YWL+2007] and [PCJ+2008]), which is able to characterise the behaviour of virus and malware. Such behaviour fits in the concept of context given in this dissertation, and antivirus software can therefore be regarded as context-aware service. Additionally, such context awareness is further enhanced, as the signature dictionary of commercial antivirus is constantly updated to address newly emerged virus and malware, so new context is constantly extracted and added to the service to assure that it stays context-aware.

2.4 Security in digital long-term preservation

As this dissertation selects digital long-term preservation (sometimes also referred to as digital long-term archiving) as the application scenario of system level context modelling for its security, this section briefly introduces the security issues in this scenario and the state of the art regarding these issues.

2.4.1 Security issues for digital long-term preservation

The heritage of human society has been presented on various materials since ancient times, e.g. stone, bones, vellum, parchment, bamboo, silk, paper, etc. The traditional objects based on such materials such as text in books or images on photographs have an important property, that they are immediately the content which can directly pass the information [BKG+2009]. However, nowadays more and more information exists in digital form, or is even born-digital. These digital objects, in contrast with traditional ones, always require their corresponding environments to render or perform them to pass the information [BKG+2009]. Such environments can be software related, e.g. software interface, data format, etc., or hardware related, e.g. physical storage media, rendering device, etc. Nevertheless, both software and hardware environments can become obsolete over time, causing digital objects not accessible any more. This is called digital obsolescence, and it started to draw the attention of the libraries and archives only in 1990s [Hedstrom1997] and directly called for the development of the concept of digital long-term preservation is developed. However, besides digital obsolescence, there are further issues that need to be solved [Giaretta2011]: there shall be some mechanism to ensure the digital objects not to be altered over time, i.e. they can be trusted that they are indeed what they are claimed to be; the legal position of the digital objects should always be tracked, as legal system changes over regions and time, resulting various constrains on possible actions to the objects; furthermore, the links to networked resources could fail or be reallocated over time, therefore they need to be properly maintained.

From the security point of view, all those exemplary issues identified for digital long-term preservation to solve reflect certain security aspects introduced in subsection 2.3.1.

Issues for digital long-term preservation to solve	*Related security aspects*
Digital obsolescence	Availability
Trusted digital objects	Integrity, authenticity
Legal position of digital objects	Confidentiality, non-repudiation
Correct links to resources	Availability

Table 3. Exemplary issues for digital long-term preservation to solve and the security aspects they reflect

As Table 3 shows, comparing to the examples given in subsection 2.3.2, digital long-term preservation is an application scenario where multiple security aspects are involved. Obsolete software and hardware environments disable the access to the digital objects, therefore compromise their availability. Link failure or reallocation can be regarded as an extension of digital obsolescence in network environment, therefore also results in the compromise of availability of the digital objects. There are two layers of requirements for archived digital objects can be trusted: first, the digital objects encompass the identical information as it originally did, i.e. its integrity can be ensured; second, the origin of the objects is clarified, i.e. its authenticity can be ensured. As for the requirement on the legal position, it ranges from the basic level of access control (therefore confidentiality) to more advanced level of being able to prove each single action employed on the objects (therefore non-repudiation).

Summarising this subsection, it can be concluded that security assurance is essential for a competent digital long-term preservation system, and this constitutes the motivation on the study in this field.

2.4.2 Previous achievements

A number of initiatives have emerged in the past few decades regarding the urge of digital long-term preservation. This section summarises some of the representative ones among them, with particular focus on their effects regarding security issues.

OAIS (Open Archival Information System)

The Consultative Committee for Space Data Systems (CCSDS) published the OAIS model in 2002 [OAIS2002] and one year later it was adopted as ISO standard ISO14721:2003. Proven to be a very useful high-level reference model, OAIS uses functional entities to describe the exchange of information during the preservation procedures. Figure 5 illustrates the functional entities in OAIS model as well as the information exchange among them.

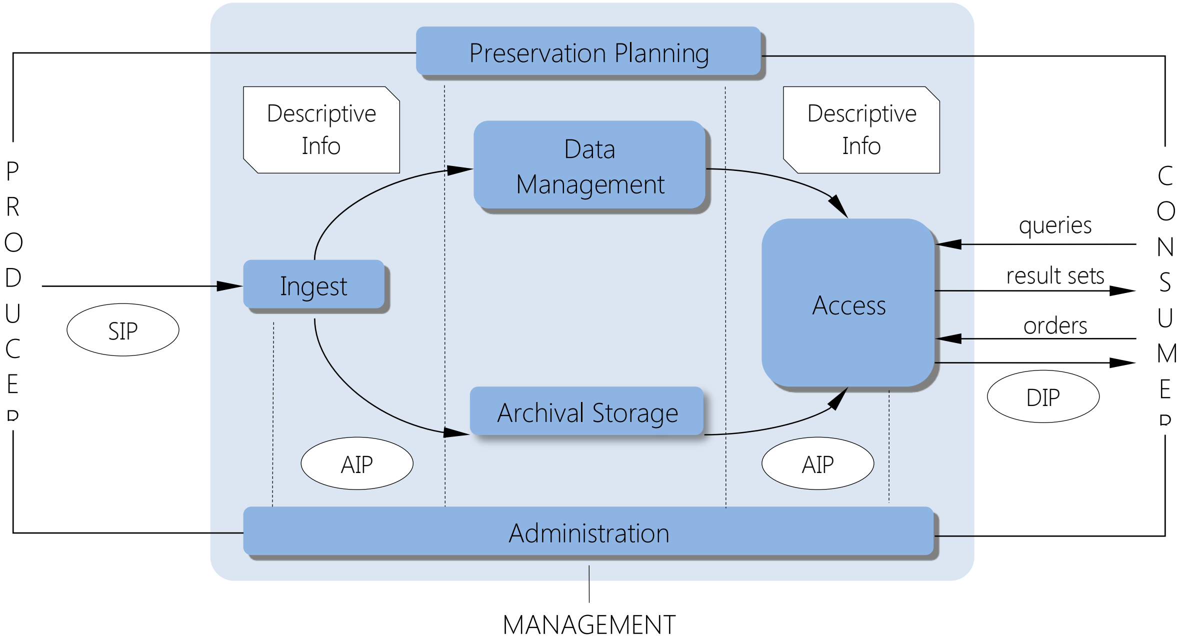


Figure 5. OAIS functional entities [OAIS2002]

OAIS uses various Information Packages (IPs) to represent digital objects in different archiving steps, and Descriptive Information (DI) to represent their metadata. The role provided by each functional entity is briefly explained as follows [OAIS2002]:

- *Ingest* entity provides the services and functions to accept Submission Information Packages (SIPs) from Producers and prepare the contents in the form of Archival Information Packages (AIPs) for storage and management within the archive.
- *Archival storage* entity provides the services and functions for the storage, maintenance and retrieval of AIPs.
- *Data management* entity provides the services and functions for populating, maintaining and accessing both Descriptive Information which identifies and documents archive holding and administrative data used to manage the archive.
- *Administration* entity provides the services and functions for the overall operation of the archival system.
- *Preservation planning* entity provides the services and functions for monitoring the environment of the OAIS and providing recommendations to ensure that the information stored in the OAIS remains accessible to the Designated User Community over time, even if the original computing environment becomes obsolete.
- *Access* entity provides the services and functions that support Consumers in determining the existence, description, location and availability of information stored in the OAIS, and allowing Consumers to request and receive information products.

Besides these functional entities, OAIS also defines several common service entities to provide pervasive services and functions. These entities include [OAIS2002]:

- *Operation system services* entity provides the core services for operating and administering the application platform as well as an interface between application software and the platform.
- *Network services* entity provides the capacity and mechanisms to support distributed applications requiring data access and applications interoperability in heterogeneous, networked environments.
- *Security services* entity provides capabilities and mechanisms to protect sensitive information and treatments in the information system.

As the preceding descriptions show, some security issues are already addressed, however they are quire underspecified [OAIS2002]: the *security services* entity is designed to include an identification/authentication service, and access control service, a data integrity service, a data confidentiality service and a non-repudiation service. Despite the these services address all the security aspects that should be covered for digital long-term preservation, the description of the services fails to propose any detail, e.g. how to implement the services, how to integrate these services with other functional entities in the information system, etc. Furthermore, in the *administration* entity, the "establish standards and policies" is responsible for establishing and maintaining the archive system standards and policies, including security policies for the contents of the archive. However the content is also in lack of details, e.g. guidance on how the security policies are established and organised.

CCSDS issued the second version of OAIS standard in 2009. Besides the content introduced above, there are some enhancement in the security-related issues in this version [OAIS2009]. It references the Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC). TRAC was issued by the Center for Research Libraries (CRL) and Online Computer Library Center (OCLC) in 2007 [TRAC2007]. It offers a set of evaluation criteria, which can be applied for audit and certification to achieve a trusted environment for digital repositories, which also takes the security-related requirements into consideration. Then, for the first time in the history of OAIS, the full definition of authenticity is introduces as the degree to which a person (or system) regards an object as what it is purported to be and it needs to be judged on the basis of evidence. This issue also make it clear that it is one of the main goals of the long-term archiving system to support the authenticity of the data. Furthermore, the issue introduces the Transformational Information Property as one of the properties of the data. This property is a contribution to authenticity when e.g. any format transformation operation needs to be applied on the archived data. A new type of Preservation Description Information is also added: Access rights provide the terms of access, including preservation, distribution and usage of Content Information. While introducing the functional entity of *ingest*, the issue describes the necessary operations on how to preserve authenticity within the scope of this functional entity, however no similar content is offered for other functional entities.

CASPAR (Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval)

Started in 2006, CASPAR project aimed at the implementation, extension and validation of the OAIS reference model proposed in [OAIS2002]. Within the scope of CASPAR, several major threats have been identified and considered as salient to be solved. Table 4 summarises these specific threats with regard to the security issues introduced in Table 3 together with the relevant security aspects their corresponding solutions that CASPAR proposes.

Security issue	*Specific threat*	*Security aspect*	*Solution in CASPAR*
Digital obsolescence	Users may be unable to understand or use the data e.g. the semantics, format, processes or algorithms involved Non-maintainability of essential hardware, software or support environment may make the information inaccessible	Availability	Archive's ability to create and maintain adequate Representation Information (RepInfo) Archive's ability to share information of hardware and software and their replacements or substitutes
Trusted digital objects	The chain of evidence may be lost and there may be lack of certainty of provenance or authenticity The ones we trust to look after the digital holdings may let us down	Authenticity, integrity	Archive's ability to bring together evidence from diverse sources about the authenticity of a digital object Certification process so that one can have confidence about whom to trust to preserve data holdings over the long term
Legal position of digital objects	Access and use restrictions may make it difficult to reuse data, or alternatively may not be respected in future	Confidentiality	Archive's ability to deal with digital rights correctly in a changing and evolving environment. Preservation-friendly rights or appropriate transfer of rights is necessary
Correct links to resources	The current custodian of the data, whether an organisation or project, may cease to exist at some point in the future	Availability	Brokering of organisations to hold data and the ability to package together the information needed to transfer information between organisations ready for long term preservation

-	Loss of ability to identify the location of data	Availability	An persistent ID resolver system

Table 4. Specific threats identified in CASPAR and corresponding solutions with regard to general security issues for long-term preservation and addressed security aspects [Giaretta2011]

CASPAR employs a document/data-centric approach, which relies heavily on the interactions between RepInfo (which is an extended concept of DI in OAIS) and IPs to describe various preservation related processes. Figure 6 on page 24 illustrates the CASPAR information flow architecture, which illustrates all such interactions in the embodied OAIS functional entities. This architecture is defined with specific focus on the transformation of a digital object (with various forms of IPs) as well as its metadata (in the form of RepInfo) being accurate, up-to-date and intact. It describes how generic interfaces with virtualisation technologies might provide "abstract interfaces on top of concrete implementations [CASPAR2007]", allowing data to be accessed and manipulated independently of discipline or platform, with preservation activities distributed across systems as shared services. Furthermore, the original OAIS representation network is logically extended to a RepInfo registry. RepInfo and its dependencies are therefore tracked, maintained and preserved.

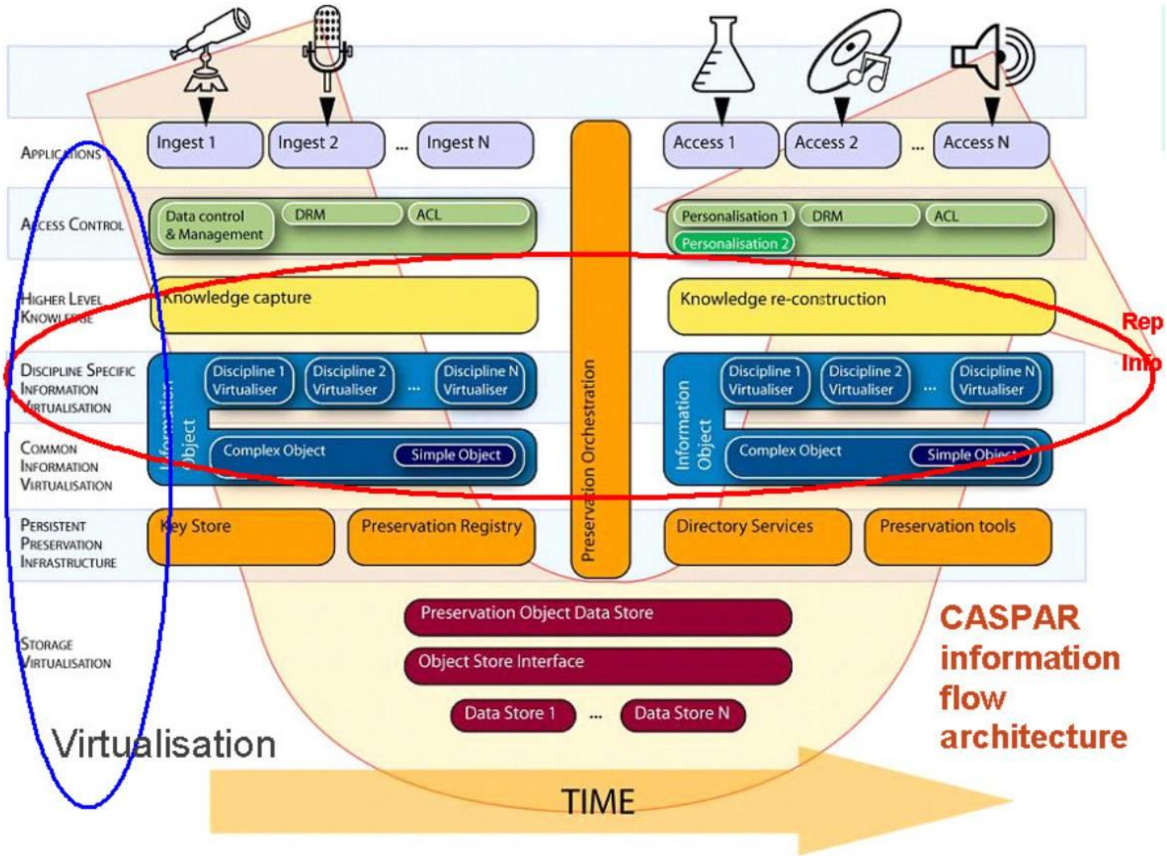


Figure 6. CASPAR information flow architecture [CASPAR2007]

Despite raising security level hardly being one of the core objectives of CASPAR, it has indeed managed to provide more specific details towards a more secure archiving environment. Access control and Digital Rights Management (DRM) are proposed for *access*, so a digital object can be properly processed when it enters the archive. Provenance management is proposed with regard to the digital rights and authenticity of digital objects. Security management is proposed to cover *preservation planning*, *data management* and *administration*, so it deals with user account/role/profile, content access permissions as well as digital rights, and also guarantees authenticity.

However, these specifications are hardly sufficient for system level security. They are scattered and lacking of systematic organisation. More importantly, as the archive is supposed to keep functioning for a long term, there is no mechanism to guarantee the flexibility of the system to adapt for either security threats or conflicts emerging over time.

e!DAL (electronical Data Archive Library)

Current studies on life sciences tend to generate large amount of raw/primary data, hence it is commonly encountered with the problem of limited storage space, and this problem can become even more complicated due to the heterogeneity of the data [ALC+2012]. In the meantime, the study of life sciences often require the publication of primary data, and the current solution is to upload the data to special domain-specific repositories (e.g. European Nucleotide Archive or the BioModels Database, where basically the findings and test sets of bio scientists used in their publications are uploaded for others to use and validate) as well as to internal storage systems. Therefore, within the scope of ongoing DPPN/EPPN (Deutsches Pflanzen Phänotypisierungsnetzwerk/ European Plant Phenotyping Network) project, e!DAL is designed as a storage system, which at the time of writing this dissertation is still under development, aiming to the support of long-term archiving, sharing and citing primary data in life sciences [ALC+2012].

e!DAL classifies the data dissemination into three domains [ALC+2012]: the private domain (raw data, typically collected from the scientists/researchers themselves), the group domain (primary data shared only within the working group), and the public domain (the selected data for publication). The framework is implemented in Java to support the workflows for dissemination, so that the data that enters into the system is disseminated to the public databases and/or the internal databases, and corresponding services are provided, including version management, metadata storage, persistent identifiers, etc .

However, in e!DAL the long-term preservation itself is only addressed in a very basic manner, and only in the sense of disseminating the data to other sites which are contractually obliged to store it for 10 years or so [ALC+2012], yet not tackling data obsolescence. From the security point of view only entity authentication is addressed, in the way that the system does not allow the deletion or modify the data after the ingest, but automatically makes a copy for every new version. However, no cryptographic means are proposed to allow the verification of the data integrity systematically, but this is rather done on per-user basis, e.g. depending on the user who ingests the data already provides hashes. Further, it is also not mentioned that any audit trail is planned to be implemented to meet further security needs.

SHAMAN (Sustaining Heritage Access through Multivalent AchiviNg)

SHAMAN project was launched in 2008, with the overall objective of a next generation digital preservation framework. Similar to CASPAR, to a large degree the framework is also constructed with the reference to OAIS. Nevertheless, as SHAMAN proposes to develop and integrate technologies to support contextual and multivalent archival/preservation processes, it achieves significantly in its conceptualisation with context, especially regarding security concerns.

Schott et al. presented in [SDV+2008] a first study on how to enforce integrity and authenticity for digital preservation beyond electronic signatures to build a solution for data emulation and migration. They analysed the integrity and authenticity requirements in SHAMAN environment with iRODS (i Rule Oriented Data System) grid and multivalent engine, summarised common integrity models including Biba model [Biba1977] and the Clark-Wilson model [CW1987], and proposed a novel approach for enforcing integrity and authenticity for digital objects by extending the Clark-Wilson integrity model. Based on this, in [SKD+2010] they introduced a more detailed investigation and showed exemplarily the influence of the application of the extended Clark-Wilson security model on the use cases and roles of the SHAMAN preservation environment. In the meantime, in [EKB+2009] Engel at al. motivated the approach of context-oriented information retrieval, based on the context appearing in a scientific archive scenario and a proposed information life cycle model. Brocks et al. introduced in [BKJ+2010] the concept of context into digital long-term preservation and proposed a generic context model to provide a formal representation for capturing all aspects of the context information of archived digital objects, to enable retracing information paths for future reuse.

An assessment framework is also developed within the scope of SHAMAN and it relates to several perspectives [SHAMAN2009]: 1) Regarding the SHAMAN project, it aims to performing the evaluation according to the objectives of the project; 2) Regarding the developed and realised SHAMAN preservation framework, it employs the relevant elements from TRAC (see Appendix B for its audit and certification criteria) and a Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) toolkit, which yields a bottom-up approach developed by the Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE) [IRM+2009][DRAMBORA2014]; 3) Regarding the individual work-packages of SHAMAN, it employs iRODs rules to test the outputs of the work-packages.

These preceding achievements in SHAMAN directly contribute to the state of the art of the work of system level context modelling for IT security, which is described in Chapter 4 in this dissertation.

2.5 DIGITAL DACTYLOSCOPY AND THE SEPARATION OF OVERLAPPED FINGERPRINTS

As this dissertation selects a specific application scenario in forensic dactyloscopy, namely the separation of overlapped fingerprints, to perform data level context modelling, this section introduces some fundamentals regarding this scenario: subsection 2.5.1 introduces very briefly about forensic dactyloscopy, subsection 2.5.2 gives a state of the art on the separation of overlapped fingerprints, subsection 2.5.3 describes the *Daubert* standard, which is the standard that forensic techniques are expected to meet to yield court evidence.

2.5.1 Forensic dactyloscopy

Forensics, or forensic science, is in general the scientific tests or techniques used in connection with the detection of crime [Oxford2013c]. As one of the most important branch of the forensics, forensic dactyloscopy, the forensic science focusing on fingerprint analysis, has been world widely applied by law enforcements and courts as authentication means for more than a century [BMW+2011]. Due to their uniqueness, the fingerprints acquired from the crime scenes, even partial ones, can help crime investigators to identify the people who left them in the first place, with support of proper infrastructure, e.g. AFIS (Automated Fingerprint Identification System).

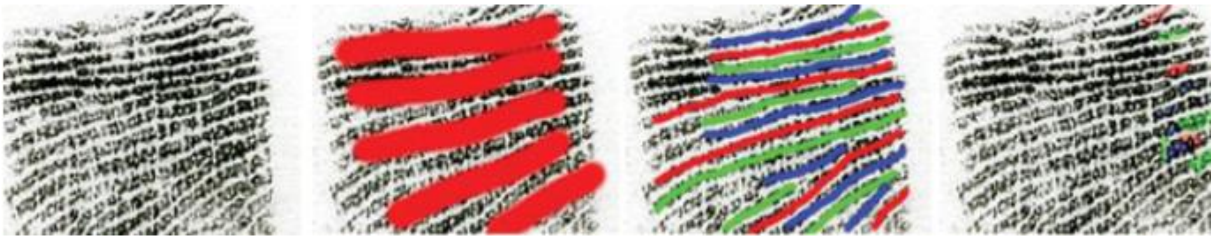
Human's friction ridge skin constitutes fingerprints, and the uniqueness of its features constitutes the uniqueness of fingerprints. Reflecting such uniqueness, three levels of detail were summarised by Ashbaugh in [Ashbaugh1999] for the fingerprint features:

- *First level detail* refers to the general overall directions of ridge flow in the fingerprint, and detail on this level solely is not considered to be unique.
- *Second level detail* refers to the path of a specific ridge, covering specific locations of where a ridge terminates at a ridge ending or bifurcation (or Galton points), which are referred

to as minutiae. The ridge path and its length with minutiae are unique, and the sequences and configurations of a series of ridge paths are also unique.

- *Third level detail* refers to the shapes of ridge structures, encompassing the morphology of the ridge, e.g. edges, textures, and pore positions. They are unique in their shapes, sequences and configurations.

Figure 7 shows the three levels of detail in one original fingerprint. As explained, while the first level detail can be solely lead to conclusive decision about the origin of a fingerprint, combining it with either of other two levels of detail, or both, can usually yield an accurate authentication.



Original fingerprint 1st level detail 2nd level detail 3rd level detail

Figure 7. Three levels of detail of fingerprints [BMW+2011]

As it is not always expected that the fingerprints left at crime scenes are simply visible, or *patent*, various techniques are introduced to acquire those hidden or unseen, or *latent*, fingerprints. In general, these techniques are either chemical or optical, or sometimes both combined. Some of the representative latent fingerprint acquisition techniques are enumerated as follows [BMW+2011]:

- Latent print powder: the conventional way for crime scene investigators to visualise latent fingerprints is with powder, or "dusting". Typically, it involves the application of finely divided particles that physically adhere to the aqueous and oily components in latent fingerprint residue [SK2011], so the fingerprint becomes more visible via reflected light, absorbed light, or luminescence. Afterwards, the fingerprint can be lifted with transparent tape or other lifters, or simply photographed.

- Ninhydrin: Ninhydrin is used to visualise latent fingerprints, because it reacts with the amino acid in the residue and forms a deep purple compound, exhibit excellent contrast and clarity [CLM+2004]. After that, the contrast can be further enhanced and photographed.

- 1,8-Diazafluoren-9-one (DFO): DFO is another reagent with the amino acid in the fingerprint residue. The reaction results in a faint red or pink fingerprint that is intensively fluorescent at room temperature [PGM1990]. The combination of DFO followed by ninhydrin develops more latent fingerprints than DFO or ninhydrin alone [WRB+2005]. Therefore fluorescence of the developed fingerprint can be photographed.

- 1,2-Indanedione: 1,2-Indanedione also reacts with amino acid, visualise the fingerprint residue by resulting fluorescence. It develops even more latent fingerprints than the combination of DFO and ninhydrin [WSS+2001].

- 5-Methlthininhydrin (5-MTN): 5-MTN reacts with amino acid and develops fingerprints under heat and humidity in purple colour. Therefore the fingerprints can be easily photographed.

- Cyanoacrylate fuming: The liquid commercial adhesive (CA) fuming is a versatile and effective latent fingerprint development technique on various surfaces, as CA vapours are extremely sensitive to fingerprint residue. Fully developed CA fingerprints are a three-dimensional matrix, often already visible to unaided eye and can be further enhanced with various techniques.

- Fluorescence examination: A forensic light source or laser is also often applied to help examine for latent fingerprints. By using the correct barrier filters that block out the light from the forensic light source being used, but not the fluorescence, a very high signal-to-noise ratio can be observed. This technique is often combined with the application of fluorescent chemicals.
- Vacuum metal deposition: As a long-established industrial technique for the application of metal coatings to components such as glass to form a mirror, vacuum metal deposition (VMD) is now also used to develop latent fingerprints. Typical choices of metal include gold, silver, cadmium, and zinc. VMD can be used on some surfaces that considered as tricky for other techniques, e.g. plastic bags and packaging and firearms, to visualise fingerprint residues so they can be easily documented by photographing.

Besides the conventional fingerprint developing techniques summarised above, with the development of modern non-invasive sensing and analysing technologies, there are new techniques derived for the acquisition, investigation of analysis of fingerprint evidence, some of them even expanding the goal of digital dactyloscopy from yielding positive identification to revealing further context information from the evidence. Endeavour has been made to visualise the latent fingerprints, which can be hardly perceptual under visible light, in invisible light spectrums. Plese et al. introduces an approach to visualise untreated latent fingerprints in the infrared (IR) spectrum between 400 and 720 nm on various substrates applying a CONDOR™ Hyperspectral Imaging System [PES2010]. Crane et al. propose in [CBP+2007] a method that applies FTIR (Fourier transformation infrared spectrometry) imaging to non-invasively detect latent fingerprint and preserve trace evidence (e.g. explosives or substance of abuse) associated with the prints. Gibson et al. compares in [GBB2012] three latent fingerprint imaging systems based on UVC (ultraviolet C) light source: 1) DEUS (digital enclosed ultraviolet imaging system), which applies home-made UVC-sensitive back-thinned CCD and camera, 2) RUVIS (reflected ultraviolet imaging system), which is a UVC-sensitive image intensifier, and 3) a flatbed scanner fitted with a UVC light source. Their work reveals that the DEUS yields the best results on porous and non-porous substrates, followed by RUVIS and the flatbed scanner. Further extensive study on reflected UV for the visualisation and enhancement of latent fingerprints can be referred to the work by Richards and Leintz described in [RL2013]. Besides IR and UV, SERS (surface-enhanced Raman spectroscopy) is also used to visualise latent fingerprints. Connatser et al. applies SERS through the targeting of lipids and amino acid components that exist in the fingerprints [CPG+2010]. Guicheteau et al. derives an approach that applies semi-automated Raman-based chemical imaging to not only visualise the latent fingerprints, but also identify threat material present in the secretions, e.g. drugs or explosives [GST+2013]. CWL (Chromatic White Light) sensor is another technology that is used in digital dactyloscopy. It makes use of the chromatic aberration of a beam of white light to generate both intensity and topography image of the sample. Therefore, it is studied to classify the substrates on which latent fingerprints are left (see [GV2011] and [GFV2012]), localise latent fingerprints (see [JHS+2012] and [MHF+2012]), acquire latent fingerprints in a contactless way (see [HDP2011]), and estimate the age of latent fingerprints left on various substrates (see [MGD+2012] and [MPD+2013]) or even identify the deposition order of a series fingerprints (see [SMD2012]). In this dissertation, CWL sensor is also used for the contactless acquisition of latent fingerprints for the approach of separating overlapped fingerprints derived in Chapter 5.

Besides proper fingerprint acquisition technique, it is also important for crime investigations to have a system that provides efficient fingerprint analysis to lead to the identities behind those fingerprints. Therefore there are currently multiple such infrastructures, so the law enforcements can have acquired fingerprints uploaded, analysed and archived. The most representative one of those is AFIS (Automated Fingerprint Identification System), which originates from early 1960s, when the FBI in the United States, the Home Office in the United Kingdom, Paris Police in France, and the Japanese National Police initiated projects to develop automated fingerprint identification systems, so the emerging electronic digital computers could be used to assist or even replace the labour-intensive processes of classifying, searching, and matching of fingerprints used for personal identification [BMW+2011]. Based on AFIS, now FBI is operating and maintaining world's largest collection of criminal history information, known of IAFIS (Integrated Automated Fingerprint Identification System). It has enormously expedited the processing of fingerprint evidence: it is able to respond to "electronic criminal transactions" (in this case the uploaded raw fingerprint images to be matched with) in less than 20 minutes, and civil background checks in less than 3 hours [BMW+2011].

2.5.2 Previous achievements on separation of overlapped fingerprints

Despite that the fingerprint related techniques have been well developed, there are always challenges in this field. One of these is the separation of overlapped latent fingerprints. It is often that the latent fingerprints acquired from crime scenes are overlaid on others. On one hand, overlapped latent fingerprints occurring in crime scenes contain useful biometric information that might lead to suspects or persons of interest. However they are difficult to process, as current fingerprint processing systems (like AFIS) assume the processed fingerprint images contain single fingerprints, therefore work poorly on overlapped ones [CFJ+2011]. On the other hand, besides identification, the application of advanced contactless, nanometre range sensing technology for fingerprint development has brought more possibilities to dactyloscopy. For example, as introduced in subsection 2.5.1, utilising a CWL sensor for non-invasive acquisition brings two aspects of benefits to the forensic work: First, the non-invasive acquisition preserves the original fingerprint evidence, so it can still be used in further invasive forensic procedures, e.g. it can be tested for its chemical composition using GC-MS (Gas Chromatography–Mass Spectrometry) [HHM+2007]. Second, the non-invasive acquisition also produces fingerprint images with a high resolution, so more details of the fingerprint can be revealed comparing to conventional methods. This in turn enables the retrieval of further context of the fingerprint evidence, e.g. the identification of the deposition order of the different fingerprints, therefore contributing to the generation of a time line [SMD2012]. However, such approach requires the a priori separation of the series of overlapped fingerprints. The convenient procedure to process overlapped fingerprints in the field work of crime investigation is to only use the nonoverlapped regions as partial fingerprints for analysis. However, as doing so means both the physical content and the biometric information in the overlapped region are abandoned, there is a salient demand for an approach of separating overlapped fingerprints while correctly preserving their features, not only for conventional biometric identification, but more importantly, also to enable further forensic analyses.

Singh et al. developed a separation approach in [STK2006] for mixed signals based on Independent Component Analysis (ICA). The approach is demonstrated with three artificially and one genuinely overlapped fingerprints. The main drawback of this approach is that it requires the source images of the fingerprint, therefore does not fit the condition of crime investigations. Bhargava et al. proposed in [BSF+2009] to use infrared spectroscopic imaging to visualise the difference of the chemical composition of the overlapped fingerprints, so they can be separated. Three difference cases are demonstrated, yet no further validation is reported to employ the proposed method on larger test set. This approach still relies on cyanoacrylate fuming for the development of the latent fingerprints, and it is expected to be limited in the case that two overlapped fingerprints have similar chemical composition. A chemical means using mass spectrometry was proposed by Tang et al. in [TLC+2010], which makes use of sputtering gold particles on the samples. However, this approach is demonstrated on only one sample and is considered as invasive, as it bears the risk of compromising the integrity of the evidence and preventing it from being further analysed by other means. Kärgel et al. introduced in [KGL+2011] a blind source separation approach based on maximum a posteriori estimation adapted from [TBS2006]. The approach is tested on 30 overlapped fingerprint samples on two different substrates, but the results showed very low resistance against noise disturbance and no error rate is reported. Chen et al. described a pattern separation approach based on relaxation labelling in [CFJ+2011]. The approach is based on image processing and requires only one source image of the overlapped fingerprints. However, the experimental evaluation is conducted with only genuine matching (i.e. match the testing fingerprint sample and the fingerprint template that belongs to the same finger) and a very limited test set, which contains only overlapped fingerprint images simulated from 4 conventionally acquired latent samples and 100 livescans ways from existing single fingerprint databases. VeriFinger 6.2 SDK is used to conduct the matching test after the separation, and a range of True Acceptance Rate (TAR) of 55% to 80% is reported corresponding to the False Acceptance Rate (FAR) ranging from 10^{-8} to 10^{-4}, yet no detail is give on the basis from which these rates are calculated. This approach is improved by Shi et al. in [SFZ2011], where the algorithm is modified and enhanced, however still tested in the same way with the same test set used in [CFJ+2011], yielding an approximate TAR range of 74% to 88% corresponding to the FAR range from 10^{-8} to 10^{-4}. Zhao and Jain took this approach a step further in [ZJ2012], which introduces more human intervention to the relaxation labelling based approach to improve separation accuracy. The approach is tested on a further extended test set, which comprises 4 real overlapping latent fingerprints, 15 simulated overlapping latent fingerprints, and 100 simulated overlapping livescan fingerprints, and the identification ranks yielded by the fingerprint matcher show obvious improvement on all three types. However, this also makes the separation prone to mistakes due to human reasons, e.g. inexperienced users, and again the improvement is only confirmed with simulated samples. Also based on the approach from Chen et al., Feng et al. described an improved algorithm and constrained it for two special cases in [FSZ2012], where a latent overlapped fingerprint database with 100 samples developed with fingerprint powder and capture with digital camera is also generated and made publicly available. The algorithm is evaluated on 100 latent overlapped fingerprints from this database together with another 100 simulated overlapped fingerprints. A TAR range of 80% to 92% is reported on the simulated database, and a TAR range of 55% to 73% on the latent database, both corresponding to the FAR range between 10^{-8} to 10^{-4}. Additional to preceding separation approaches, Filax et al. introduce in [FKW+2014] a series of evaluation metrics to assess the overlapped latent fingerprint images to estimate the expected separation performance using the algorithm based on [CFJ+2011], however no detail information is reported regarding specific criteria or details error rates.

As the algorithm introduced in [CFJ+2011] serves as the basis of the context-aware approach developed in Chapter 5, it is summarised with more details in the following table.

Processing steps	*Input*	*Output*
1. Input overlapped fingerprint image	Simulated fingerprint image $I = I_{sim}$	$I: H_{image} \times W_{image}$
2. Manually assign fingerprint masks	I, M_1, M_2: $H_{image} \times W_{image}$	$I \circ M_{N1}, I \circ M_O, I \circ M_{N2}$
In this step the investigator manually assigns two fingerprint masks M_1, M_2, both (0,1) matrices, on the image imported in step 1, resulting three region masks: an overlapped region mask M_O, and two		

nonoverlapped regions masks M_{N1}, M_{N2}. These are then applied to I by using the Hadamard product $\circ$.

3. Apply window-wise DFT on the image	I, M_1, M_2	$\{\widehat{win}_1, \widehat{win}_2, \dots, \widehat{win}_n\}$

Fixed block size $b = 16$ and DFT window size $w = 64$ are applied. A series blocks are then generated (Block function) on the masked fingerprint area:

$$\{bl_1, bl_2, \dots, bl_n\} = \mathrm{Block}(I \circ M_1 \circ M_2, b), bl_i: b \times b, 1 \le i \le n$$

Subsequently a series of DFT windows are generated (Window function) in the way that each block locates in the centre of the corresponding DFT window:

$$\{win_1, win_2, \dots, win_n\} = \mathrm{Window}(\{bl_1, bl_2, \dots, bl_n\}, w), win_i: w \times w$$

For each Gaussian filtered window its two-dimensional spectrum is calculated:

$$\widehat{win}_i = \mathrm{DFT}\big(\mathrm{Gauss}(win_i, \sigma)\big)$$

4. Extract the dominant orientations for each block in non-overlapped and overlapped regions	$\{\widehat{win}_1, \widehat{win}_2, \dots, \widehat{win}_n\}$	$\forall \widehat{win}_i: f_{max1,i}, f_{max2,i},$ $O_N, O_O: \left\lceil \frac{H_{image}}{b} \right\rceil \times \left\lceil \frac{W_{image}}{b} \right\rceil \times 2$ $= H_{blocks} \times W_{blocks} \times 2$

The two frequencies with the highest and the second highest amplitudes are selected in each window for its corresponding blocks, and contribute to the orientation field $O(p, q, k)$, where p and q denote the position of the block and k the label:

$$f_{max1,i} = \underset{f \in \widehat{win}_i}{\mathrm{argmax}} |\widehat{win}_i(f)|$$

$$f_{max2,i} = \underset{f \in \widehat{win}_i \setminus \{f_{max1,i}\}}{\mathrm{argmax}} |\widehat{win}_i(f)|$$

$$O_N(p, q, k) = \widehat{win}_{p \cdot b + q}\left(\arg\big(f_{max1, p \cdot b + q}\big)\right), \text{where } M_{N[k]}(p \cdot b, q \cdot b) = 1, k = \{1,2\}$$

$$O_O(p, q, k) = \widehat{win}_{p \cdot b + q}\left(\arg\big(f_{\max[k], p \cdot b + q}\big)\right), \text{where } M_O(p \cdot b, q \cdot b) = 1$$

5. Perform relaxation labelling on the overlapped region	O_O	$O_O': H_{blocks} \times W_{blocks} \times 2$

The relaxation labelling is an iterative approach of updating labelling probabilities $p_{pq,k1}$ and $p_{pq,k2}$, based on the calculated compatibilities R between orientations of block pairs (pq and $p'q'$), using supports s – in this case normalised difference between two orientations. As the result of this step, the orientations within O_O are relabelled, yielding O_O' as the output.

$$s = 1 - \frac{\delta(|O_O(p, q, k) - O_O(p', q', k')|)}{\frac{\pi}{2}}, k, k' \in \{1,2\}$$

$$\delta(x) = \begin{cases} x & \text{if } x \le \frac{\pi}{2} \\ \pi - x & \text{otherwise.} \end{cases}$$

$$R_{pq,p'q'} = \begin{cases} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & \text{if } |p - p'| > 6 \text{ or } |q - q'| > 6 \text{ or } (p = p' \text{ and } q = q' \text{ and } k = k') \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \text{if } p = p' \text{ and } q = q' \text{ and } k \ne k' \\ \begin{bmatrix} s & 1 - s \\ 1 - s & s \end{bmatrix} & \text{otherwise} \end{cases},$$

$$p, p' = 1,2, \dots, H_{blocks}, q, q' = 1,2, \dots, W_{blocks}$$

$$p_{pq,k1}(t + 1) = p_{pq,k1}(t) + \frac{\alpha}{N} \sum_{p'} \sum_{q'} R_{pq,p'q'}\big(k_{pq}, k_{p'q'}\big) \cdot \Big(1 - p_{pq,k1}(t)\Big),$$

$$p_{pq,k2}(t+1) = p_{pq,k2}(t) - \frac{\alpha}{N}\sum_{p'}\sum_{q'} R_{pq,p'q'}\left(k_{pq}, k_{p'q'}\right) \cdot \left(p_{pq,k2}(t)\right) ,k1 \neq k2, t \leq t_{max}$$

This step involves parameters α, and t_{max}, whereas the values of both left open in [CFJ+2011].

6. Merging the two separated orientation fields with non-overlapped regions	$O'_O, O_N, M_{N1}, M_O, M_{N2}$	O_1, O_2

The merging step involves the boundaries $B'_k(x,y,D)$ between overlapped and nonoverlapped regions, which are defined as

$$B'_k(x,y,D) = \begin{cases} 1 & \text{if the block is inside the boundary with a range of } D \\ 0 & \text{otherwise} \end{cases}.$$

The output of this step is a pair of merged orientation fields (O_1, O_2), and the merging decision is based on the compatibility c_k between the dilated overlapped region and the boundaries in both of the nonoverlapped regions.

$$c_k = \frac{1}{2}\left(\begin{array}{c} \frac{1}{\sum_p \sum_q B_1(p,q,D)} \sum_p \sum_q \delta(|O'_O(p,q,k) - O_N(p,q,1)|) \cdot B_1(p,q,D) + \\ \frac{1}{\sum_p \sum_q B_2(p,q,D)} \sum_p \sum_q \delta(|O'_O(p,q,2-k+1) - O_N(p,q,2)|) \cdot B_2(p,q,D) \end{array} \right)$$

$$O_k(p,q) = \begin{cases} O'_O(p,q,k) & M_O(p \cdot b, q \cdot b) = 1 \\ O_N(p,q,1) & (k = 1 \text{ and } c_1 < c_2) \text{ or } (k = 2 \text{ and } c_1 \geq c_2) \\ O_N(p,q,2) & (k = 1 \text{ and } c_1 \geq c_2) \text{ or } (k = 2 \text{ and } c_1 < c_2) \end{cases}$$

7. Apply Gabor filter to render the separation results	O_1, O_2	I'_1, I'_2

At last the merged orientation for each block from step 6 are used as the input of a Gabor filter applied on the corresponding block of the fingerprint.

$$I'_k(p,q) = \text{Gabor}\left(bl_{p \cdot b+q}, O_k(p,q)\right)$$

Table 5. Summary of the separation approaches in [CFJ+2011] (reproduced from [QSZ+2014])

2.5.3 Admissibility of forensic results in court

Besides the weaknesses summarised at the end of last subsection, there is another essential aspect of general requirement for forensic techniques, towards which little endeavour has been made in the research field of separating overlapped latent fingerprints: as the ultimate purpose for developing forensic techniques is to yield statements regarding evidence that can be accepted in court as expert testimony [HS2010], it is also necessary to conduct relevant study on the evaluation and enhancement of the separation approach with regard to existing criteria regarding the admissibility of forensic results as scientific and technical evidence. Therefore, this subsection briefly introduces such criteria used in the U.S., as it has the most active legal system in the world, resulting in one of the most sophisticated and strictest systems of rules regarding the admissibility of evidence.

Prior to 1923, there was no specific rule governing the admissibility of scientific evidence. In 1923, in the *Frye* case, i.e. *Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923)*, for the first time the court stated that, with respect to novel scientific evidence, besides the common relevancy standard, an additional hurdle must be overcome [HS2010]. The court ruled [Congress1923]:

> *"Just when a scientific principle of discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidence force of the principle must be recognized, and, while courts will go a long way in admitting expert testimony*

> *deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field to which it belongs."*

Based on this ruling, a significant principle of *general acceptance* was set to govern the admissibility of novel scientific/technical evidence, derived by new scientific techniques. However, what constitutes general acceptance has never been clearly decided. Still, in the following decades, a number of novel scientific techniques were subject to "*Frye* challenges" in various courts in U.S., including voiceprint spectroscopy, blood spatter pattern analysis, polygraph analysis, and even DNA typing techniques [HS2010].

On Jan. 2, 1975, U.S. Congress approved an evidence code for the first time, known as the Federal Rules of Evidence (FRE). The FRE became effective on Jul. 1, 1975 and contained a specific article regarding expert and opinion testimony. Under the rules in this article, specifically Rule 702, the proponent of expert testimony has the burden of demonstrating that the expert is qualified and that the opinion evidence would be helpful to the fact finder, i.e. the judge or jury [HS2010]. This rule at that time read [HS2010]:

> *"If scientific, technical, or other specialized knowledge will assist the trier-of-fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training or education, may testify thereto in the form of an opinion or otherwise."*

This actually caused the divided opinion among federal and state courts, on whether *Frye* or the new FRE should be used. The question has not been addressed and settled till 1993, when U.S. Supreme Court ruled in *Daubert v. Merrel Dow, 509 U.S. 579 (1993)*, that federal courts could not use the *Frye* rule any more. In interpreting the Rule 702 and other relevant rules, e.g. 401, 402, 403 and 701, 703, and 704, the Supreme Court indicated that the judge must be the "gatekeeper" who decides when scientific evidence is admissible [USC1993]. The Court suggested several criteria that a judge could use in the gatekeeper role, namely falsifiability, knowledge of error rates, peer review, and general acceptance [HS2010]. Based on these findings in *Daubert*, U.S. Congress changed some of the rules in FRE, including the Rule 702, which currently reads [Congress2011]:

> *"A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if: (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue; (b) the testimony is based on sufficient facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert has reliably applied the principles and methods to the facts of the case."*

Since the adjudication of *Daubert*, it was later refined by two important following cases, *General Electric v. Joiner, 552 U.S. 136 (1997)* and *Kumho Tire Co. v. Carmichael, 119 S.Ct. 1167 (1999)*. In both cases the judge exercised his discretion as the "gatekeeper" under *Daubert*, and excluded the testimony [HS2010]. These three cases, also known as the "*Daubert* Trilogy", serve as the basis of the *Daubert* criteria that governs the admissibility of the scientific/technical evidence, especially those derived using forensic techniques. A further interpretation of the *Daubert* criteria is summarised in [LII2013]: *"(1) whether the expert's technique or theory can be or has been tested—that is, whether the expert's theory can be challenged in some objective sense, or whether it is instead simply a subjective, conclusory approach that cannot reasonably be assessed for reliability; (2) whether the technique or theory has been subject to peer review and publication; (3) the known or potential rate of error of the technique or theory when applied; (4) the existence and maintenance of standards and controls; and (5) whether the technique or theory has been generally accepted in the scientific community."*

Daubert criteria indeed set a high standard for forensic techniques, including digital dactyloscopy. In fact it was seriously challenged by *Daubert* in the 1999 case of *United States v. Mitchell* (Cr. No. 96-407-1), when defense moved before trial to bar the government's fingerprint experts from testifying. The admissibility of fingerprint evidence was only reapproved after a 5-day hearing, which ended with the rule that it satisfied all *Daubert* criteria [BMW+2011]. Therefore, the forensic approaches must be as transparent as possible about how it processes the evidence, as the courts "may, and many will, require the experts to show that they know what the scientific method consists of and provide the scientific basis for their conclusions [BMW+2011]" and "can relitigate the admissibility of a certain type of expert evidence if a litigant can make a credible argument that there has been no previous scientific inquiry of the validity of the assumptions on which a forensic field has long rested [BMW+2011]." Despite the evidence based on latent fingerprints is in general long accepted and admitted in courts, this by no means automatically grants the admissibility for the separation results of overlapped fingerprints, unless the separation approaches which yield them satisfy the *Daubert* criteria. This actually implies the requirement of high standard non-repudiation, in the way that all the details about how the evidence is processed are expected to be clarified, confirmed, and understood. However, the existing approaches introduced in subsection 2.5.2 lack concerns in this matter. Therefore, the data level context modelling is proposed and applied to improve an overlapped latent fingerprint separation approach in Chapter 5, making effort towards the direction of meeting *Daubert* criteria.

2.6 Summary

Plenty of academic achievements have been introduced in this chapter, meant to serve as the state of the art for the scientific contributions made by this dissertation (see section 1.3). Section 2.1 and 2.2 introduce the definitions that are used as the fundaments of Chapter 3. Formula (2.1) is used as the basis for more sophisticated formulations, covering other introduced definitions, e.g. scientific modelling and context modelling. Lee's contextual information framework is used for verifying the correctness and generilisability of the derived formulations, and the categorisation from Perera et al. inspires on the aspect of data level context modelling. Additionally, the evaluation criteria proposed by Bettini et al. serve as basis towards a series of evaluation metrics developed in Chapter 6. Section 2.3 describes the security relevant concepts, which not only serve as the fundaments for the modelling theory in Chapter 3, but also form the basis for analysing the security issues in the specific application scenarios in Chapter 4 and 5. Section 2.4 and 2.5 serve respectively for Chapter 4 and 5, analysing the motivation to improve the work in these fields from the security point of view and offering the state of the art.

However, before moving on to later chapters, it is necessary to recall the research challenges that have been identified at the beginning of this dissertation (see section 1.1) and ask one question: to what degree can those achievements respond to these research challenges already?

The first research challenge is to collect useful information and organise it to form security-related context for modelling. While the conception regarding security is quite maturely developed (see section 2.3), it is not yet clear that under what circumstances that which information constitutes context. As introduced in subsection 2.1.2, various definitions (e.g. in [Dey2001], [Lee2011]) and categorisations (e.g. in [SAW1994], [Henricksen2003], and [PZC+2014]) of context have been proposed, but mostly only specialise in their own specific scenarios in the field of pervasive computing, and none of these scenarios has its focus on security assurance. Therefore, despite that the state of the art provides some theory dots, the framework that connects these dots, i.e. clarifies the information considered as context in security scenarios and effectively organise/utilise it to serve the scenarios, is still missing.

The second research challenge is to generalise the context modelling methodology for secure data processing systems. Such general methodology should be developed from the induction based on existing modelling cases. As introduced in subsection 2.2.2, similar to the state of the art on the theories regarding context, current study on context modelling is also mainly restrained in the field of pervasive computing, and security is seldom the primary concern. Exploring the existing cases of applying context modelling to address security issues in pervasive computing, it is not hard to notice that they deal with either confidentiality/privacy (in [JL2002], [ZQZ+2007], and [SLC2009]) or authenticity (most realised with access control, in [BBG2004], [MK2005], [FN2006], [FM2008], [JSG+2011], and [KS2012]), or both (in [HAC+2004], [JSB2008], and [Johnson2009]). This reflects the limited security focus of their application scenarios. Outside this field, the applications have a slightly higher diversity in the involved security aspects (e.g. [DT2009] tackles besides confidentiality and authenticity also integrity). However, as the cases are rather scattered and focus little on generalising the approach, it is not only difficult to systematically analyse the context which is relevant to security, but also improbable for the user to estimate the outcome of the application of context modelling with regard to security assurance and enhancement. Therefore, what in the first place is missing here, is the specific cases on the application of context modelling in various scenarios with systematic security requirements, which enables the induction towards the general methodology.

The last research challenge is to evaluate the performance of security-oriented context models. Little effort has been put into this direction in literature. A possible basis could be the series of factors raised in [BBH+2010] (see subsection 2.2.3). However, no security-related factors are included in this work, so such challenge remains at the time of writing this dissertation.

Furthermore, as digital long-term preservation and digital dactyloscopy are the two application scenarios that this dissertation selects, some research gaps revealed by the state of the art (see section 2.4 and 2.5) are also summarised here:

- In the field of **digital long-term preservation**, there are various theories and even prototypes of archives developed. However, from the security point of view, they still need improvements. For example, as introduced before, although the 2009 issue of the OAIS model references TRAC and makes some improvement to the covered security aspects compared to the 2002 issue, the expression of how to ensure the security for long-term preservation of data is still rather scattered and disorganised. Therefore, it is essential to develop a framework which can be integrated into the existing OAIS standard and particularly aims for system-wide integration and management of security. Addressing this point, system level context modelling is introduced in Chapter 4, to identify security context in the field and construct a security framework to compensate existing system models, especially OAIS.
- With regard to **digital dactyloscopy**, in specific the separation of overlapped latent fingerprints, the existing approaches are nowhere near to Daubert compliance: they have not been thoroughly tested, the reliability has not been investigated, the rather high error rates come with limited statistical significance, and more importantly, no endeavour has been made to clarify the processing steps of potential evidence for the sake of requirement of non-repudiation. Besides these, some of them are also too complicated for crime investigators to use, or not suitable with the conditions of crime investigation at all, or involving invasive acquisition procedures which restrain further forensic processing. Addressing this point, data level context modelling is introduced in Chapter 5, to identify the context during the processing and use it to improve the separation performance.

3 THEORETICAL FRAMEWORK

This chapter proposes a theoretical framework for context modelling. It comprises a meta-model that formalises the conception of security-oriented context modelling. Furthermore, based on the meta-model, it also specifies the meta-model on both system and data levels, according to which context models can be generated, contributing to the applicability of the framework.

Section 3.1 formulates a general meta-model of context models. It starts with a basic logic structure describing scientific model and enriches it by gradually specifying the context and security related concepts in the structure to form more sophisticated meta-models. After that, section 3.2 takes the derived meta-model and projects it further on both system and data levels. On the system level, a hierarchical structure for security policies is proposed to identify and clarify the security requirements, while on the data level the concept of primary and secondary context in both acquisition and processing environments are defined, yielding the effective execution of security mechanisms. As section 3.1 and 3.2 together complete the conception of security-oriented context modelling, section 3.3 proposes a descriptive scheme which approaches a step further to the application of such context models. At last, section 3.4 summarises the whole chapter.

3.1 A GENERAL META-MODEL OF CONTEXT MODELS

Based on the fundamentals on modelling and context introduced in section 2.1, this section derives a general meta-model to describe context models. Subsection 3.1.1 formulates a series of concepts that contribute to the meta-model, and then subsection 3.1.2 derives the meta-model, identifying various types of context.

3.1.1 Concepts towards the meta-model

In section 1.1, a research question is raised: how to collect and organise information that constitutes security related context? Interpreting this question for the sake of context modelling, this question can be rephrased in a more explicitly extensional way: in a given particular scenario where a context model is to be built, what constitutes modelling target and what to its context that is relevant to security issues, and how to organise the context for building the model? As summarised in the end of Chapter 2, this question has not yet been answered due to the gaps between the concepts of context, modelling, and security. To close those gaps, this dissertation proposes to start with a formalised representation of a meta-model, which integrates the concepts of context and security into that of scientific modelling.

As introduced in subsection 2.1.1, scientific models are generated to cover one or more of the following aspects: phenomena, data, and theory. While there is no formalised representation of the model in the first two cases, Hodges proposes in [Hodges1997] to describe model of theory with a structure of 3-tuple

$$\boldsymbol{S} = (U, O, R), U \neq \emptyset, R \neq \emptyset \tag{2.1}$$

which is widely used in modern logic [FH2012]. As a matter of fact, as such structure is formalised on a highly abstract level, it is also suitable to describe a selected part of world, with proper adjustment of the definitions of its elements. Furthermore, as the existing context modelling approaches (see subsection 2.2.2) all emphasise the representation of the relationships existing among those being modelled, this structure already possesses a solid potential to be further extended for context modelling: among the three elements of the 3-tuple, E covers what being modelled, R identifies the connections on E, whereas O depicts the nature of the connections. Therefore, in this dissertation, such structure is also proposed to represent scientific model, again on a highly abstract level:

$$\boldsymbol{M} = (E, O, R), E \neq \emptyset, R \subseteq E \times \ldots \times E, R \neq \emptyset \qquad (3.1)$$

where 1) the non-empty set E is the domain of $\boldsymbol{M}$, and it denotes all the entities contained in the modelling scenario regarding the modelling target, 2) O is an indexed set which can be empty, and it denotes the operations on E, and 3) R is a non-empty indexed set and denotes the relations on E. Similar to the definition given by [Hodges1997], if the interpretation of the elements of $\boldsymbol{M}$ holds with the modelling scenario, $\boldsymbol{M}$ is a model of it. In scientific modelling, depending on the nature of the modelling target, the modelling approach can either start from the very basic knowledge and be progressed by gradually adding necessary information to achieve proper complexity ("top-down") or from the vivid representation of the physical world and be progressed by gradually removing the redundancy to achieve the proper simplicity ("bottom-up"). Nevertheless, both ways yield scientific model with proper granularity, which can be generally described by the structure of 3-tuple in formula (3.1).

After the representation of general scientific model is defined, it can be further specified for *computer science*, where this dissertation lies in. As a scientific discipline, *computer science* is defined in [Britannica2013b] as "the study of computers, including their design (architecture) and their use for computations, data processing and system control". This definition implicitly points out the core concern of computer science, at the same time what distinguishes it from other scientific disciplines: the processing of data or information (which can be regarded as data with semantics). Interestingly enough, this is in fact reflected by how the discipline is named in various European countries, e.g. *Datalogi* (Danish), *Informatik* (German), *informatique* (French). Therefore, when modelling is applied within the range of such discipline, the modelling target is therefore restrained to either data processing procedures in small scales, or data processing systems (one encompasses multiple data processing procedures) in large scales. In this dissertation, the former is referred to as *data level*, whereas the latter as *system level.* Nevertheless, on both levels the resulting model, as a special case of the scientific model defined in formula (3.1), can be further defined using the following meta-model:

$$\boldsymbol{M_{CS}} = (E, O, R), E \neq \emptyset, O \neq \emptyset, R \subseteq E \times \ldots \times E, R \neq \emptyset \qquad (3.2)$$

where 1) E is the domain of $\boldsymbol{M_{CS}}$ and denotes a non-empty set of entities contained in the data processing procedure/system, 2) O denotes the operations on E, and it is a non-empty indexed set, as data processing always yields operations on data in case of computer science, 3) R is a non-empty indexed set which denotes the relations on E. Notice the highlighted specified part in formula (3.2) comparing to (3.1). Similarly, if the interpretation of the elements of $\boldsymbol{M_{CS}}$ holds with the modelled data processing procedure/system, $\boldsymbol{M_{CS}}$ is a model of it.

3.1.2 The meta-model identifying the context

As mentioned at the beginning of this chapter, now that a meta-model is derived for the general scientific model in computer science, the follow-up step is to reveal the context in it. Based on the preceding definitions of modelling and context, the concept of *context modelling* can be derived within the academic discipline of computer science and regarded as the modelling approach with *context awareness*, i.e. the resulting model recognises and takes into consideration the context of its target, so that it becomes flexible, adaptive and capable of autonomously handling the evolved context. To integrate the concept of context introduced in section 2.1, i.e. to make the model context-aware, it is necessary for the domain of $\boldsymbol{M_{CS}}$ to cover at the same time the modelling target and its context in the modelling scenario. In other words, for a context model, an entity within the scope of the modelling target can be either target entity (TE) or context entity (CE). Furthermore, as O and R together actually describe the connections among the entities, they can both be regarded as context.

Therefore, a context model $\boldsymbol{CM}$ in computer science is defined using following meta-model:

$$\begin{gathered}\boldsymbol{CM} = (E, O, R), E \neq \emptyset,\ \ O \neq \emptyset, R \subseteq E \times \ldots \times E, R \neq \emptyset \\ E = \{E_T, E_C\}, C = \{E_C, O, R\}\end{gathered} \tag{3.3a}$$

while the sets E_T and E_C respectively denoting TE and CE and C being the generalised context. Notice the highlighted specified part in formula (3.3a) comparing to (3.2).

To further elucidate the definition given in (3.3a), an example is provided here in the scenario of evidence processing pipeline: suppose a latent fingerprint is first acquired from a crime scene, and then analysed in the forensic lab, then in the end it enters an evidence archive before it can be used as court evidence. In an extremely simplified case, i.e. the evidence archive contains only this piece of evidence and only very basic forensic steps are deployed on it, the processing pipeline can be described as follows, based on (3.3a):

Example 1. *Let $\boldsymbol{CM}' = (E', O', R')$ denote the evidence archive, then an exemplary entity set of E' of $\boldsymbol{CM}'$ can be further specified as:*

$E' = \{E'_T, E'_C\}, E'_T = \{e_t\}, E'_C = \{e_{c0}, e_{c1}, \ldots, e_{ci}\}, i = 0,1, \ldots, 11$
e_t refers to the fingerprint
e_{c0} refers to the time of acquisition
e_{c1} refers to the location of acquistion
e_{c2} refers to the crime scene investigator who acquired the fingerprint
e_{c3} refers to the crime case the fingerprint is related to
e_{c4} refers to the time when the evidence arrives in the forensic lab
e_{c5} refers to the address of the forensic lab
e_{c6} refers to the fingerprint expert who analysed the fingerprint
e_{c7} refers to the identification of the person of interest that the fingerprint leads to
e_{c8} refers to the address of the evidence archive
e_{c9} refers to the time the fingerprint enters the archive
e_{c10} refers to the location of the fingerprint in the archive
e_{c11} refers to the chain-of-custody document of the fingerprint

while its corresponding operation set O' and relation set R' can be specified as:

$O' = \{o_0, o_1, \ldots, o_n\}, R' = \{r_0, r_1, \ldots, r_n\}, n = 0,1, \ldots, 7$
o_0 refers to responding, $r_0 = (e_{c2}, e_{c3})$
o_1 refers to fingerprint acquisition, $r_1 = (e_{c2}, e_t, e_{c0}, e_{c1})$
o_2 refers to entering, $r_2 = (e_t, e_{c5}, e_{c4})$
o_3 refers to fingerprint analysis, $r_3 = (e_{c6}, e_t)$
o_4 refers to yielding, $r_4 = (e_t, e_{c7})$
o_5 refers to entering, $r_5 = (e_t, e_{c8}, e_{c9})$
o_6 refers to documenting,
$r_6 = (e_t, e_{c11}, e_{c0}, e_{c1}, e_{c2}, e_{c4}, e_{c6}, e_{c9}, e_{c10})$
o_7 refers to storing, $r_7 = (e_t, e_{c11}, e_{c10})$

As Example 1 shows, a context model (in this case $\boldsymbol{CM}'$) always covers its TEs (in this case E'_T, which contains only one element e_t) and the CEs to their corresponding TEs (in this case E'_C, which covers i items that serve as the CEs to e_t). The connections between the TEs and CEs are identified by the operations (in this case O') together with the relations (in this case R'). Both indexed sets comprises n elements, so the nth element of R' identifies the involvement of TE and CEs in a connection, while its corresponding nth element of O' defines the nature of this connection. Notice here while the semantics in this example is clearly defined, the syntactic rules on the expression of the relations remain open, i.e. r_n is only defined in general as a finitary relation, whereas its grammar of expression is not regulated. This flexibility is considered as necessary for such definition on a rather abstract level. Also notice neither o_n nor r_n is supposed to be atomic, i.e. an operation or relation element defined here can be further split. This is also shown later in this subsection.

To further validate the sufficiency of the meta-model of context model in (3.3a) regarding its general representability of various types of context, here the nine-class framework introduced in [Lee2011] (see subsection 2.1.2) is selected due to its generality, and the meta-model can be further specified to cover each of its nine different context classes:

- As either actual or virtual item to be modelled, a context entity in E_C can appear as either *object*, *agent*, or *concept*, depending on the restraining context that comes with it. Furthermore, as *time* and *space* also possess specific nature of existence, the coverage of E_C can be extended on these two classes. In other words:

  $$E_C = Object \cup Agent \cup Concept \cup Time \cup Space \tag{3.3b}$$

  For instance, in Example 1, among all the entities identified in E', e_t and e_{c11} are classified as *object*, e_{c2} and e_{c6} as *agent*, e_{c3} and e_{c7} as *concept*, e_{c0}, e_{c4}, and e_{c9} as *time*, e_{c1}, e_{c5}, and e_{c8} as *space*.

- As *relationship* describes general association between two or more context entities, it can be described by an unspecific operation and its corresponding relation, i.e. an operation element of indexed set O describes the nature of the association, and a relation element of indexed set R points out the involved entities in E, as shown as follows:

  $$\begin{aligned} &Relationship = \{(o_n, r_n) | o_n \in O, r_n \in R, n \geq 0, n \in \mathbb{N}\} \\ &r_n = (e_0, e_1, \ldots, e_i), e_0, e_1, \ldots, e_i \in E, i \geq 0, i \in \mathbb{N} \end{aligned} \tag{3.3c}$$

  where r_n denotes an element in R, demonstrating an i-ary relation on E.

  In Example 1, the relationship in $\boldsymbol{CM'}$ is simply

  $$Relationship' = \{(o_n, r_n) | o_n \in O', r_n \in R', n = 0,1, \ldots ,7\}$$

- Within the scope of the meta-model, *occurrence* can be regarded as a special case of relationship described by as follows:

  $$\begin{aligned} &Occurrence = \{(o_n, r_n^O) | o_n \in O, r_n^O \in R, n \geq 0, n \in \mathbb{N}\} \subseteq Relationship \\ &o_n \text{ refers to specific process or event} \\ &r_n^O = \begin{cases} (e_{sub}, e_{obj}, e_{time}, e_{space}) & \text{in case of action} \\ (e_{sub}, e_{time}, e_{space}) & \text{in case of general phenomenon} \end{cases} \\ &e_{sub}, e_{obj} \in E, e_{time} \in Time \subset E_C, e_{space} \in Space \subset E_C \end{aligned} \tag{3.3d}$$

  In the formula, o_n refers to the specific process or event, e_{sub} , e_{obj} , e_{time} , and e_{space} respectively denote subject entity, object entity, time entity, and space entity. Therefore, $o_n$and r_n^Otogether can describe either an action, in terms of that e_{sub}takes an action (specified by o_n) on e_{obj}at the time of e_{time} and at the location of e_{space}, or a general phenomenon, in terms of that e_{sub}occurs (specified by o_n) at the time of e_{time} and at the location of e_{space}. Both e_{sub} and e_{obj} can be either TE or CE, and can also be extended to a series of entities of their kinds, while e_{time} and e_{space} sometimes can be omitted, depending on the actual situation.

In Example 1, among all the 2-tuples in $Relationship'$, an example of occurrence is the (o_1, r_1), which describes an action that the crime scene investigator (denoted by e_{c2}) acquires (specified by o_0) a fingerprint (denoted by e_t) at the time of e_{c0} and at the location of e_{c1}. Similarly, among the rest of the elements in $Relationship'$, the tuples (o_2, r_2), (o_3, r_3), (o_4, r_4), (o_5, r_5), (o_6, r_6), and (o_7, r_7) also describe actions.

- *Form of expression* can similarly be regarded as a special case of relationship:

$$\begin{aligned} &Form = \{(o_n, r_n^F) | o_n \in O, r_n^F \in R, n \geq 0, n \in \mathbb{N}\} \subseteq Relationship \\ &o_n \text{ refers to particular expressive forms} \\ &r_n^F = \begin{cases} (e_{sub}, e_{obj}) & \text{in case of expression} \\ (e_{sub}) & \text{in case of property} \end{cases}, e_{sub} \in E, e_{obj} \in E_C \end{aligned} \tag{3.3e}$$

so it can be used either to describe an expressive relation between entities as introduced in [Lee2011], or for an extension to the property of certain entity, which is not specified in [Lee2011] yet quite important. In the former case it describes that the subject entity $e_{sub}$is expressed with the object entity e_{obj}, which can also be extended to a series of entities of its kind if needed. In the latter case, r_n^F degenerates to a unary relation to identify the referred entity e_{sub}, while o_n specifies the property. An example of form of expression can be given here, extended from the evidence archive scenario which is already set in Example 1.

Example 2. *Let $e_{c12} \in E_C'$ denote the minutiae extracted from the fingerprint e_t, an exemplary form of expression can be defined as*

$$\begin{aligned} &form = (o_8, r_8), o_8 \in O', r_8 \in R' \\ &o_8 \text{ refers to biometric expression} \\ &r_8 = (e_t, e_{c12}) \end{aligned}$$

where the minutiae are connected to the corresponding fingerprint in the form of its biometric expression.

- As *purpose* mainly answers the "why" question, it can be induced to a special form of relationship between entity and concept, which in this case specifies the "mandates, norms, values, intentions, rules, standards, virtues, functions [Lee2011]" that explain for the entity, i.e.

$$\begin{aligned} &Purpose = \{(o_n, r_n^P) | o_n \in O, r_n^P \in R, n \geq 0, n \in \mathbb{N}\} \subseteq Relationship \\ &o_n \text{ refers to conformity} \\ &r_n^P = (e, e_c), e \in E, e_c \in Concept \subset E_C \end{aligned} \tag{3.3f}$$

An example of purpose can be given from Example 1 by the 2-tuple (o_0, r_0), which describes the crime scene investigator (denoted by e_{c2}) responds, because of the committed crime (denoted by e_{c3}).

- Furthermore, for a particular entity e to be modelled, it is possible that it has its own surrounding context and at the same time also serves as context for other entities, in other words:

$$\exists e \in E_T \cap E_C \tag{3.3g}$$

To demonstrate this situation, the scenario set in Example 1 and 2 is now further adapted:

Example 3. *Suppose the fingerprint e_t is acquired from the scene on its substrate, a piece of paper, and it enters the forensic lab together with its substrate, where both of* *them are examined. Let* e'_t *denotes the substrate,* $e_{c13} \in E'_C$ *the trace evidence expert. Then the analysis of fingerprint should now be described as*

$$o_3 \text{ refers to fingerprint analysis}, r_3 = (e_{c2}, e_t, e'_t)$$

where e_t *is trivially the TE, whereas* e'_t *serves as a CE, because the substrate needs to be considered in the selection of specific analysing method. Yet similarly, the analysis of the substrate for trace evidence can be described as*

$$o_9 \text{ refers to trace analysis}, r_9 = (e_{c13}, e'_t, e_t)$$

where this time e'_t *is the TE, while* e_t *serves as a CE, because the trace analysis method must be designed in the way that the fingerprint evidence is not compromised.*

Therefore, among all the entities defined in the two subsets of E'

$$E'_T \cap E'_C = \{e_t, e'_t\}$$

In summary, by flexibly specifying the modelled items denoted by E_T and E_C, the content of operation O as well as the arity and coverage of the relations in R, the meta-model proposed in formula (3.3) is capable of describing any context model in general. A typical embodiment of such meta-model would appear as follows:

$$\begin{aligned}
&\boldsymbol{CM} = (E, O, R), E \neq \emptyset, O \neq \emptyset, R \subseteq E \times \ldots \times E, R \neq \emptyset \\
&E = \{E_T, E_C\}, C = \{E_C, O, R\} \\
&E_T = \{e_t | e_t \text{ refers to target entities}\} \\
&E_C = \{e_c | e_c \text{ refers to context entities}\} \\
&O = \{o_0, o_1, \ldots, o_n\}, R = \{r_0, r_1, \ldots, r_n\}, n \geq 0, n \in \mathbb{N} \\
&r_n = (e_0, e_1, \ldots, e_i), e_0, e_1, \ldots, e_i \in E, i \geq 0, i \in \mathbb{N}
\end{aligned} \tag{3.4}$$

where the extension is highlighted comparing to the definition in (3.3a).

3.2 Security-oriented context modelling

As the meta-model is defined and also validated for general context modelling in computer science, this section further integrates the concepts regarding IT security into the meta-model developed in last section, deriving the concept of security-oriented context modelling. Subsection 3.2.1 specifies the security context in the meta-model, subsection 3.2.2 focuses on system level security and formulates security policies on various granularities, subsection 3.2.3 focuses on data level security and identifies various types of context in the both data acquisition and processing environments.

3.2.1 Recognising security context in the meta-model

As more and more security related requirements are raised in various IT application scenarios, the modelling applied in those scenarios subsequently involves more and more security relevant context. Therefore to emphasise the existence of such context, this dissertation proposes the concept of *security-oriented context modelling*, which is defined as the context modelling recognising security context applied for security assurance of large scale data processing systems. Specifying the existence of security context, the meta-model described in (3.4) can be further extended to describe the concept:

$$\begin{aligned}
&\boldsymbol{CM_S} = (E, O, R), E \neq \emptyset, O \neq \emptyset, R \subseteq E \times \ldots \times E, R \neq \emptyset \\
&E = \{E_T, E_C\}, E_C = \{E_{GC}, E_{SC}\} \\
&E_T = \{e_t | e_t \text{ refers to target entities}\} \\
&E_{GC} = \{e_{gc} | e_{gc} \text{ refers to general context entities}\} \\
&E_{SC} = \{e_{sc} | e_{sc} \text{ refers to security context entities}\} \\
&O = \{O_G, O_S\}, R = \{R_G, R_S\} \\
&O = \{o_0, o_1, \ldots, o_n\}, R = \{r_0, r_1, \ldots, r_n\}, n \geq 0, n \in \mathbb{N} \\
&r_n = (e_0, e_1, \ldots, e_i), e_0, e_1, \ldots, e_i \in E, i \geq 0, i \in \mathbb{N} \\
&C = \{E_C, O, R\}, C_S = \{E_{SC}, O_S, R_S\}
\end{aligned} \tag{3.5a}$$

where the highlighted part specifies the security-relevance in the definition: E_C is further split into general context entities E_{GC} and security context entities E_{SC}, O into general operations O_G and security-oriented operations O_S, R into general relations R_G and security-oriented relations R_S. E_{SC}, O_S together with R_S form the security context C_S.

Recalling the scenario presented in Example 1, based on (3.5a) now the elements in $\boldsymbol{CM'}$ can be further categorised according to their security relevance:

Example 4. *Suppose in such forensic scenario the security requirement covers only the processed fingerprint, i.e. its confidentiality, authenticity, integrity, availability, and non-repudiation, let* $\boldsymbol{CM'_s}$ *be the security-oriented context model, it can then be derived by clarifying the security relevance of the elements in* $\boldsymbol{CM'}$*:*

$\boldsymbol{CM'_s} = (E', O', R')$
$E' = \{E'_T, E'_C\}, E'_C = \{E'_{GC}, E'_{SC}\}$
$E'_T = \{e_t\}, E'_C = \{e_{c0}, e_{c1}, \ldots, e_{ci}\}, i = 0,1, \ldots, 11$
$E'_{GC} = \{e_{c5}, e_{c8}\}, E'_{SC} = \{ e_{c0}, e_{c1}, e_{c2}, e_{c3}, e_{c4}, e_{c6}, e_{c7}, e_{c9}, e_{c10}, e_{c11}\}$
e_t refers to the fingerprint
e_{c0} refers to the time of acquisition
e_{c1} refers to the location of acquistion
e_{c2} refers to the crime scene investigator who acquired the fingerprint
e_{c3} refers to the crime case the fingerprint is related to
e_{c4} refers to the time when the evidence arrives in the forensic lab
e_{c5} refers to the address of the forensic lab
e_{c6} refers to the fingerprint expert who analysed the fingerprint
e_{c7} refers to the identification of the person of interest that the fingerprint leads to
e_{c8} refers to the address of the evidence archive
e_{c9} refers to the time the fingerprint enters the archive
e_{c10} refers to the location of the fingerprint in the archive
e_{c11} refers to the chain-of-custody document of the fingerprint
$O' = \{o_0, o_1, \ldots, o_n\}, R' = \{r_0, r_1, \ldots, r_n\}, n = 0,1, \ldots, 7$
$O' = \{O'_G, O'_S\}, R' = \{R'_G, R'_S\}$
$O'_G = \{o_0\}, O'_S = \{o_1, o_2, o_3, o_4, o_5, o_6, o_7\}$
$R'_G = \{r_0\}, R'_S = \{r_1, r_2, r_3, r_4, r_5, r_6, r_7\}$
o_0 refers to responding, $r_0 = (e_{c2}, e_{c3})$
o_1 refers to fingerprint acquisition, $r_1 = (e_{c2}, e_t, e_{c0}, e_{c1})$
o_2 refers to entering, $r_2 = (e_t, e_{c5}, e_{c4})$
o_3 refers to fingerprint analysis, $r_3 = (e_{c6}, e_t)$
o_4 refers to yielding, $r_4 = (e_t, e_{c7})$
o_5 refers to entering, $r_5 = (e_t, e_{c8}, e_{c9})$
o_6 refers to documenting,
$r_6 = (e_t, e_{c11}, e_{c0}, e_{c1}, e_{c2}, e_{c4}, e_{c6}, e_{c9}, e_{c10})$
o_7 refers to storing, $r_7 = (e_t, e_{c11}, e_{c10})$

$$C = \{E_C, O, R\}, C_S = \{E_{SC}, O_S, R_S\}$$

As shown above, among all the CEs, the address information of forensic lab (e_{c5}) and that of evidence archive (e_{c8}) is classified as general context, as it is not directly related to the security assurance of the TE. All the other CEs are classified into security context, e.g. the time and space information constitutes the chain-of-custody, which is directly related to integrity and non-repudiation, the human agents, i.e. the crime scene investigator and fingerprint experts, both are related to the authenticity of the fingerprint. Similarly, the operations and relations are also classified in this way.

As already established in section 1.3, a well-developed data processing system should achieve its security on two levels: on the system level the security requirements are analysed and clarified, then on the data level specific processing procedures are realised to meet those security requirements. Therefore despite its nature being consistent as described in (3.5a), the content of C_S varies on system and data levels, and this is further discussed in later subsections.

Based on the difference content of security context on system and data levels, the security-oriented context modelling also plays different roles on different levels in their contribution to the overall security assurance of the whole data processing system.

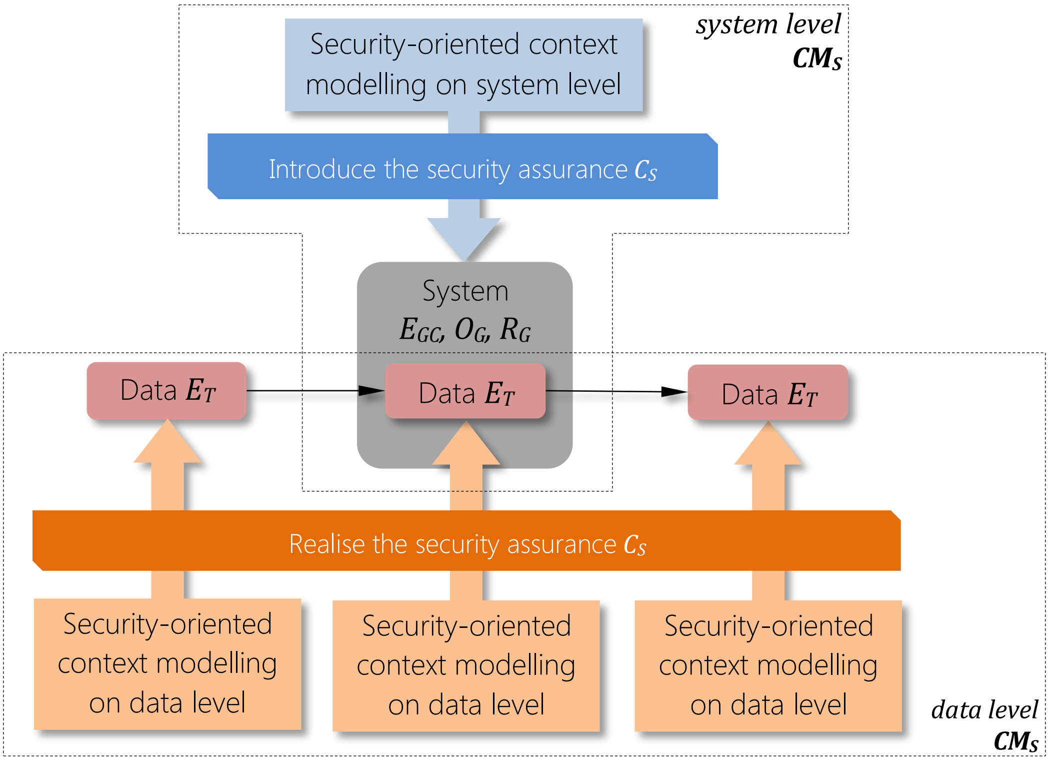

system level
CM_S
Security-oriented context modelling on system level
Introduce the security assurance C_S
System
E_{GC}, O_G, R_G
Data E_T
Data E_T
Data E_T
Realise the security assurance C_S
Security-oriented context modelling on data level
Security-oriented context modelling on data level
Security-oriented context modelling on data level
data level
CM_S

Figure 8. Security-oriented context modelling contributes on both levels to the security assurance of data processing system

As Figure 8 illustrates, the aim of system level security-oriented context modelling is to identify and clarify the context C_S on the system level, so the security assurance can be introduced via C_S either to improve an already existing system model or in the early stage of the development of a system model. In the meantime, data level security-oriented context modelling is applied to identify and interpret security context C_S on the data level, so the data processing procedures can be better designed, organised and implemented to eventually realise the security assurance. Embracing the different types of entities defined in (3.5a), in a digital data processing system, the data is always considered as E_T, and while its E_{GC} together with O_G, R_G is usually introduced within the scope of the system, its E_{SC}, O_S, and R_S are introduced by security-oriented context modelling on both system and data levels.

The following subsections describe how security context is identified, extracted, organised, and expressed on system and data levels, respectively.

3.2.2 Contextualise the security on the system level

As defined previously, system level security focuses on identifying and clarifying the security requirements with regard to various system components. Since security policy is defined as statement of what is allowed and what is not [Bishop2002] to regulate what in the system should be protected [BS2002], it serves ideally to describe the required security assurance on various granularities in a system. To better clarify the content of security context on different granularities, the following series of concepts are derived from the original definition of security policy, reflecting various context classes described in subsection 3.1.2.

On the level of coarsest granularity, *abstract security policy* (ASP) can be defined as

$$\begin{aligned}
&\boldsymbol{ASP} = (E, \{o\}, \{r_S\}) \\
&E = \{E_T, E_C\}, E_C = \{E_{GC}, E_{SC}\} \\
&E_T = \{\text{the overall system}\} \\
&E_{SC} = \{\text{authenticity, integrity, confidentiality, availability, non-repudiation}\} \\
&o \text{ refers to conformity} \\
&r_S \in R_S \subseteq E_T \times E_{SC}
\end{aligned} \tag{3.5b}$$

specifying the definition in (3.5a) by integrating the context class of purpose described in (3.3f). By definition, abstract security policy simply identifies the security context C_S by regulating the security aspects that the overall data processing system needs to be in conformity with. Despite its simplicity, the abstract security policy implicitly regulates the conditions under which the data objects are processed. Taking the forensic scenario set in Examples 1, 2, 3, and 4 for instance, if the forensic processing pipeline being modelled is regarded as a system, then an exemplary ASP for this system can be:

Example 5. ASP1: *The forensic system shall be in conformity with data authenticity.*

then the forensic system, which is always the TE in abstract security policy, shall be at least designed in the way as the CE requires, in this particular policy i.e. the provenance/authenticity information of the data objects (i.e. the fingerprint evidence in the scenario set in the examples) shall be processed, and they should even be actively authenticated on regular basis.

It is expected that the proper processing of data objects involves interactions among different objects (including data objects and system components), sometimes human agents or even virtual concepts, restrained by time and/or space. Therefore, *specific security policy* (SSP) is subsequently defined to regulate those interactions in the forms of security-oriented occurrence or form of expression (see (3.3b-e)). It is formulated as follows:

$$\begin{array}{l}
\boldsymbol{SSP} = (E, O_S, R_S) \\
E = \{E_T, E_C\}, E_C = \{E_{GC}, E_{SC}\} \\
E_T = \left\{ e_t \middle| \begin{array}{c} e_t \text{ refers to target system components, data objects,} \\ \text{human agents, or virtual concepts.} \end{array} \right\} \\
E_{GC} = \{ e_{gc} | e_{gc} \text{ refers to relevant system components, time, space, data objects, etc.} \} \\
E_{SC} = \{ e_{sc} | e_{sc} \text{ refers to security-oriented system components, data objects, etc. } \} \\
O_S = \{ o_s | o_s \text{ refers to security-oriented processes, events, or expressive forms. } \} \\
R_S \subseteq E \times \ldots \times E
\end{array} \tag{3.5c}$$

In formula (3.5c), o_s denotes a security-oriented operation which reflects requirement from abstract security policies. e_t refers to the TE of the operation, which can be either object (either data object or system component), human agent, or virtual concept. e_{sc} is security context entity (SCE) that developed only for security assurance, and it can similarly appear as either object, agent, or concept. e_{gc} is general context entity (GCE), which is involved with the operation yet exists for non-security related system functionality. Therefore besides object, agent, and concept, it can cover time and space as restraining factors.

Taking the previous ASP_1 for example, a possible specific security policy reflecting it can be:

Example 6. *Suppose a human archive managing agent takes charge of all the actions within the evidence archive, an exemplary SSP derived from* ASP_1 *can be:*

SSP_1: *When the fingerprint enters the evidence archive, the archive managing agent shall store it in an indexed location and assures its authenticity by updating its chain-of-custody document with this location information.*

Let e_{c14} denote the human archive managing agent, interpreting SSP_1 with regard to formula (3.5), Table 6 analyses the semantics of SSP_1, using notations established in Examples 1, 2, 3, and 4, and the entity types are also marked in brackets according to the nine-class classification from [Lee2011].

Context entities					*Target entities*
r	e_{gc}		e_{sc}	o_s	e_t
(e_{c14}, e_t, e_{c9})	Entering time e_{c9} (time)		Archive managing agent e_{c14} (human agent)	store o_7	Fingerprint e_t (object)
$(e_{c14}, e_t, e_{c9}, e_{c11})$	Chain-of-custody document e_{c11} (object)	Entering time e_{c9} (time)		documento_6	Fingerprint e_t (object)

Table 6. Identify the context and target entities of an exemplary specific security policy

The definition of SSP bears high flexibility regarding its granularity, therefore suitable for systems with different scales and complexities. In fact, when the complexity of a data processing system reaches a certain level, it is necessary to introduce certain mechanism, e.g. a policy hierarchy, to organise the large number of SSPs. This is further explained and implemented in Chapter 4.

Comparing to ASP, SSP in general encompasses more details, but its flexible structure constrains its potential to be directly executed. Therefore, on the finest granularity the concept of *security rule* (SR) is defined, as an atomic SSP that describes strictly one operation targeting one TE from one SCE, if necessary restrained by GCE/GCEs, described as follows:

$$\begin{aligned}
&\boldsymbol{SR} = \{E, \{o_s\}, \{r_s\}\}\\
&E = \{e_{sc}, e_t, e_{gc0}, e_{gc1}, \dots, e_{gci}\}, i \geq 0, i \in \mathbb{N}\\
&e_{gc0}, e_{gc1}, \dots, e_{gci} \in E_{GC} \subset E_C, e_{sc} \in E_{SC} \subset E_C, e_t \in E_T\\
&o_s \in O_S\\
&r_s = (e_{sc}, e_t, e_{gc0}, e_{gc1}, \dots, e_{gci}), i \geq 0, i \in \mathbb{N}
\end{aligned} \tag{3.5d}$$

Based on this definition, an exemplary SR can be easily derived from SSP$_1$ directly:

Example 7. SR$_1$: *When the fingerprint enters the evidence archive, the archive managing agent shall store it in an indexed location.*

Similar to Table 6, the following Table 7 identifies the semantics in SR$_1$:

Context entities				*Target entities*
r	*e*$_{gc}$	*e*$_{sc}$	*o*$_s$	*e*$_t$
(e_{c14}, e_t, e_{c9})	Entering time e_{c9} (time)	Archive managing agent e_{c14} (human agent)	store o_7	Fingerprint e_t (object)

Table 7. Identify the context and target entities of an exemplary security rule

As shown in Table 7, SR$_1$ strictly projects one operation on one TE, describing an atomic action. Comparing to SSP$_1$, it gains a huge advantage with regard to the simplicity for further execution. In the case of a data object as TE of the SR (like SR$_1$), the SR can be regarded as an atomic data processing procedure. Therefore, all such SRs together describe the overall data processing environment in the system, serving as the basis for the further data level security-oriented context modelling.

Summing this subsection up, to introduce the security assurance to a data processing system, the system level security-oriented context modelling contextualises the security requirements on various levels of granularity with a top-down structure. As illustrated in Figure 9, security-oriented context modelling takes the system level security context, which comes in a rather raw form (hence covered by general sets of E, O, and R in the figure) as input. Then it uses ASP to identify the general security aspects for the overall system on the coarsest granularity, SSP to clarify the specified security operations with regard to specified roles in the system on intermediate granularities, and SR to regulate the atomic actions to be executed on the finest granularity. As SR comprises atomic operations involving individual entities, i.e. (o_s, r_s) that connects all *e*$_t$, *e*$_{gc}$, and *e*$_{sc}$ (see (3.5d)), it directly describes the processing environment (see the definition of ***PE*** in (3.5e) in subsection 3.2.3) and serves as the basis of data level analysis. A further instantiation of system level security-oriented context modelling is described in Chapter 4.

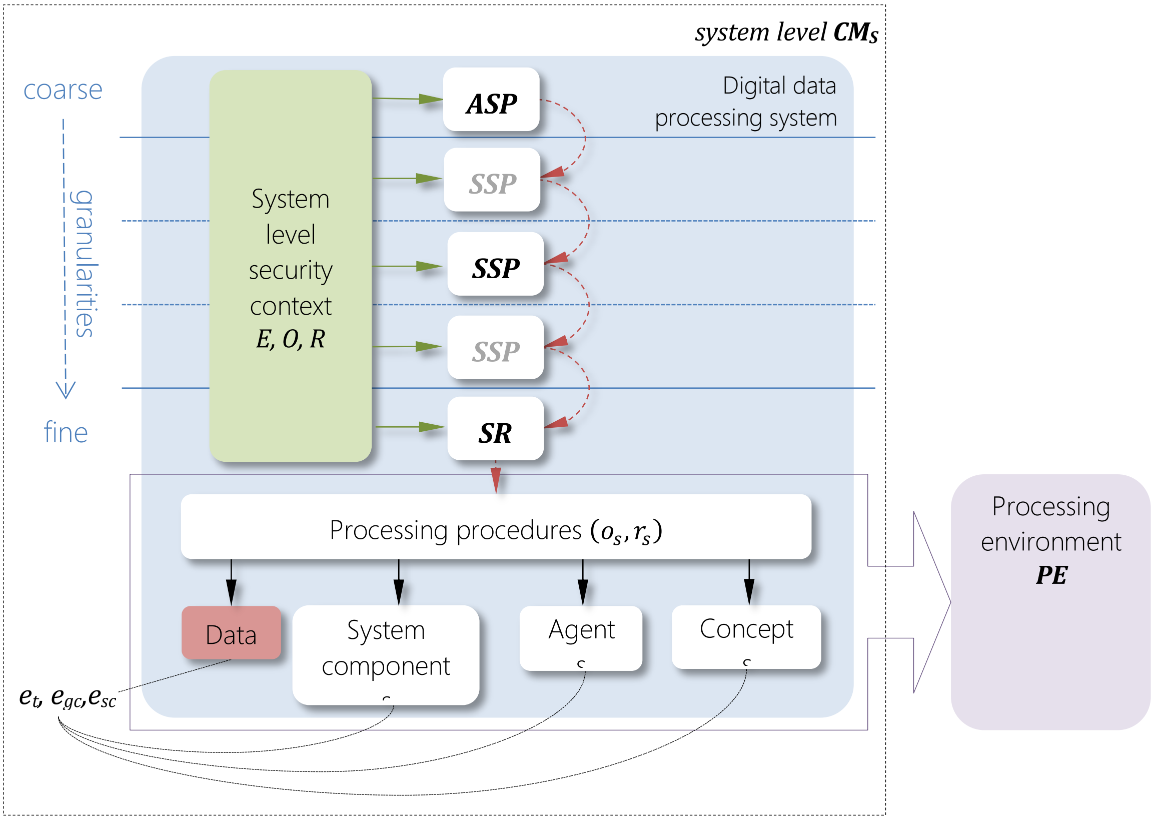


Figure 9. System level security-oriented context modelling introduces security assurance using security policies on various granularities

3.2.3 Contextualise the security on the data level

While the security assurance of the system is introduced on the system level by the development of security policies on various granularities, it is realised on the data level by specific mechanisms. As defined in last subsection, SRs on the finest granularity on the system level yield processing procedures on the TEs, including the data objects. Therefore, for a particular data object that the system processes, its processing environment can be described by clustering all the SRs with this object as TE (see also Figure 9):

$$\begin{aligned}
&\boldsymbol{PE} = \{SR_0, SR_1, \dots, SR_m\}, m \geq 0, m \in \mathbb{N} \\
&SR_m = (E_m, \{o_{sm}\}, \{r_{sm}\}) \\
&E_m = \{e_t, e_{sc}^m, e_{gc0}^m, e_{gc1}^m, \dots, e_{gcn}^m\}, n \geq 0, n \in \mathbb{N} \\
&e_t \text{ refers to data object} \\
&C_p = \{(E_0 \cup E_1 \cup \dots \cup E_m) \setminus \{e_t\}, \{o_{s0}, o_{s1}, \dots, o_{sm}\}, \{r_{s0}, r_{s1}, \dots, r_{sm}\}\}
\end{aligned} \tag{3.6a}$$

where ***PE*** refers to the processing environment of the data object e_t, and its context in this environment is defined as *processing context*, which is denoted as C_p.

Besides the processing environment, for some systems it is also necessary to consider the acquisition environment. When the data object is acquired before it enters the processing environment, it is usually expected that additional information is also acquired simultaneously. Such information is defined here as *acquisition context*. Let ***AE*** be the acquisition environment, it can be formalised as the following 3-tuple:

$$\begin{aligned}&\boldsymbol{AE} = (E_A, O_A, R_A)\\&o_a \in O_A, o_a \text{ refers to acquistion}\\&R_A \subseteq E_A \times \ldots \times E_A\\&E_A = \{e_t\} \cup E_{object} \cup E_{envi}\\&C_a = \{E_{object} \cup E_{envi}, O_A, R_A\}\end{aligned} \tag{3.6b}$$

where the domain E_A comprises three subsets: the target entity, in this case the data object e_t being processed, together with E_{object}, which denotes the context entities that represent the property of e_t, and E_{envi}, which denotes the context entities introduced to e_t by the acquisition environment itself. For instance, in the forensic scenario set in earlier examples, if a fingerprint is acquired in a crime scene using a digital camera, its own properties like size or clarity of details would constitute E_{object}, while the further information introduced by the acquisition environment like temperature, humidity, substrate, or camera settings would constitute E_{envi}. The operation set O_A not only covers the specific action of acquisition (denoted here by o_a) but also specifies further relationships existing among the entities in E_A, while the relation set R_A identifies the involved entities in those relationships. Therefore, regarding the acquired data object e_t, its *acquisition context* is defined as C_a in the formula.

The context in processing and acquisition environment can be further specified, based on the categorisation from Perera et al. in [PZC+2014]. In the processing environment, the GCEs and SCEs that come directly from the environment belong to the *primary processing context* with regard to the processed data object. Subsequently, the *secondary processing context* can be derived by further interpreting the primary one (see Chapter 5 for more details). Additionally, the processing environment can be organised differently based on different GCEs in the primary processing context. For instance, based on time and/or space entities, a processing environment can appear as temporal/geographical processing pipeline. Similarly, in the acquisition environment, *primary acquisition context* is defined as the additional information which is directly collected during the acquisition of the data object and covers in general all entities that are relevant to the concerned application scenario in the acquisition environment. Despite that it does not necessarily interact with the data object immediately in the phase of acquisition, it enters the processing environment together with the data object. Therefore, similar to the secondary processing context, the *secondary acquisition context* can also be derived in the processing environment by interpreting the primary acquisition context. The two kinds of secondary context together yield the execution of the processing procedures, which directly interact with the processed data object.

Figure 10 illustrates contextualisation of the security on the data level. In the acquisition environment, the primary acquisition context is generated simultaneously with the acquisition of the data object. After it enters the processing environment together with the acquired data object e_t, the secondary acquisition context is derived from it. In the meantime, the primary processing context is generated by clustering the related SRs from the system level context modelling, sometimes also by further interpreting the secondary acquisition context. The secondary processing context is then derived from the primary one. Therefore, the two kinds of secondary context are together used to guide the specific execution of the processing procedures on the data object e_t. In this way, not only can the security assurance introduced by the system level modelling be fully realised on the data level, additionally it gains the potential of addressing non-repudiation for the system, as the data level context model possesses the ability of clarifying all the interaction between the concerned data object and other system entities. Chapter 5 describes a further instantiation of the data level security-oriented context modelling, especially on how to interpret primary context to derive secondary context.

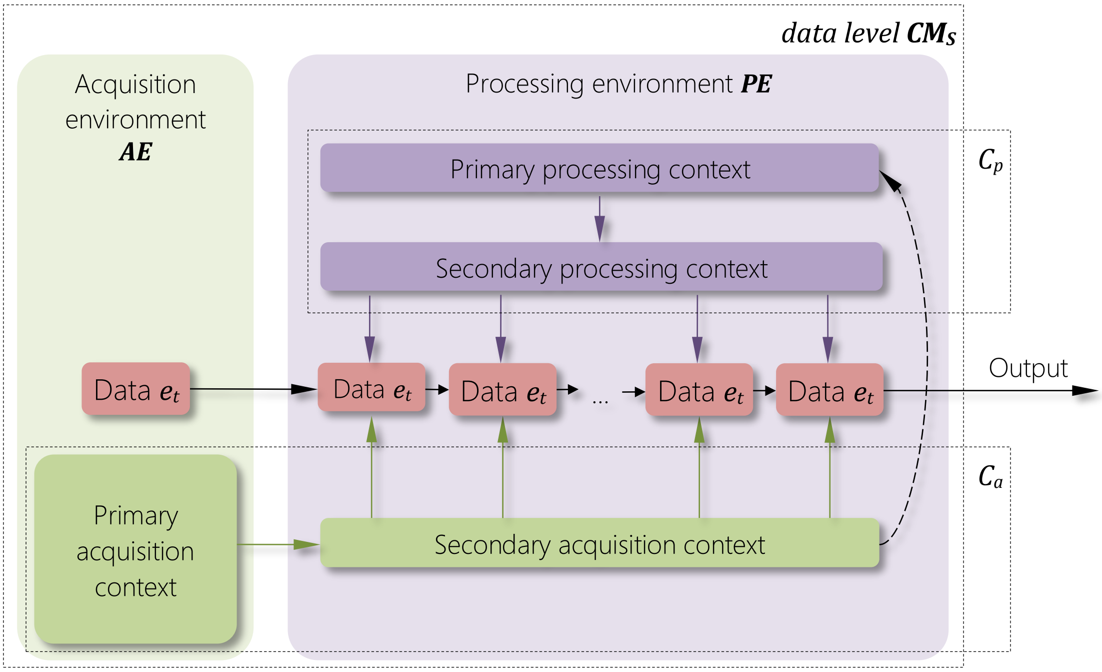

data level CM_S
Acquisition environment AE
Processing environment PE
Primary processing context
Secondary processing context
C_p
Data e_t
Data e_t
Data e_t
...
Data e_t
Data e_t
Output
C_a
Primary acquisition context
Secondary acquisition context

Figure 10. Data level security-oriented context modelling realises security assurance by collecting primary context and derive it to secondary context (adapted from [QSZ+2014])

3.3 Descriptive scheme of security-oriented context models

The last two sections elucidate the conception of security-oriented context modelling. Following the conception, security-oriented context models can be derived. However, to close the last gap between the derived context models and the application of them for data processing systems or procedures, this section proposes a descriptive scheme. As shown in Figure 11, the descriptive scheme of security-oriented context model comprises the following four parts:

- *Identifier* describes the application scenario that the context model is planned to be used for.
- *Version* describes the status of the context model, as the model evolves with the evolvement of its context, resulting in an iterative modelling process.
- *Security-oriented context model* is the result of the modelling technique, and as shown respectively in Chapter 4 and 5 in two different yet connected application scenarios, it can either be a system level model or a data level one.
- *Evaluation metrics* cover two aspects: the *scenario-unspecific evaluation metrics* assess the quality of a context model, regarding its universal properties irrelevant to the nature of its application scenario, whereas *the scenario-specific evaluation metrics* assess the performance of the context model, i.e. the data processing system that the model yields on the system level, or the data processing procedure it yields on the data level, both with regard to its specific application scenario. According to subsection 2.2.3, the scenario-unspecific metrics are yet to be formed, thus this is addressed in Chapter 6. As introduced in subsection 2.4.2 and 2.5.2, the scenario-specific metrics vary from approach to approach and scenario to scenario. The relationship between the application scenario and the two aspects of evaluation metrics is also further explored in Chapter 6.

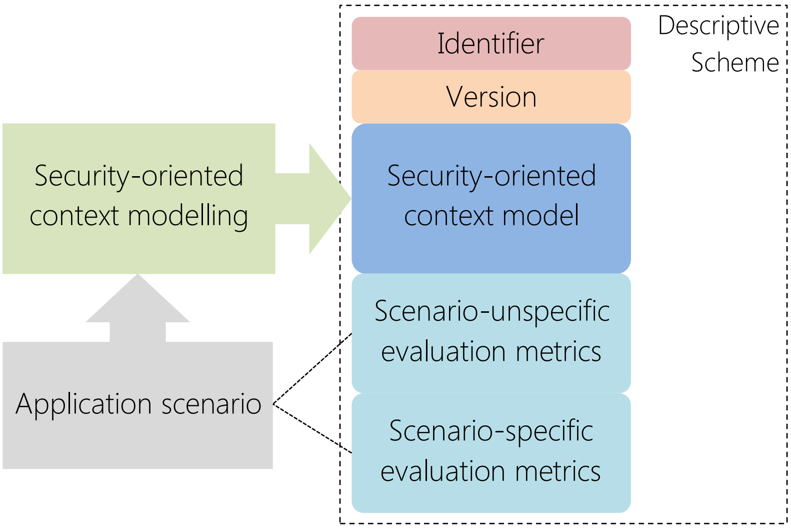

Identifier
Version
Descriptive Scheme
Security-oriented context modelling
Security-oriented context model
Scenario-unspecific evaluation metrics
Application scenario
Scenario-specific evaluation metrics

Figure 11. Descriptive scheme of security-oriented context model

3.4 Summary

This chapter depicts the theoretical framework of security-oriented context modelling proposed in this dissertation. The framework comprises a series of concepts, which are founded on existing definitions on general modelling and enriched by gradually introducing further related ones regarding both context and IT security. The resulting meta-model covers all the general context types identified in the state of the art, and it also handles IT security on two different levels: the security assurance is introduced on the system level, and then realised on the data level. On the system level, security policies on various granularity levels are derived from system level security context as the carrier of specific security requirements, yielding processing procedures of various entities in the system. On the data level, security context are collected and interpreted from both acquisition and processing environments, yielding specific operations on the processed data objects. As the "hook" connecting both levels, the security policy on the finest granularity level, namely the security rule, is on one hand the expression of atomic actions derived from system level security context, and on the other hand the specific requirement to be implemented on the data level, contributing to the processing environment as well as the data level security context in it. Furthermore, this chapter also proposes a descriptive scheme for the security-oriented context model, clarifying the role of its evaluation metrics.

The framework derived in this chapter answers to research challenge 1) with the formalised meta-model, and it also serves as the theoretical basis of the general methodology, which answers to research challenge 2). Despite it does not directly answer to the research challenge 3) regarding the evaluation issue, a descriptive scheme additional to the framework is proposed to identify the association of the evaluation metrics to the model, as well as the functionality of the metrics summarised in two aspects.

4 CONTEXT MODELLING ON THE SYSTEM LEVEL

This chapter describes a system level instantiation of the security-oriented context modelling in the application scenario of digital long-term archiving. Such instantiation directly answers to the two events of security compromise in digital archiving introduced in Chapter 1. The theoretical framework derived in Chapter 3 is applied in this scenario, resulting in a system-level context model of an archiving environment, which is capable of securely preserving data for a relatively long-term of time. The model is based on the OAIS standard and specific use-cases and derives security context from them. The security context is then further interpreted, so that corresponding system components can be derived to introduce security assurance through a security policy hierarchy.

The evolving digital era presents an unusual challenge for law enforcement: not only more and more evidence is found digital, due to the ever-increasing involvement of digital equipment in the crimes, but a growing amount of digitised evidence also emerges with the development of modern forensic technology. Therefore it is becoming a salient concern not only regarding the collection and processing of such electronic evidence but also its storage and long-term preservation. Despite the fact that such concern has been raised in U.S. at least since 2006 [Schilling2006], until recently the law enforcement personnel there is still guided to preserve electronic evidence in the conventional way, i.e. transferring digital information onto a CD-R or other write-once read-many (WORM) media and preserve it in a chain of custody, just like other physical evidences [Cameron2011]. Similar preservation means are also enforced by law enforcements in China [NPC1996] [SAA1999] [SAA2012] and Germany [RFJ2007]. Applying conventional preservation methods on electronic evidence fulfils certain demand for security, yet it has some severe drawbacks [Garfinkel2010]. First, the storage medium itself, like CD-R, is vulnerable and prone to damage. Second, the hardware and software interface for the storage medium can become obsolete in the case of long-term preservation. Third, the formats in which the digital data is stored can also become obsolete in the case of long-term preservation. At last, neither the currently used hardware nor the software is intended for efficient management of huge amounts of electronic evidence. All the above drawbacks would eventually lead to the inaccessibility of the data, which can cause fatal consequences if the preserved data is a significant piece of electronic evidence, and this is unfortunately already reflected by the dismissal DEA case against the fugitive doctor, as described and analysed in Chapter 1. To some degree, employing naive preservation strategies is fatal as much as no preservation strategies at all. For example, the WORM (write-once read-many) concept gets effectively undermined by simply copying data from one obsoleting storage device to an up-to-date one, as from the point of view of data the "write-once" principle is violated here.

Therefore, it is necessary to develop a digital long-term preservation environment, which meets certain security requirements in managing large amounts of data. This necessity is almost salient, given the lack of endeavour towards this direction reflected by the state of the art (see section 2.4). As a first step towards such environment, based on the state-of-the-art achievement within the scope of SHAMAN (Sustaining Heritage Access through Multivalent AchiviNg, see subsection 2.4.2), this chapter proposes an OAIS (Open Archival Information System)-compliant framework which applies system level security-oriented context modelling as solution for security policy generation, implementation and enforcement: Section 4.1 presents the design of the framework for contextualisation and policy-based security realisation. Section 4.2 demonstrates the application of the framework by applying it to one functionality entity of the archival system. Section 4.3 demonstrates a security assessment based on selected criteria from SHAMAN assessment framework. Section 4.4 summarises the whole chapter.

The research presented in this chapter was sponsored within the scope of SHAMAN project[1], and parts of it have been published in [QSK+2011a], [SQK+2011], and [QSK+2011b].

4.1 Design of the contextualisation framework

This section describes the proposed framework for contextualisation of security for digital long-term archiving. Considering the application scenario of the framework, there are some basic requirements: first, the framework must be able to introduce the security assurance in all aspects within a long-term archiving system for electronic evidence, especially those barely covered by the OAIS standard. Second, it must provide effective mechanism to manage a large quantity of security policies coming with a rather complex system. Third, it must at the same time serve as a seamless integration with the OAIS standard to ensure the accessibility and availability of all concepts and functions defined by the OAIS standard. At last, the framework must provide central management and audit service, similar to OAIS standard as well as any other complex systems.

As embodiment of the system level security-oriented context modelling explained in subsection 3.2.2, a contextualisation framework is developed with four major functional blocks, each aiming at one of the preceding four basic requirements [QSK+2011b]: a) *context analysis*, b) *policy hierarchy*, c) *Information Package (IP) processing*, and d) *central control*. The context analysis block consists of two distinct parts: global (system-wide) and local context analysis, both helping the system identify security-related requirements from its application scenario. The policy generation hierarchy is a hierarchy of stages beginning at the top with the generation of system-wide global policies and ending in the deviation and invocation of rules for IP processing operations. As shown later, the policy hierarchy is a satisfactory solution for large quantity of policies. The IP processing itself identifies the IPs (or related system data) to be processed, which can be various pieces of electronic evidence, and applies the rules as sequences of atomic data processing operations. Depending on their status in the archival system, IPs can be further specified as Submission IPs (SIPs) which are submitted to the archive, or Archival IP (AIPs) which are preserved in the archive, or Dissemination IPs (DIPs) which are disseminated from the archive on request. This block is based on the already existing functions defined by OAIS standard thus offers the seamless integration required. The control block controls the context analysis and the policy generation hierarchy and acts as a central policy repository as well as a central audit service for the overall system.

Figure 12 visualises the embodiment of the system level security-oriented context modelling in the application scenario of digital-term preservation discussed in this chapter, specifying the concepts in the theory using the entities appearing in the scenario. As shown in the figure, here three parts contribute to the system level security context: 1) the already established OAIS standard (see [OAIS2002] and [OAIS2009]), 2) the Information Lifecycle Model (ILM) developed within the scope of SHAMAN (see [BKJ+2010]), 3) the use-cases derived within the scope of SHAMAN. All three together cover the involved context domain E, as well as all the operations O and relations R on the domain. From the system level security context, a policy hierarchy is then derived, covering various granularity levels from the system to its module, then to the functional entities, and eventually to the

[1] Funded in part by the European Commission through the FP7 ICT Programme under Contract FP7-ICT-216736 SHAMAN.

functions, where SRs yield specific processing procedures (which can be described by 2-tuple (o_s, r_s) according to preceding definitions) on various entities processed on the function level, denoted by e_t, e_{gc}, and e_{sc} depending on the roles that the entities play in the specific procedures. The following subsections describe further in details how to model the context, as well as how to generate, implement, and enforce security policies based on use-cases from a data intensive, complex, security-oriented data processing system, like an archive for digital long-term preservation.

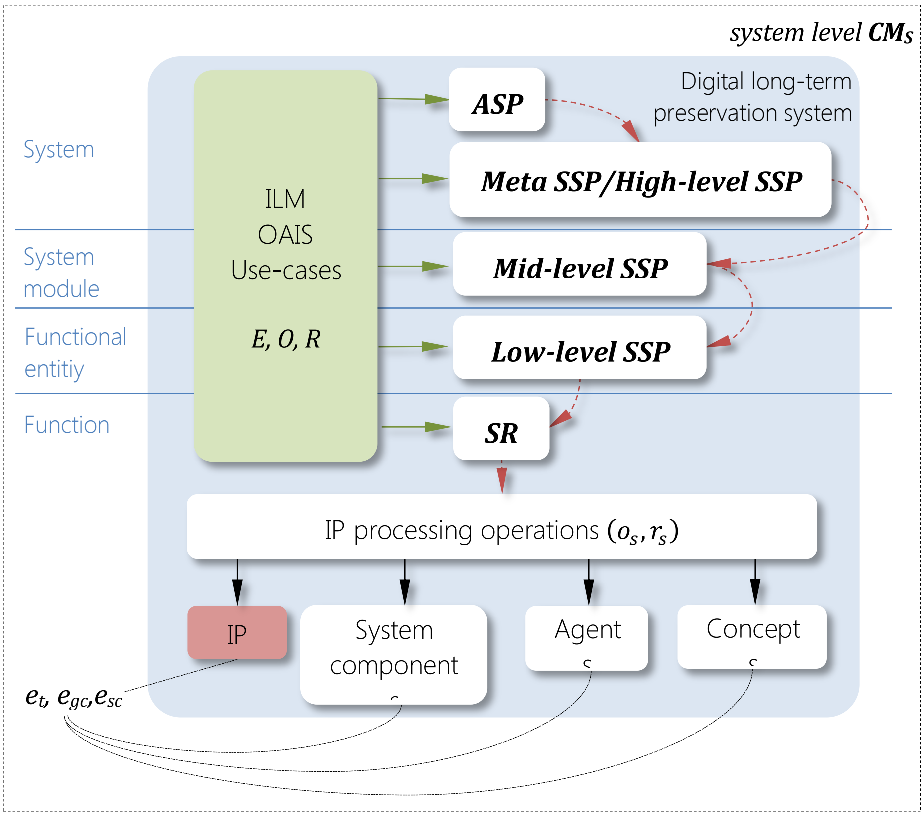


Figure 12. Embodiment of system level security-oriented context modelling in the application scenario of digital long-term preservation

4.1.1 System level context analysis and interpretation

A complex data processing system usually contains a large amount of components with different types of relationships among the components. Therefore, the "bottom-up" modelling approach is not suitable, as achieving a complete and vivid representation as a starting point in context modelling is not feasible under these circumstances. Instead, it is more appropriate to first extract typical functions (workflows) from use-cases in such systems and then gradually extend these into a fully developed context model, which clearly depicts the system components as well as relationships among them, expressed here in the form of policies and rules. In other words, the progress of system level security-oriented context modelling here is a progress of gradually specifying all the elements (i.e. e_t, e_{gc}, e_{sc}, o_s and r_s) in the system level context (i.e. E, O and R) introduced into the scenario at the beginning. As the first step towards in modelling progress, the system level context is analysed and interpreted here in this subsection.

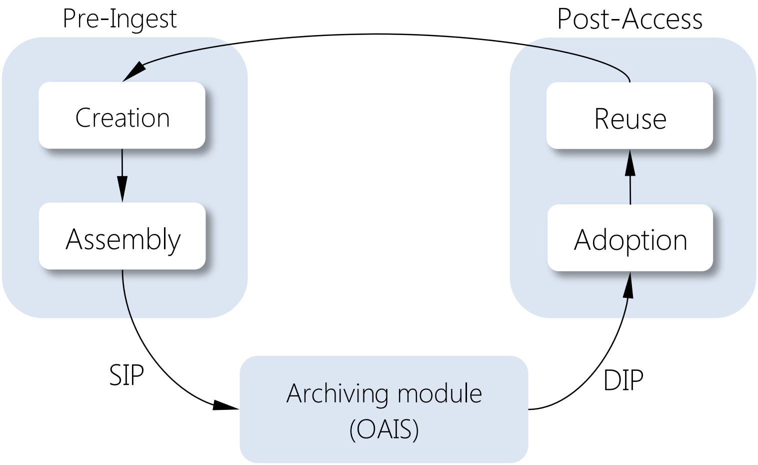


Figure 13. Basic structure of the archiving system based on the ILM introduced in [BKJ+2010]

As mentioned previously, Brocks et al. extended in [BKJ+2010] the OAIS model by introducing an extended ILM, as shown in Figure 13. In this the OAIS-compliant archiving module including all its functional entities are considered as one phase of the lifetime of a data object. The ILM extends this by including the "life" of data objects before and after being archived within the module. The phase before a digital object enters an archival system is called "Pre-Ingest", which further contains the actual *creation* of the data later to be ingested and its *assembly* into a package supported by the archive. The phase after a digital object leaves the module is called "Post-Access" and it contains *adoption* where the received data is unpacked, examined, transformed, displayed or in short all tasks that are needed for repurposing the content and *reuse* where the content is actually exploited. Notice that *reuse* may also include the re-ingest of this data object or a derivation thereof into an archival system, leading to a real lifecycle as shown in Figure 13. Such connection between *reuse* and *creation* is especially the case for collaborative environments. In the discussed scenario in this chapter, such ILM contributes to the system level context in the way of serving as the skeleton of the secure long-term archive to be developed.

As a further extension of the work of Brocks et al., in this chapter the considerations are limited on the central phase of the ILM, i.e. the functional entities within the OAIS-compliant archiving module, together with the corresponding security considerations. As one part of the system level context, OAIS together with its already established system components (which in both [OAIS2002] and [OAIS2009] are referred to as "functional entities", see Figure 5) provides the skeleton of the central archiving phase of the ILM, which is the core of the secure long-term archive to be developed. However, as explained previously in subsection 2.4.2, the original OAIS standard lacks detailed information about security requirements, thus it needs to be enhanced in this regard. To apply context modelling towards such enhancement, the application scenario is analysed with its security concerns so its context can be extracted and processed. In the scope of SHAMAN, the application scenario is described by the use-cases provided by corresponding project partners. These use-cases are regarded as the other part of system level context that describes the secure long-term archive in the discussed application scenario (as illustrated in Figure 13), and the first step to process it is to review all these use-cases and select those security-related ones, yielding the *global security context*. The results of this step can be referred to in see Appendix A.

As Appendix A shows, the global security context reveals the security related events expected for the archiving system, so the basic security aspects involved in this application scenario can be easily identified, trivially leading to the ASPs of the system, following the formalisation described in formula (3.5b):

ASP$_1$: *The archiving system must be in conformity with both data origin and entity authenticity.*

ASP$_2$: *The archiving system must be in conformity with integrity.*

ASP$_3$: *The archiving system must be in conformity with confidentiality.*

ASP$_4$: *The archiving system must be in conformity with availability.*

ASP$_5$: *The archiving system must be in conformity with non-repudiation.*

These ASPs serve as the basis and reference for further security policies on finer granularities.

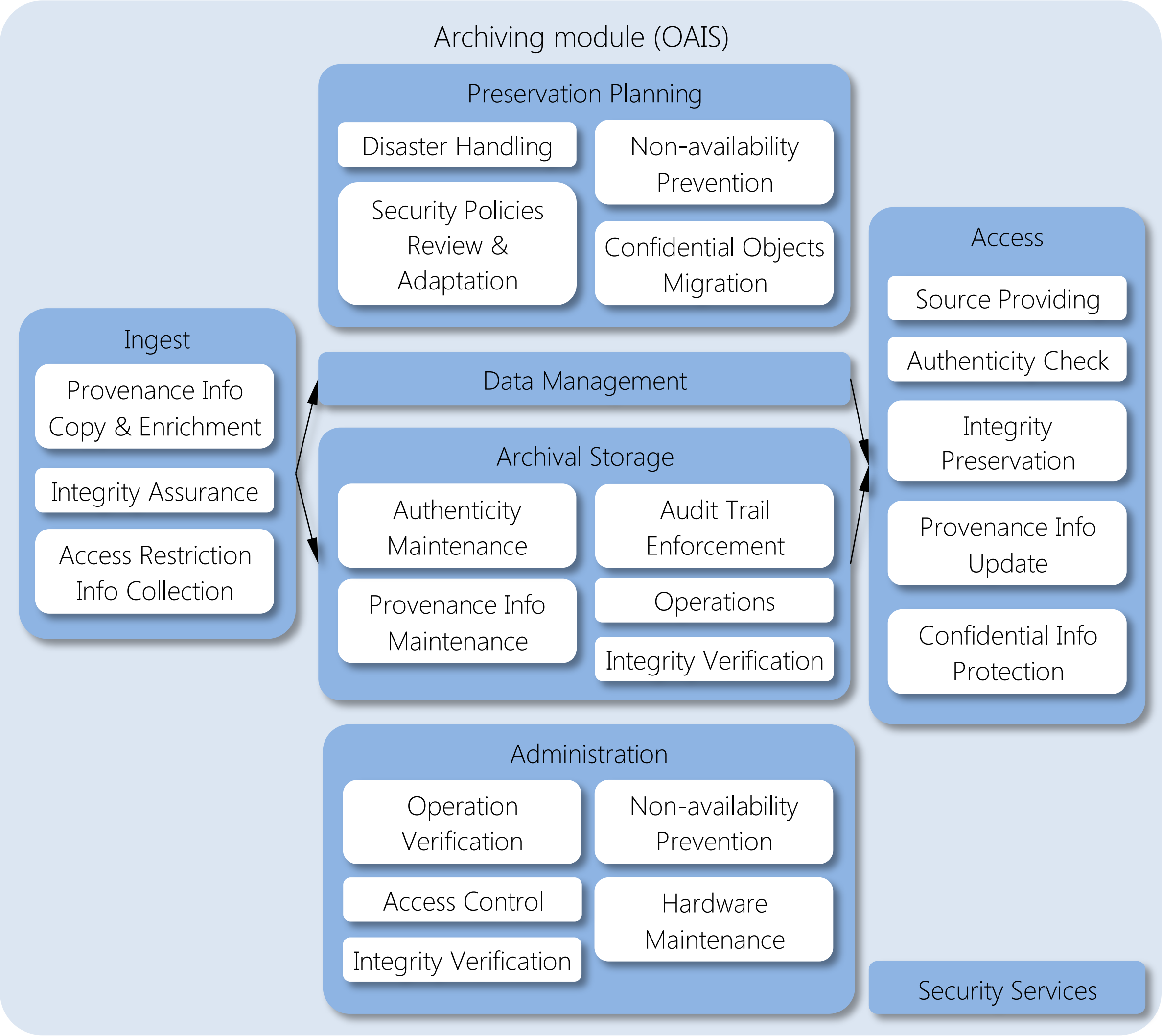


Figure 14. Required extension of the OAIS functional entities from a security point of view (adapted from [QSK+2011b])

More importantly, as here use-cases can be further formulated to clarify the involved entities (marked as "roles" and "objects" in Appendix A), the global security context can be conveniently interpreted into *local security context*, i.e. the use-cases can be categorised into their corresponding functional entities existing in OAIS model, and as such the additional specified security-related functions and services (e.g. hash calculation, certificate management, logging service, trusted time-stamping, etc.) can be revealed for these established functional entities, including the *security service* entity, to provide better context representation in terms of security. Figure 14 shows these newly derived functions extended in [QSK+2011b] to meet the security requirements regulated by the ASPs for the original OAIS functional entities, whereas their original functions from [OAIS2002] and [OAIS2009] are omitted for the sake of clarity.

4.1.2 Security policy hierarchy

As explained in subsection 3.2.2, the ASPs (in this case together with the global context) are used to derive SSPs on finer granularities, eventually yielding executable SRs. For the sake of more vivid representation of the complicated relationships among the entities in complex data processing systems as well as the implementation of the means of governance or orchestration, a hierarchical organisation of the policies is necessary. Therefore, a security policy hierarchy is adapted from [BS2002] and introduced here, serving for the generation, implementation and enforcement of huge amount of SSPs, which covers on various granularities of IPs in the ILM. As already shown in Figure 12, such security policy hierarchy comprises the following levels of SSPs: meta-SSP ($\boldsymbol{SSP_{meta}}$)/high-level SSP ($\boldsymbol{SSP_{high}}$), mid-level SSP ($\boldsymbol{SSP_{mid}}$), and low-level SSP ($\boldsymbol{SSP_{low}}$):

$$\boldsymbol{SSP} = \{\boldsymbol{SSP_{meta}},\ \boldsymbol{SSP_{high}},\ \boldsymbol{SSP_{mid}},\ \boldsymbol{SSP_{low}}\} \tag{4.1a}$$

Within the scope of OAIS standard, the policy generation is the responsibility of the function of “develop preservation strategies and standards” within the *preservation planning* functional entity. Integrating with the theory of system level security-oriented context modelling introduced in subsection 3.2.2, such generation starts on a global system level with the most abstract types of SSPs – meta-SSPs and high-level SSPs. Meta-SSPs make statements about other SSPs, which can be described based on the previous definition of general SSP given in (3.5c):

$$\begin{aligned}
&\boldsymbol{SSP_{meta}} = (E_{meta}, \{o\}, R_{meta})\\
&E_{meta} = \{E_{metaT}, E_{metaC}\}, E_C = \{E_{metaGC}, E_{metaSC}\}\\
&E_{metaT} = \boldsymbol{SSP_{high}} \cup \boldsymbol{SSP_{mid}} \cup \boldsymbol{SSP_{low}}\\
&E_{metaGC} = \{e_{gc} | e_{gc} \text{ refers to general requirements as virtual concepts. }\}\\
&E_{metaSC} = \{e_{sc} | e_{sc} \text{ refers to security-oriented requirements as virtual concepts.}\}\\
&o \text{ refers to conformity}\\
&R_{meta} \subseteq E_{meta} \times \dots \times E_{meta}
\end{aligned} \tag{4.1b}$$

where either general or security-related requirements (denoted respectively by E_{metaGC} and E_{metaSC}) are regulated for the conformity with the TEs of $\boldsymbol{SSP_{meta}}$, namely all other SSPs.

High-level SSPs make statements about general security goals and acceptable procedures on a system-wide perspective throughout the lifecycle of IPs. Similar to (4.1b), it can also described as

$$\begin{aligned}
&\boldsymbol{SSP_{high}} = (E_{high}, \{o\}, R_{high})\\
&E_{high} = \{E_{highT}, E_{highC}\}, E_C = \{E_{highGC}, E_{highSC}\}\\
&E_{highT} = \{e_t | e_t \text{ refers to the virtual concepts on the system level.}\}\\
&E_{highGC} = \{e_{gc} | e_{gc} \text{ refers to general requirements.}\}\\
&E_{highSC} = \{e_{sc} | e_{sc} \text{ refers to general security goals as virtual concepts.}\}\\
&o \text{ refers to conformity}\\
&R_{high} \subseteq E_{high} \times \dots \times E_{high}
\end{aligned} \tag{4.1c}$$

while its TE is mostly the whole system, or unspecific operation/data object on the system level.

As shown in (4.1b) and (4.1c), both meta-SSP and high-level SSP regulates requirements on a quite coarse granularity level and can be derived from either the global security context, or directly the ASPs that reveal general understanding of the system and/or the application scenario.

Comparing to the three-layer policy model introduced in [BS2002], a new layer of mid-level SSP is added on a finer granularity than that of meta-SSPs and high-level SSPs, for better handling of larger complex system modules. This reflects the fact that many use-cases in the global security context do not make assertions about the system that covers the ILM as a whole, but about certain functionalities. Such use-cases are restricted to larger system modules (here equivalent to OAIS, the central archiving phase in ILM) and their domain of various functions. Therefore, the mid-level SSP can be described similar to (4.1b) and (4.1c) as:

$$\begin{aligned}
&\boldsymbol{SSP_{mid}} = (E_{mid}, O_S, R_{mid}) \\
&E_{mid} = \{E_{midT}, E_{midC}\}, E_{midC} = \{E_{midGC}, E_{midSC}\} \\
&E_{midT} = \left\{e_t \middle| \begin{array}{c} e_t \text{ refers to OAIS archiving module as system components,} \\ \text{data objects or virtual concepts on the module level.} \end{array}\right\} \\
&E_{midGC} = \left\{e_{gc} \middle| \begin{array}{c} e_{gc} \text{ refers to general virtual concepts,} \\ \text{time, or space.} \end{array}\right\} \\
&E_{midSC} = \{e_{sc} | e_{sc} \text{ refers to security-oriented virtual concepts. }\} \\
&O_S = \left\{o_s \middle| \begin{array}{c} o_s \text{ refers to security-oriented processes, events,} \\ \text{or expressive forms.} \end{array}\right\} \\
&R_{mid} \subseteq E_{mid} \times \ldots \times E_{mid}
\end{aligned} \tag{4.1d}$$

In the policy generation hierarchy these mid-level SSPs on the one hand serve as a process-based filter for the large quantity of system global security context, and on the other hand they also serve to verify if the high-level SSPs themselves make sense by not contradicting the existing context (i.e. verify the consistency between global and local context modelling).

The mid-level SSPs are used to act as the basis for the generation of low-level SSPs, which provide sufficient information within particular OAIS functional entities about what needs to be implemented as a SR in the enforcement. Similar to (4.1b–d), low-level SSP can be described as:

$$\begin{aligned}
&\boldsymbol{SSP_{low}} = (E_{low}, O_S, R_{low}) \\
&E_{low} = \{E_{lowT}, E_{lowC}\}, E_{lowC} = \{E_{lowGC}, E_{lowSC}\} \\
&E_{lowT} = \left\{e_t \middle| \begin{array}{c} e_t \text{ refers to data objects, human agents,} \\ \text{or virtual concepts in one fuctional entity.} \end{array}\right\} \\
&E_{GC} = \left\{e_{gc} \middle| \begin{array}{l} e_{gc} \text{ refers to fdata objects, human agents, virtual} \\ \text{concepts, time, or space in one functional entity.} \end{array}\right\} \\
&E_{SC} = \left\{e_{sc} \middle| \begin{array}{l} e_{sc} \text{ refers to security-oriented data objects, human} \\ \text{agents, or virtual concepts in one functional entity.} \end{array}\right\} \\
&O_S = \left\{o_s \middle| \begin{array}{c} o_s \text{ refers to security-oriented processes, events,} \\ \text{or expressive forms.} \end{array}\right\} \\
&R_{low} \subseteq E_{low} \times \ldots \times E_{low}
\end{aligned} \tag{4.1e}$$

In the ideal case, such low-level SSPs are precise enough for the direct derivation of SRs in a formalised language (see formula (3.5d)) from them.

As proposed in [QSK+2011b], for the sake of clarity as well as the traceability of the policy origins, a SSP derived from a higher level SSP comes with an identifier indicating its parent SSPs. If high-level SSPs have identifiers of the format Px (with x being an unique identifier), their children mid-level SSPs are appointed with identifiers that include their parent's identifiers (e.g. Px-y), and similarly the grandchildren low-level policies are derived with identifiers identifying their parent and grandparent policies (e.g. Px-y-z). As policies need to be updated or even removed at certain times, this form of traceability eases the browsing of the hierarchical tree structure of the policies that would be required in these cases. This is further shown in section 4.2.

For highly complex systems some issues raise regarding the implementation and enforcement of policies: First, when introducing a new policy into such systems, there could be multiple possible methods to implement it. Thus it requires specific analysis (e.g. complexity-based) to identify the optimal method. Second, complex systems are with a high probability also heterogeneous, therefore considerations have to be included on the interoperability, distribution and orchestration of policies and policy descriptions (for instance how to interpret between possible different policy syntaxes used in different parts of a heterogeneous system). Third, due to the quantity and complexity of the policies, it is necessary to develop an assurance and auditing mechanism to make sure that all the policies are enforced properly.

Within the range of proposed framework in this chapter, the SSPs are basically descriptions in natural language of what the archiving system (or system components) does, which creates barriers for actual enforcement. Therefore, in this framework, after the SSPs are generated by the function "develop preservation strategies and standards" in the scope of *preservation planning* functional entity, they are sent to the function "establish standards and policies" within the scope of *administration* functional entity, where the SSPs are implemented by applying a manual and iterative procedure which turns low-level SSPs into enforceable SRs. The procedure is described in [QSK+2011a], [SQK+2011], and [QSK+2011b] as follows:

- *SR creation*: This turns low-level SSPs, which define what needs to be done, into SRs, which define how the SSP is enforced. It analyses the statement in a SSP by utilizing validation criteria that consist for the significant properties, format validation, organisational and domain information. Then a sequence of steps is derived, describing specific actions. Each step shall be as atomic as possible, ideally performing one action and also verifiable, so it can be considered as one abstract SR. Optionally a SR can comprise sub-SRs if one of the steps is too complex to be described as a single action. Therefore the output here is a sequence of abstract SRs.
- *SR instantiation*: Abstract SRs are not executable as they only describe actions in natural language. Therefore it is necessary to derive executable SRs from abstract ones. Templates containing the grammar and syntax for rule-engines can be used by a SR instantiation tool to create realisations of the abstract SRs. Such tool must also keep track of the realisation process so that it is possible to track from an executable SR back to the abstract SR and then back to the low-level SSP. Additionally, similar to Event-Condition-Action (ECA) rules which always have the form of if…then…else, the executable SRs are formalised with regard to the structure described in formula (3.5d), thus each SR becomes an executable atomic data processing operation.
- *SR validation*: Here it is ensured by validation that the instantiated executable SRs are correct implementations of the policies. The functionality of the used validation tools would be defined by the validation criteria, which are the adherence to the global and local security context (developed in subsection 4.1.1). After a SR passes the validation, it is deployed with records of its deployment time and intended deployment enforcement point in the production system and ready to be enforced.

As the policy enforcement in this framework is designed to be realised by enforcing the formalised and validated SRs, a specific solution for such realisation is also proposed in [QSK+2011b] by adapting the concepts introduced in [YPG2000]: A Rule Decision Point (RDP) is set as a component of the function "establish standards and policies", where all the enforceable SRs are developed. Rule Enforcement Points (REPs) are set as components of all the security-related functions within OAIS (see Figure 14). A RDP takes responsibility of making rule decisions and those decisions are sent to corresponding REPs to be enforced. The basic interaction between these components begins with the REP, when the function it belongs to requires enforcing a SR. Then the REP formulates a request for a rule decision and sends it to the RDP. The RDP generates the proper rule decision based on the received request and returns it to the REP, and then the REP executes the rule decision.

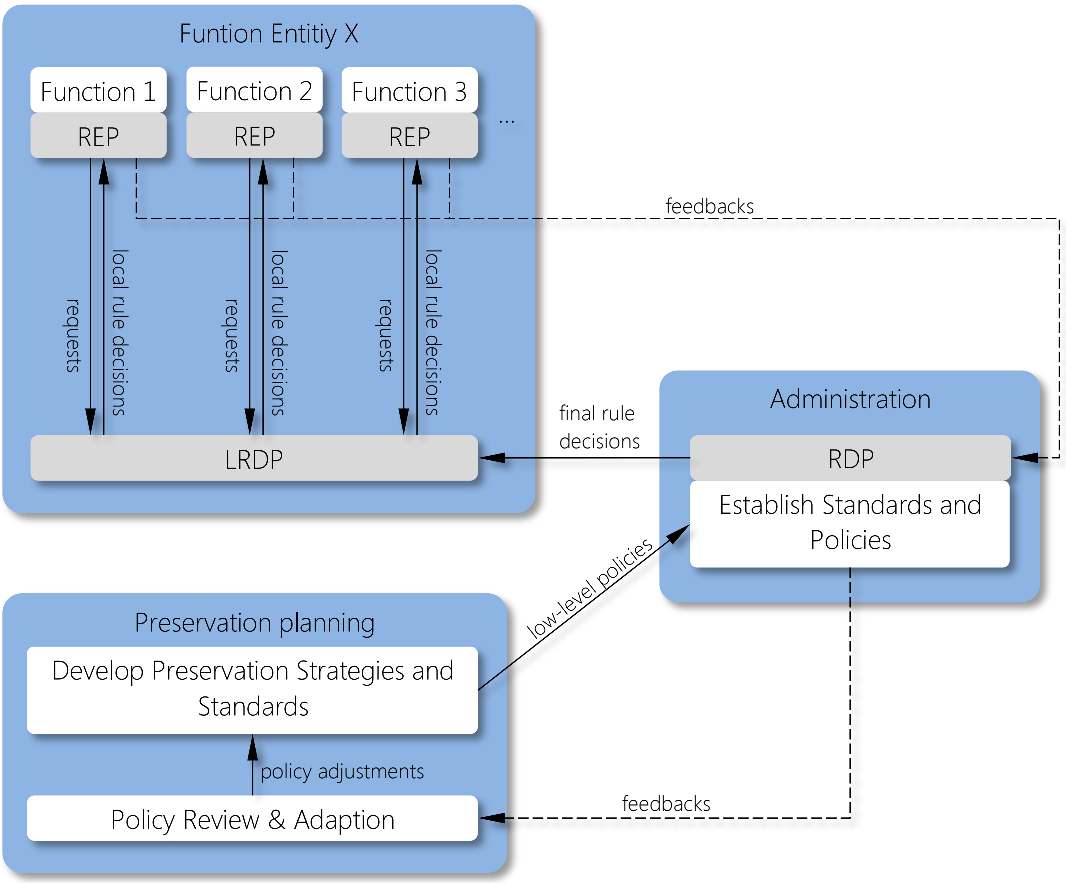


Figure 15. Security policy implementation and enforcement (re-sketched from [QSK+2011b])

In complex systems like the archiving system discussed in this chapter, it is expected that the quantity of rules may overwhelm the RDP, thus it is necessary to introduce the concept of Local Rule Decision Points (LRDPs), which locate in different functional entities and are responsible for generation local rule decisions within the scope of the entity. In this case, the request for rule decision is sent to the corresponding LRDP first and a local rule decision is returned, whereas the RDP remains as final authority and it gives final rule decisions if necessary, which override the LPDPs. Furthermore, when an unusual situation occurs (e.g. a conflict between two rule decisions is detected), the related information is collected and sent back to the function "develop preservation strategies and standards" through the RDP. Then the function sends a feedback about the situation to the function "security policy review and adaptation", which has the authority to make necessary adjustment to involved policies and communicates with the function "develop preservation strategies and standards" with policy adjustments. The technical details on the communication among RDP, REPs and LPDPs are out of the scope of this dissertation, therefore not discussed here. However interested readers are recommended to refer to COPS protocol described in [BCH+2000], as this protocol can be served as a valid fundamental in this case. Figure 15 visualises the interactions among involved functions, RDP, LRDP and REPs during the security policy implementation and enforcement.

It should be noted that as the archiving system discussed here aims for long-term preservation, it is expected that the expression and format of the policies and rules evolve during the preservation with the alternation of global and local context. Therefore, the related components should also evolve accordingly. Furthermore, as in the proposed framework the distribution, communication and encoding of policies in the form of rules is solely done via the rule points (LRDP

↔ REP, RDP → LRDP), any changes necessary due to the long time periods involved are isolated from the other functional parts of the system.

4.1.3 IP processing and central control

In the *IP processing* block a function within a system functional entity enforces rules on IP from the archival system and/or system data (like search indexes, the user database, etc.). The result of the enforcement has to be communicated by the responsible REP to the central audit service. This central audit is part of the functionality of the *control* block. Besides this audit functionality there are also mechanisms for the storage of the policy tree (all policies are communicated to this storage during the construction of the *policy hierarchy*) as well as the policy conflict analysis and conflict resolve. The corresponding OAIS authority responsible for these operations would be the function "security policies review & adaptation" in the *preservation planning* entity (see Figure 14). It is designed to keep track of all the policies to ensure they operate properly, especially no policy from one phase conflicts with those from other ones.

4.1.4 Security framework construction by combination of functional blocks

Figure 16 on page 61 extends the low detail description of the contextualisation framework explained in the preceding three subsections by the data, information and control flows described for the four functional blocks introduced in the beginning of section 4.1.

In the figure the importance of the *control* block especially sticks out as a dominant factor. Each context modelling block, the different stages of the policy generation hierarchy and the IP processing communicate their actions to the *control* block. This is on one hand done to audit all operations for purposes of transaction control and non-repudiation of transactions as. On the other hand this functional block also acts as central policy storage repository and performs policy conflict analysis and resolve.

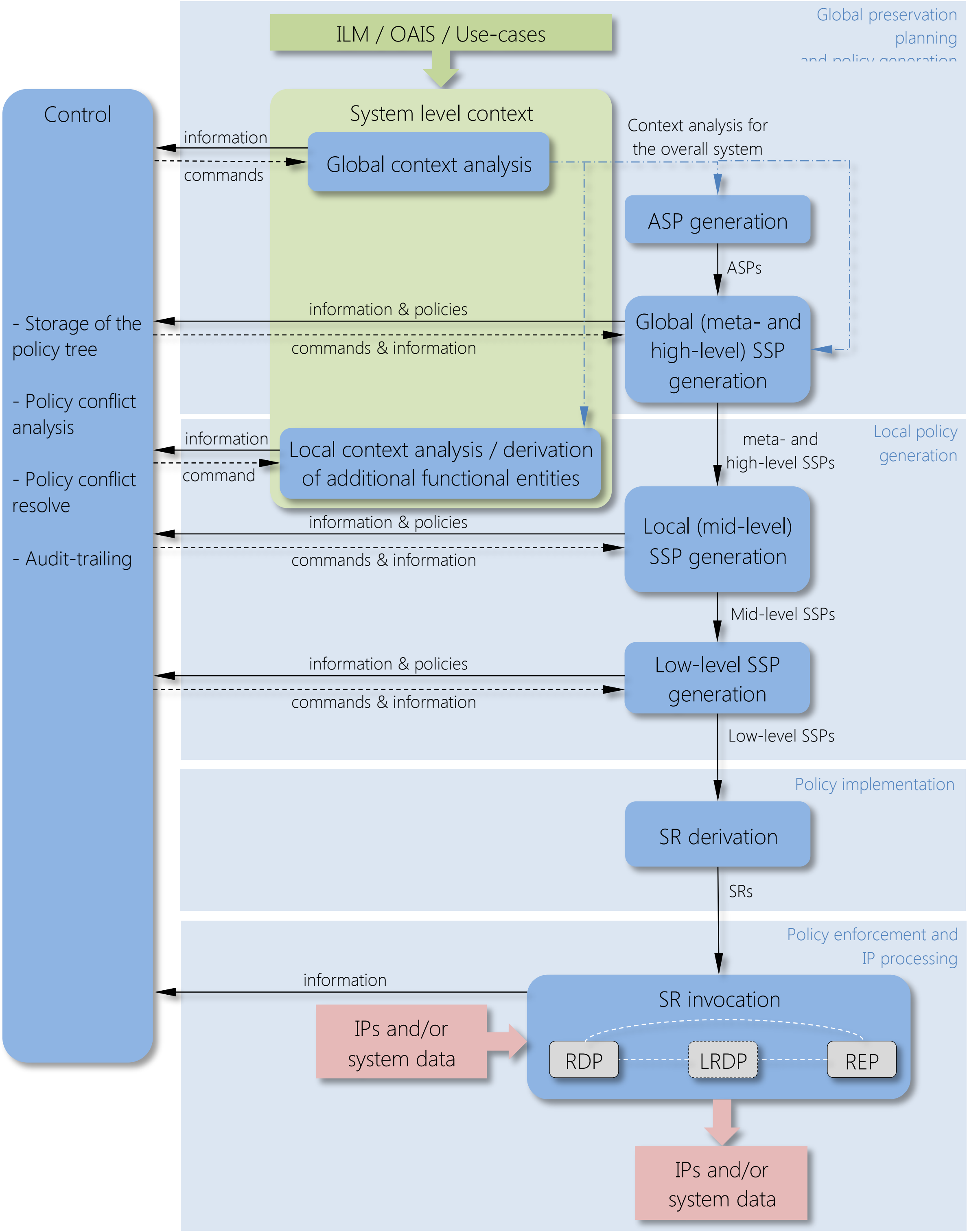


Figure 16. General overview over the security contextualisation framework (adapted from [QSK+2011b])

4.2 APPLICATION OF THE CONTEXT MODEL

This section shows the application of the proposed framework on one of the functional entity within the extended OAIS described in subsection 4.1.1. As digital long-term preservation systems like many other are rather complex with a large amount of functional entities, for demonstration purposes here the considerations are restricted only on the *ingest* functional entity.

The application is performed in four stages, following the framework overview shown in Figure 16.

4.2.1 Global preservation planning and policy generation

For the preservation planning the precise functionality of the *ingest* functional entity needs to be specified at first. The functions from the original OAIS standard include [OAIS2009]:

- "Receive submission" where the producer uploads its content as a SIP into the system.
- "Quality assurance" validates this SIP, checking whether it is conform to the specification of the system, whether it is a valid SIP, and if any security related issues like non-integer transmissions are solved.
- "Generate AIP" transforms one SIP into one or more AIPs and generates audit information.
- "Generate descriptive info" can receive the generated AIP, create or extract metadata of the AIP, and indexes to aid the later search mechanisms.
- "Coordinate updates" stores the AIPs as well as the descriptive information in the archival storage.

Furthermore, as introduced in subsection 4.1.1 there are extended security-related functions:

- "Provenance info copy and enrichment" ensures the completeness of the provenance information contained in the metadata as part of authenticity assurance.
- "Integrity assurance" prevents data from being manipulated either by accidence or ill will.
- "Access restriction info collection" collects necessary information for access control.

Additionally, this functional entity also applies some of the functions which are contained in the *security services* entity defined by OAIS standard [OAIS2009]:

- "Identification/authentication service" identifies the identities of the producers.
- "Access control service" restricts access to sensitive resources, using the information collected by "access restriction info collection".
- "Data confidentiality service" prevents the disclosure of information.
- "Non-repudiation service" provides irrefutable proof that two subjects have communicated with each other and exchanged certain objects (data).

All these functions are either provided with global/local security context or are meant to be representing certain use-cases in the context. As such, all these functions serve a double purpose of being the system outline (or enforcement points for SSPs/SRs) and their specifications and description provide a basic context which is used here to derive from, align to and finally enforce the policies.

Following the method explained in subsection 4.1.2 the following high-level SSPs can be derived from the global and local security context on the system level that covers the ILM, as shown in Table 8. Following the definition given in formula (4.1c), different parts of the semantics are highlighted, representing TEs, GCEs and SCEs as well as the corresponding operations, thus also implicitly revealing the relations. In this case no meta-policy is derived, only because none of the use-cases from the system level context leads to such policies.

Policy	*Description*
P01	The system provides the means to authenticate all objects by providing/identifying the sources it was created from.
P02	The system ensures authenticity of digital objects for all steps of processing.
P03	Each operation must be logged including what the operation has processed, on whose behalf, when and with which result.
P04	To ensure their correct working, operations must be validated/verified that they correspond to the policies.
P05	The system provides mechanisms to authenticate subjects.
P06	Originals of Content Information in the system must not be altered but only copies thereof.
P07	The integrity of object must be guaranteed for all processing steps.
P08	The system's integrity must be verified periodically.
P09	The system must be able to recover from integrity violations.
P10	The actual performance of the system's integrity preservation must be audited by an independent mechanism.
P11	The system must employ access restriction mechanisms.
P12	The system's access control must allow backups, data replication and other prevention and recovery measures for disaster handling.
P13	The system must review its security policies at certain time intervals and adapt it to newly identified risks.
P14	The system shall meet or exceed specified availability requirements.
P15	The system must employ measures to prevent non-availability.
P16	The system must employ measures to detect actions aiming at non-availability.
P17	The system must employ measures to recover from non-availability.
P18	The audit trail must be available at any time, conversely the system must not operate without an audit trail although this will result in non-availability.
P19	Confidential information in the system must not be disclosed.

Table 8. High-level SSPs for the archival system that covers the ILM, with TEs, GCEs, SCEs and operations highlighted respectively in blue, green, red, and orange (based on [SQK+2011])

4.2.2 Local policy generation

In the next stage, based on the derived high-level SSPs, the layer of mid-level SSPs are subsequently generated as explained in subsection 4.1.2. These SSPs refer to the larger system module, which in this case is equivalent to the OAIS archiving module. Thus the mid-level SSPs refine the requirements of the high-level SSPs for the archiving module. They act as a mediator between the general high-level SSPs which consider the system scope and low-level SSPs which are very concrete policies that provide sufficient information what needs to be implemented as a SR in an enforcing mechanism to adhere to the high-level SSPs. In the following only selected and refined policies are shown, i.e. the considerations are limited to the *ingest* functional entity of the archiving module and high-level SSPs which also have sufficient power of expression or are otherwise directly usable for this level are omitted as the derived mid-level SSPs would have the same description. For the sake of clarity a mid-level SSP derived from a high-level SSP Px is noted as Px-y. The actual mid-level SSPs in the OAIS-compliant archiving module are:

Policy	*Description*
P02-01	The archiving module ensures authenticity of digital objects for all steps of processing.
P03-01	Each operation in the module must be logged including what the operation has processed, on whose behalf, when and with which result.
P07-01	Objects where an operation is applied on, must be checked if they and their references are integer anymore, at least for operations with write access.
P07-02	If SIP is ingested and thus becoming one or more AIP or AIP are accessed and delivered as one or more DIP, the archiving model must preserve the integrity, especially the semantic and referential integrity between the SIP to AIP and the AIP to DIP conversion processes.
P08-01	The module's integrity must be verified periodically.
P09-01	The module must be able to recover from integrity violations.
P11-01	The module must employ access restriction mechanisms.
P19-01	Confidential information in the module must not be disclosed.

Table 9. Mid-level SSPs in the archiving module/OAIS, with their TEs, SCEs, GCEs and operations highlighted respectively in blue, green, red, and orange (based on [SQK+2011])

Similar to Table 8, highlighting is applied to denote the TEs, GCEs, SCEs and corresponding operations in the mid-level SSPs shown in the table. These mid-level SSPs act as the basis for the generation of low-level SSPs in specific functional entities. The following Table 10 shows the derived low-level SSPs for the *ingest* functional entity. In the table a low-level SSP derived from a mid-level SSP Px-y is noted as Px-y-z. Similarly, highlighting is also used here to mark the TEs, GCEs, SCEs, together with operations in the SSPs.

Policy	*Description*
P02-01-01	If SIP is ingested and thus becoming one or more AIP, the ingestion must preserve the authenticity/provenance, by including the SIP Provenance information and enriching it with

	data about the ingest (authenticated Provider, time, etc.).
P02-01-02	If external systems are responsible for ingestion, these preserve the actual provenance but must also be referred to in the provenance information.
P03-01-01	The ingestion must ensure it is provable which ingest preparation policies were applied.
P07-01-01	All metadata linking to other SIP or their metadata must link to the corresponding AIP or their metadata to preserve the authenticity/provenance information.
P07-01-02	If objects are migrated, for example conversion into a newer format, the integrity of the old and the new version must both be enforced, and the newer version should be at least semantic integer with the older version directly after the conversion.
P07-02-01	If SIP is ingested and thus becoming one or more AIP, the ingestion must preserve the integrity which especially includes semantic and referential integrity that the SIP are ingested as AIP completely.
P08-01-01	To preserve the integrity of objects, prevention and recovery measures must at least include data replication and backups, while after a recovery the whole system must be checked for integrity.
P09-01-01	Although the ingestion should be as fault tolerant as possible, SIP must be ingested either wholly/completely or not at all, but not partially.
P11-01-01	The ingestion must consider access restrictions provided stated within the SIP and include these for the actual access restrictions of the AIP.
P19-01-01	If SIP is ingested and thus become one or more AIP, the appropriate AIP must contain the confidentiality conditions.

Table 10. Low-level SSPs in *ingest* function entity, with their TEs, SCEs, GCEs and operations highlighted respectively in blue, green, red, and orange (based on [SQK+2011])

4.2.3 Policy implementation

As explained in subsection 4.1.2, the implementation of the low-level SSPs is realised by further deriving executable SRs. Based on the low-level SSPs listed in Table 10, the corresponding executable SRs can then be derived. The following table shows these derived SRs, where a SR derived from a low-level SSP Px-y-z is noted as Rx-y-z-w. These SRs are formulated with regard to the definition introduced in formula (3.5d) in subsection 3.2.2, so the comprised TEs, SCEs and GCEs as well as the operations on them are explicitly clarified.

Rule	***Context entities***			***Target entity***
	e_{gc}	e_{sc}	o_s	e_t
R02-01-01-01	SIP ingestion (time)	Identification/authentication service (function)	identify	Authenticity info of the SIP
R02-01-01-02	SIP ingestion (time)	Provenance info copy and enrichment (function)	retrieve	Authenticity/provenance info of the SIP

R02-01-01-03	SIP ingestion (time), Generate descriptive info (function)	Authenticity/provenance info of the SIP	update	Metadata of AIP/AIPs
R02-01-02-01	Involvement of external system (time)	Provenance info copy and enrichment (function)	retrieve	Actual provenance info from the external system
R02-01-02-02	Involvement of external system (time) Generate descriptive info (function)	Actual provenance info from the external system	update	Metadata of AIP/AIPs
R03-01-01-01	Generate AIP (function)	SIP to be ingested	generate	Info regarding the ingest preparation policies for the SIP
R03-01-01-02	-	Non-repudiation service (function)	document	Info regarding the ingest preparation policies for the SIP
R07-01-01-01	SIP ingestion (time), Generate descriptive info (function)	Metadata of SIP	link	Corresponding metadata of AIPs
R07-01-02-01	AIP migration (time), Quality assurance (function)	Old AIP, new AIP	verify	Integrity report
R07-01-02-02	AIP migration (time)	Integrity assurance (function), Integrity report	accept /reject	New AIP
R07-02-01-01	SIP ingestion (time), Quality assurance (function)	SIP, AIP/AIPs	verify	Integrity report
R07-02-01-02	SIP ingestion (time)	Integrity assurance (function), integrity report	accept /reject	AIP/AIPs
R08-01-01-01	-	Coordinate updates (function), AIP	store	Redundant copy/copies of AIP
R08-01-01-02	System recovery (time)	Integrity assurance (function)	verify	Various copies of AIP
R09-01-01-01	SIP ingestion (time)	Integrity assurance (function), Integrity report	accept /reject	SIP to be ingested
R11-01-01-01	SIP ingestion (time)	Access restriction info collection (function)	collect	Access restriction info of the SIP to be ingested

R11-01-01-02	SIP ingestion (time)	Access control service (function)	restrict	Access to corresponding AIP/AIPs
R19-01-01-01	SIP ingestion (time)	Data confidentiality service (function)	enforce	Confidentiality conditions of AIPs

Table 11. Formulated SRs in *ingest* function entity, the expressions of relation r_s are omitted as they all conform with that in formula (3.5d)

As shown in Table 11, each of these SRs describes an atomic action sourced from one function on one target entity in *ingest* functional entity, so they are ready to be directly enforced.

4.2.4 Policy enforcement/IP processing

As depicted in the theory of system level security-oriented context modelling in subsection 3.2.2 and also reflected in 4.2.3, the SRs represent atomic processing procedures on the data, which in this case is IPs in various forms. Therefore the framework enforces the policies by actually executing the SRs derived from them. Figure 17 on page 68 illustrates the enforcement of SRs derived in Table 11 within the scope of *ingest* functional entity.

As shown in the figure, the SRs are processed by the LRDP of the functional entity and enforced by the appropriate REPs of the respective functions. Some functions may receive no SRs (e.g. "Receive submission", which nevertheless would certainly receive general rules to process SIPs) whereas others may receive multiple ones (e.g. "Generate descriptive info"). Furthermore, the enforcement of certain SRs also involves the general *security services* defined in OAIS standard, therefore they are also shown in the figure.

Although not shown, all the operations and interactions within the scope of this functional entity are monitored by the central audit service, hence managed by the central management service. Both of these central services are provided by the *control* block of the framework, as already shown in Figure 16 on page 61.

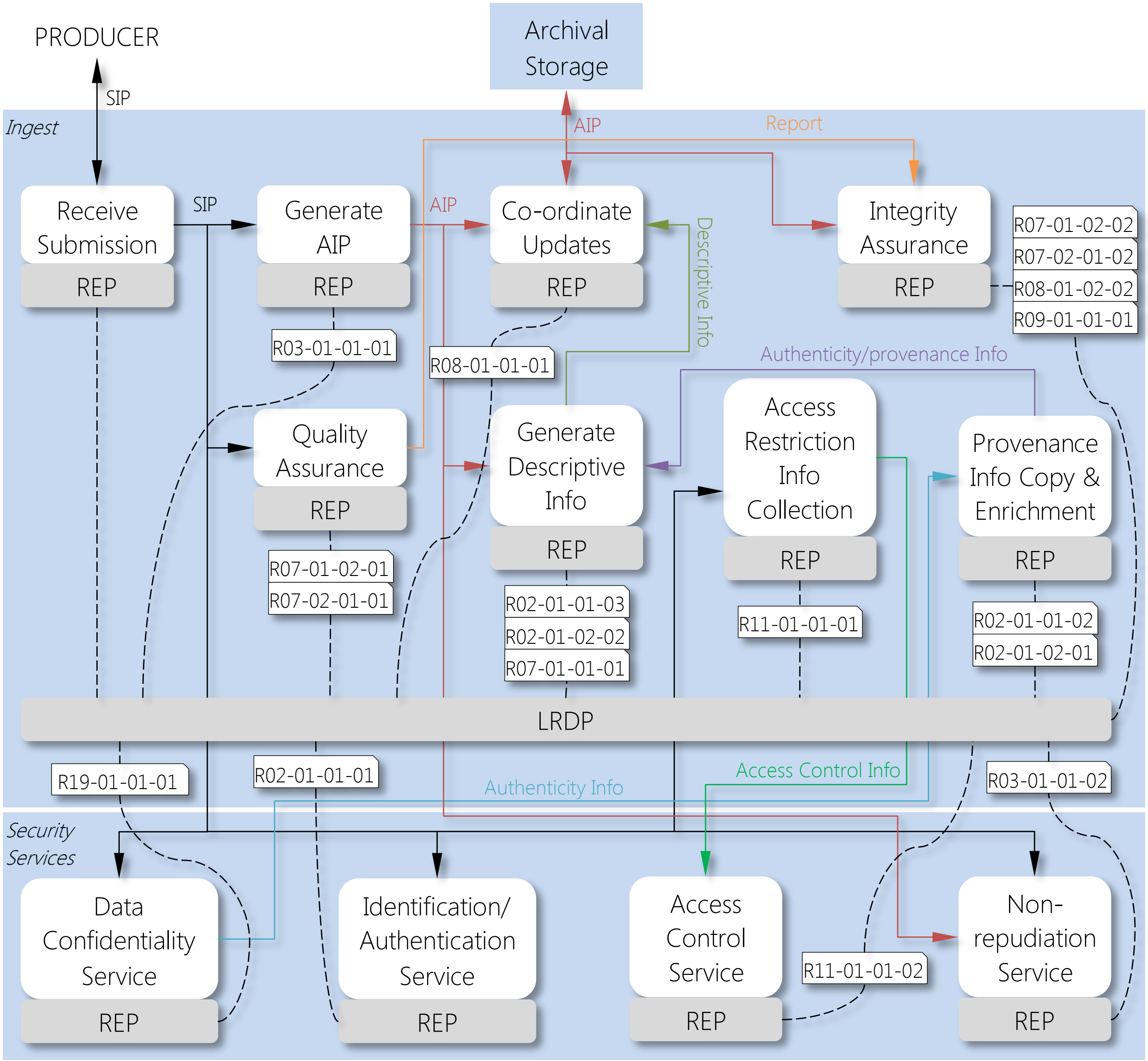


Figure 17. Functions and SRs related to the *ingest* functional entity

4.3 Exemplary security assessment

As introduced in subsection 2.4.2, an assessment framework has been developed in SHAMAN to evaluate the performance of the long-term archiving system that realises the preservation framework, consisting of three parts: 1) the general assessment of the SHAMAN project, 2) the assessment of SHAMAN preservation system based on both TRAC (Trustworthy Repositories Audit & Certification: Criteria and Checklist) and DRAMBORA (Digital Repository Audit Method Based on Risk Assessment) toolkit, and 3) the assessment of individual SHAMAN work-packages based on iRODS (i Rule Oriented Data System) rules. As the security framework developed in this chapter aims at its ultimate realisation towards a preservation system, only the second part of the assessment framework is considered as relevant here.

However, as DRAMBORA toolkit applies a bottom-up approach which meant for the assessment of the outcome of the implementation of the SRs, it is not directly applicable on the SRs. Therefore, this section selects a group of audit and certification criteria in TRAC from Appendix B and uses them to demonstrate a security assessment on the SRs developed in section 4.2 as an example.

As shown in Appendix B, the TRAC audit and certification criteria cover three main categories [TRAC2007]:

- A. *Organisation infrastructure*, in which the criteria regulate all the organisational attributes, used as indicators of the digital archive's comprehensive planning, readiness, ability to address its responsibilities, and trustworthiness.
- B. *Digital object management*, in which the criteria regulate both organisational and technical aspects of the requirements for the functions, processes, and procedures of handling digital objects, grouped under OAIS functional entities.
- C. *Technologies, technical infrastructures, & security*, in which the criteria regulate the requirements for general best practices for data management and security.

To evaluate the SRs derived for the *ingest* functional entity, the criteria developed in category B2 *Ingest: creation of the archivable package* are subject to further selection, resulting the security-related ones. Table 12 summarises the assessment using these security-related criteria. It identifies the related security aspects for selected criteria, and then uses them to assess the SRs derived in section 4.2, to see whether the SRs adequately address them or not.

Security-related TRAC criteria	***Related security aspect***	***Assessment***	***Corresponding SRs***
B2.3 Repository has a description of how AIPs are constructed from SIPs.	Authenticity	Addressed	R02-01-01-01, R02-01-01-02, R02-01-01-03, R07-01-01-01
B2.4 Repository can demonstrate that all submitted objects (i.e., SIPs) are either accepted as whole or part of an eventual archival object (i.e., AIP), or otherwise disposed of in a recorded fashion.	Integrity	Addressed	R09-01-01-01
B2.5 Repository has and uses a naming convention that generates visible, persistent, unique identifiers for all archived objects (i.e., AIPs).	Authenticity	Addressed	R02-01-01-01, R02-01-01-02, R02-01-01-03
B2.6 If unique identifiers are associated with SIPs before ingest, the repository preserves the identifiers in a way that maintains a persistent association with the resultant archived object (e.g., AIP).	Authenticity	Addressed	R02-01-01-01, R02-01-01-02, R02-01-01-03
B2.7 Repository demonstrates that it has access to necessary tools and resources to establish authoritative semantic or technical context of the digital objects it contains (i.e., access to appropriate international Representation Information and format registries).	Authenticity	Addressed	R02-01-02-01, R02-01-02-02

B2.8 Repository records/registers Representation Information (including formats) ingested.	Authenticity	Addressed	R02-01-01-01, R02-01-01-02, R02-01-01-03, R07-01-01-01, R07-01-02-01, R07-01-02-02
B2.9 Repository acquires preservation metadata (i.e., PDI) for its associated Content Information.	Authenticity	Addressed	R07-01-01-01
B2.10 Repository has a documented process for testing understandability of the information content and bringing the information content up to the agreed level of understandability.	Availability	Addressed	R03-01-01-01, R03-01-01-02
B2.11 Repository verifies each AIP for completeness and correctness at the point it is generated.	Integrity	Addressed	R07-02-01-01, R07-02-01-02
B2.12 Repository provides an independent mechanism for audit of the integrity of the repository collection/content.	Integrity, non-repudiation	Not addressed	
B2.13 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (AIP creation).	Non-repudiation	Not addressed	

Table 12. Exemplary security assessment of the SRs derived for *ingest* functional entity using selected TRAC criteria

As shown in Table 12, 11 out of 13 criteria are addressed by one or multiple SRs derived previously for the *ingest* functional entity. Furthermore, it should be noted that the requirement for non-repudiation reflected in the last two criteria, which are not directly addressed by the SRs, should nevertheless be adequately covered by the audit service provided on the module level (see Figure 16).

4.4 SUMMARY

As an embodiment of the system level security-oriented context modelling explained in Chapter 3, this chapter instantiates the proposed theory into the application scenario of digital long-term archiving, resulting a theoretical framework developed in a "top-down" way. As beneficial complement to the theory, the system level security context appears in the form of use-cases, and by analysing and categorising these use-cases, the context is further interpreted into global security context and local security context. Aiming at closer integration with the OAIS standard, the context model takes into consideration the security-related issues introduced in both versions of the standard [OAIS2002] [OAIS2009]. A solution for security policy generation, implementation and enforcement by adapting the concepts introduced in [BS2002] for policy-based admission control to the OAIS standard, therefore the SSPs on the granularities between coarsest level of ASPs and finest level of SRs can be better organised. Furthermore, with a specific example it is also illustrated that how the policies within one functional entity are generated hierarchically before they are implemented to be ready for the enforcement.

To sum up, this chapter demonstrates that the proposed framework is ready for its application on a security-oriented archival system, which 1) can be constructed with more reasonable storage solution for huge amounts of data (e.g. RAID) instead of WORM media, 2) ensures specific security requirements using specific security policies, 3) manages large amounts of security policies reflecting its high complexity using policy hierarchy, 4) introduces context awareness so the system evolves itself with evolving context (e.g. obsolescence issue) by introducing and adopting new policies while abolishing old policies if necessary. Therefore, such system can not only handle the data obsolescence issue (as introduced in Chapter 1 that NASA encountered), but also pose as a revolutionary solution for the salient need of electronic evidence archive that is capable of preserving large amount of data (as also introduced in Chapter 1 that DEA requires). Furthermore, this chapter also shows an exemplary security assessment, using the relevant criteria selected from an existing assessment framework.

However, securing evidence is never that easy: while the work in this chapter closes the gaps regarding the security issues on the system level, there are further concerns that how to ensure the security of evidence throughout its specific processing procedures, and this rests on a further detailed data level. Therefore, in next chapter the author proposes another instantiation of applying data level security-oriented context modelling. Furthermore, as already shown in the demonstrated assessment section, the current existing assessment framework is designed to be applied on an already realised preservation system, other than the system model itself. This gap is therefore addressed later in Chapter 6.

5 CONTEXT MODELLING ON THE DATA LEVEL

In the last chapter an instantiation of the conception of security-oriented context modelling is presented on the system level. As proposed in Chapter 3, the conception can be applied on both system and data levels, resulting either a data processing system or certain data processing procedures. Therefore, this chapter describes a data level instantiation in the application scenario of forensic dactyloscopy, yielding an overlapped latent fingerprint separation approach and its application.

As summarised in Chapter 2, besides the identified weaknesses existing for the state of the art separation approaches for overlapped fingerprints, there is a huge gap for them to contribute to sound forensic techniques, as very limited effort has been made to have them meet certain standards, e.g. *Daubert* criteria. Therefore in this chapter, as a first attempt towards *Daubert* compliance, a context-based approach is derived, adapted from the relaxation labelling based technique originally proposed in [CFJ+2011], for the high-resolution samples of overlapped fingerprints captured by a CWL sensor and aiming at an improved separation performance. For the sake of *Daubert* compliance, it is essential that all influences on potential evidence are well identified and clarified. This provides the non-repudiation, which significantly contributes to the admissibility of such techniques in court. As introduced in Chapter 2, it is already proven in the field of pervasive computing that large-scale systems with context awareness are capable of handling sophisticated, sometimes even evolving influential factors. In this chapter, such context awareness is also introduced to the development of a digitised forensic[1] approach: the influences are reflected by various context entities along the processing procedures, so they can be processed for further improvement of the approaches.

Following the theoretical framework proposed in Chapter 3, the digitised forensic approach is considered as a cluster of various data processing procedures in the processing environment, serving for its corresponding SRs, within the scope of a forensic framework on the system level (see Figure 9 on page 47 and Figure 10 on page 49). Despite that this dissertation does not focus on such framework, as its development would be another system level instantiation like the one described in detail in Chapter 4, some exemplary SRs can be derived here based on the information provided in section 2.5 as well as expert advices from the field work of forensic investigation[2], to regulate the separation environment of overlapped latent fingerprints in a digitised forensic scenario.

Example 8. SR_0: *A fingerprint expert must oversee the whole process of the separation.*

SR_1: *The fingerprint expert must observe the overlapped fingerprint image to be processed and decide if separation is potentially plausible.*

[1] In this dissertation, *digitised forensics* refers to the forensic science that works with the digitised version of conventional physical evidence (e.g. digital scan of fingerprints or fibres), in comparison to *digital forensics* that works with digital evidence found in digital device (e.g. data captured from confiscated hard disk).

[2] The author thanks Sabine Wabnitz from the State Criminal Police Office of Saxony-Anhalt (*Landeskriminalamt Sachsen-Anhalt*) for her professional consultation as forensic expert during the derivation of the exemplary security rules.

SR$_2$: *If separation is considered as plausible, the fingerprint expert shall decide which method to apply for the separation.*

SR$_3$: *The fingerprint expert must select a separation method that is suitable to process the fingerprint images according to their origin (e.g. means of acquisition).*

SR$_4$: *The fingerprint expert must select a separation method which is able to maintain the fingerprint features on different levels (see subsection 2.5.1) to the largest degree according to the requirements of further processing steps after the separation.*

SR$_5$: *The selected separation method shall be subject to proper standard (e.g. Daubert criteria) regarding the admissibility of the separation results as court evidence.*

SR$_6$: *The fingerprint expert shall apply the separation method selected by SR$_4$ and document the separation results.*

SR$_7$: *The separation results must be able to be reproduced, as a second fingerprint expert must verify the documented separation results.*

SR$_8$: *Only the separation results verified by SR$_7$ can be stored and used for further processing.*

SR$_9$: *If further processing requires the original overlapped latent fingerprint sample to remain intact, the selected separation method shall not employ invasive processing steps.*

SR$_{10}$: *All decisions and actions made by the fingerprint experts must be documented and updated to the chain-of-custody information.*

By no means would these SRs precisely describe the situation, if the separation of overlapped fingerprint were to be conducted in actual forensic work. However, they nevertheless conform to general requirements of forensics. Therefore, they can be considered here as the atomic actions with finest granularity on the system level, which need to be executed by their data level realisation, i.e. the context-aware separation approach this chapter presents.

In this chapter, section 5.1 describes the contextualisation of the forensic scenario and proposes the overlapped fingerprint separation approach with context awareness. Section 5.2 introduces the evaluation on the performance of the proposed approach. Section 5.3 subjectively assesses the developed approach with regards to the preceding exemplary SRs, comparing to the state of the art approaches. Section 5.4 summarises this chapter.

The research presented in this chapter was partially funded within the scope of research project DigiDak[3], and parts of it have been published in [QSS+2012], [QSD2013], and [QSZ+2014].

5.1 A CONTEXT-BASED SEPARATION APPROACH FOR OVERLAPPED LATENT FINGERPRINTS

Embracing a cutting-edge nanometre range contactless sensing technology, in this section a separation approach is developed with compliance with the utilisation of CWL sensor for non-invasive acquisition of the fingerprint evidence. Therefore, the general theoretical framework introduced in Chapter 3 is applied for the scenario of separating overlapped fingerprints in high-resolution images, clarifying various context entities, so an improved separation algorithm can be derived to address them. As the improved algorithm involves parameters which are mutually connected hence

[3] Funded in part by the German Federal Ministry of Education and Science (BMBF) through the research program under the contract no. FKZ: 13N10818.

complicate the situation, the approach also includes a parameter optimisation by further interpreting the context, so the scale of the testing needed can be significantly reduced.

5.1.1 Contextualisation of the forensic scenario

The first step towards a context-based forensic approach is to contextualise its application scenario, i.e. the acquisition and processing environments (see Figure 10 on page 49), the latter also described by the SRs in Example 8.

Aiming at the embodiment of data level security-oriented context modelling proposed in subsection 3.2.3, Figure 18 on page 74 illustrates the forensic scenario discussed in this chapter, with SRs from Example 8 marked at where they are supposed to be executed.

In the acquisition environment, the overlapped latent fingerprints are acquired in a contactless way by using a CWL sensor, yielding high-resolution (typically 2540 ppi) intensity and topography images. In the processing environment, the separation approach processes the acquired intensity image and yield two separated fingerprint. Such separation results can then be used for further forensic investigations, e.g. non-invasive ones like aging estimation, or invasive ones like chemical composition analysis using GC-MS (Gas Chromatography–Mass Spectrometry). Figure 18 also shows that in alternative scenarios the fingerprints can also be acquired from other methods in forms of simulated [CFJ+2011] or conventionally captured overlapped fingerprints [FSZ2012].

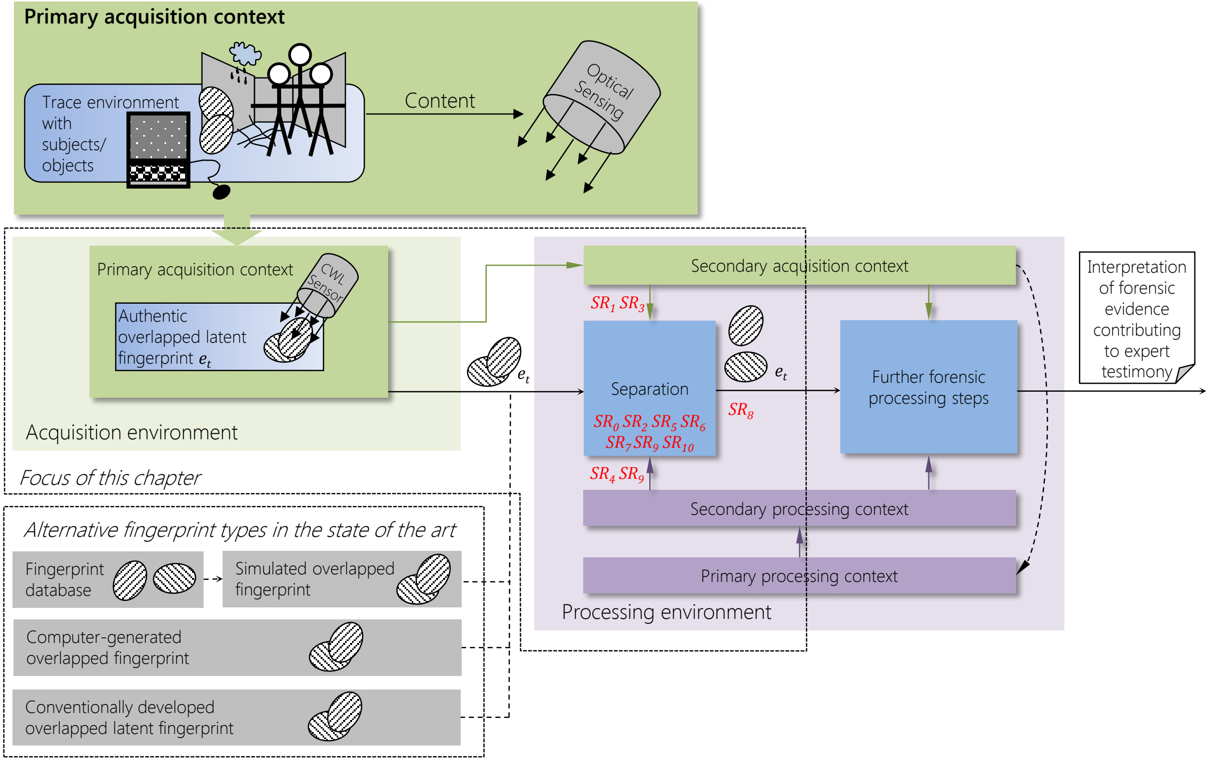

Primary acquisition context
Trace environment with subjects/ objects
Content
Optical Sensing
Primary acquisition context
CWL Sensor
Authentic overlapped latent fingerprint e_t
Acquisition environment
Focus of this chapter
e_t
Secondary acquisition context
SR_1 SR_3
Separation
SR_0 SR_2 SR_5 SR_6
SR_7 SR_9 SR_{10}
SR_4 SR_9
e_t
SR_8
Further forensic processing steps
Interpretation of forensic evidence contributing to expert testimony
Secondary processing context
Primary processing context
Processing environment
Alternative fingerprint types in the state of the art
Fingerprint database
Simulated overlapped fingerprint
Computer-generated overlapped fingerprint
Conventionally developed overlapped latent fingerprint

Figure 18. Embodiment of data level security-oriented context modelling in the forensic scenario, in comparison with other acquisition approaches in literature (adapted from [QSZ+2014])[4]

Recalling the formula (3.6b) from subsection 3.2.3 that defines the acquisition environment and its context:

[4] The original context visualisation was sketched by Jana Dittmann and is/will be part of different figures of several publications to visualise context aspects in different research fields of her research group.

$$\begin{aligned}
&\boldsymbol{AE} = (E_A, O_A, R_A)\\
&o_a \in O_A, o_a \text{ refers to acquistion}\\
&R_A \subseteq E_A \times \ldots \times E_A\\
&E_A = \{e_t\} \cup E_{object} \cup E_{envi}\\
&C_a = \{E_{object} \cup E_{envi}, O_A, R_A\}
\end{aligned} \tag{3.6b}$$

the acquisition context C_a regarding its TE, which in this case is the overlapped fingerprint image and denoted bye_t, comprises information introduced to the image from two aspects:

- the own properties of the fingerprint sample itself (denoted by E_{object}), e.g. the appearance of the fingerprint sample,
- the properties of the acquisition mechanism (denoted by E_{envi}), e.g. the acquisition devices and the operating human agents.

Such information comes from the acquisition environment with a coarse granularity. Therefore, it is considered as *primary acquisition context*, as previously defined in subsection 3.2.3. It covers a general cluster of context entities E_A (in this case the cluster of various properties) as well as the operations O_A and relations R_A on E_A. As E_A, O_A and R_A are unspecified in the primary acquisition context, they need to be interpreted to form the *secondary acquisition context*, so it can be further addressed by specific processing procedures.

Table 13 summarises such interpretation: the specific CEs in E_A that are relevant to the proposed separation approach are interpreted from E_A, and listed with remarks on if they are also addressed in the state-of-the-art literature, while O_A and R_A together are specified to identify the connection between those entities and the fingerprint image e_t. The table also highlights the $\mathrm{e}_{\mathrm{sc}}^{\mathrm{Ai}}$ occurring only in the acquisition environment particularly discussed in this chapter. As shown by the table, the secondary acquisition context shares the same TE (the overlapped fingerprint image e_t) with the primary acquisition context.

E_A in primary acquisition context	***Specified E_A for secondary acquisition context***		***Specified O_A, R_A for secondary acquisition context***	***Remarks***
Information regarding properties of the fingerprint sample E_{object}	Type of fingerprint sample e_{sc}^{A0}		Various types of overlapped fingerprint images (e.g. authentic/simulated/computer generated) contributes to different levels of difficulty for separation	Test sets in previous publications consist of single or multiple types of fingerprints.
	Overlapping behaviour e_{sc}^{A1}	Contact pressure e_{sc}^{A11}	Image intensity reflecting the contact pressure	Most image processing based approaches work on intensity images, e.g. [TWZ+2001], [STK2006], [CFJ+2011], [FSZ2012]
		Overlapping angle e_{sc}^{A12}	Various intersection situations of the ridges in the image	-

	Smearing e_{sc}^{A13}	Smeared regions on the fingerprint images	Smearing affects chemical approaches less, e.g. [TLC+2010]
	Overlapping percentage e_{sc}^{A14}	Area of overlapped and nonoverlapped regions	Important for relaxation labelling based approaches, e.g. [CFJ+2011], [FSZ2012]
	Trace source e_{sc}^{A2}	Degree of similarity among the images reflecting the trace source (same finger, different fingers from same person, or different fingers from different persons)	-
	Chemical trace composition e_{sc}^{A3}	Image intensity partially reflecting the chemical composition	Tackled in [BSF+2009]
Information regarding properties of the acquisition environment E_{envi}	Substrate e_{sc}^{A4}	Extra noise in the image reflecting the substrate that the fingerprint is left on	-
	Fingerprint development technique (in case of latent fingerprint) e_{sc}^{A5}	Image properties (e.g. quality, resolution, appearance) dominated by the development technique employed on latent fingerprints	[FSZ2012] involves different development techniques, yet does not discuss further.
	Sensor settings e_{sc}^{A6}	Image quality reflecting the parameter settings of the sensor	-

Table 13. Acquisition context entities (CEs) for separating overlapped fingerprints, with those occurring only in the acquisition environment in this chapter highlighted in blue (adapted from [QSZ+2014])

As revealed by the SRs in Example 8, multiple aspects of security requirements are expected to be met regarding the TE (e.g. regarding its integrity as stated in SR4), whose quality is directly influenced by the CEs identified in Table 13. Therefore, in this case all such CEs are security relevant and considered as e_{sc}, denoted in the table as e_{sc}^{An}, where $n = 0,2,\dots,6$.

After the secondary acquisition CEs are clarified, they can contribute to the formation of primary processing context by analysing the outcome of processing them. In the case, the processing environment is regulated by the SRs Example 8. To meet the requirements of these SRs, the relaxation labelling based processing approach introduced in [CFJ+2011] is selected as the basis, into which further enhancements with context awareness is introduced to develop the processing environment in this chapter. Comparing to other state-of-the-art separation approaches, the relaxation labelling based on shows the best compatibility with the contextualised forensic scenario: First, it requires only a digital representation of the overlapped latent fingerprints, so it is not mandatory to apply invasive methods to acquire the fingerprints – in fact it does not even require for the physical presence of the evidence. Second, it is based on image processing, so it also does not introduce further invasive manipulation of the evidence. At last, as depicted in Table 5 on page 32, it does not introduce any sort of interpolation or other similar ways to generation new data, so it does not introduce any risk of potentially tampering of the evidence. Therefore, due to the fact that the basic processing method already exists in a coarse way, it is relatively easier in this case to analyse the outcome of processing the acquisition context with the method. Among all the e_{sc}^{An} listed in Table 13, it is the overlapping behaviour entity e_{sc}^{A1} that poses the main issue to be solved. The relaxation labelling that the processing procedures utilise is an iterative method that sorts extracted fingerprint ridge orientations (see Table 5 in subsection 2.5.2 on page 32). Orientation being the core element of relaxation labelling, the overlapping behaviour of two fingerprints is eventually represented by that of two orientations. Therefore, as the outcome of using the processing procedures described in Table 5 on the overlapping behaviour entity, the corresponding primary processing context can be derived, in the form of three separation error classes, which constitutes the primary processing CEs e_{sc}^{pn} $(n = 0,1,2)$:

- the class of *orientation extraction errors* e_{sc}^{p0} denotes the incorrect extraction of the dominant orientations in the overlapped region, despite they are perceptual in the image,
- the class of *labelling errors* e_{sc}^{p1} denotes the correct extraction yet incorrect labelling of the dominant orientations, and
- the class of *merging errors* e_{sc}^{p2} denotes the false matching of the two orientation fields achieved by the relaxation labelling to wrong nonoverlapped regions.

Addressing the identified primary processing context, an improved algorithm is developed to introduce context awareness, with highlights on its improvement with regard to the one in [CFJ+2011]. It is described in the following Table 14.

<table>
<tr><th><em>Processing steps</em></th><th><em>Input</em></th><th><em>Output</em></th></tr>
<tr><td>1. Input overlapped fingerprint image</td><td>Authentic overlapped latent fingerprint image acquired using CWL sensor</td><td>$I_{auth}: H_{image} \times W_{image}$</td></tr>
<tr><td>2. Manually assign fingerprint masks</td><td>I_{auth}, M_1, M_2:
$H_{image} \times W_{image}$</td><td>$I_{auth} \circ M_{N1}, I_{auth} \circ M_O, I_{auth} \circ M_{N2}$</td></tr>
<tr><td>3. Context-based parameter calculation</td><td>$I_{auth}, M_1, M_2, M_{ref1}, M_{ref2}$</td><td>$\{\widehat{win}_1, \widehat{win}_2, \dots, \widehat{win}_n\}$</td></tr>
<tr><td colspan="3">The step starts with the estimation of the block size parameters depending on the ridge distances. The investigator assigns two additional masks, M_{ref1} and M_{ref2}, respectively locating in each nonoverlapped regions, covering areas with mainly parallel and smoothly curved ridges. Then Fourier representations of both masked areas are achieved with a reference window size w_{ref}, which is calculated based on the image resolution R_I in ppi and the statistical average ridge distance 0.46 mm [Moore1989]:

$$w_{ref} = \frac{0.46 \cdot R_I}{2540} \cdot 4 \cong R_I/1380$$

$$RR_k = \{rr_{k,1}, rr_{k,2}, \dots, rr_{k,m}\} = \text{Block}(I_{auth} \circ M_{ref[k]}, w_{ref}),$$
$$rr_{k,i}: w_{ref} \times w_{ref}, 1 \leq i \leq m, k \in \{1,2\}$$
$$\widehat{rr}_{k,i} = \text{DFT}(rr_{k,i})$$

Then the frequency vector with highest amplitude is extracted in each window and used to</td></tr>
</table>

calculate the actual block size, so each block covers one ridge:

$$d_k = \frac{1}{m}\sum_{i=1}^{m}\frac{w_{ref}}{\left|\underset{\vec{f}\in\widehat{rr}_{k,i}}{\operatorname{argmax}}\left|\widehat{rr}_{k,i}(\vec{f})\right|\right|}$$

$$b = \max(d_k)$$

Afterwards a series blocks are generated on the masked fingerprint area:

$$\{bl_1, bl_2, \dots, bl_n\} = \operatorname{Block}(I_{auth} \circ M_1 \circ M_2, b), bl_i\colon b \times b, 1 \le i \le n$$

Subsequently a series of DFT windows $\{win_1, win_2, \dots, win_n\}$ are generated as previously written, using $w = 4b$. At last, for each Gaussian filtered window its two-dimensional spectrum is calculated, with a σ selected so that two ridges in the centre of each window get clearly filtered:

$$\sigma = \frac{3}{2}b \cdot \frac{1}{\sqrt{-2\ln 0.35}}$$

$$\widehat{win}_i = \operatorname{DFT}\big(\operatorname{Gauss}(\mathrm{win_i}, \sigma)\big)$$

4. Extract the dominant orientations for each block in non-overlapped and overlapped regions	$\{\widehat{win}_1, \widehat{win}_2, \dots, \widehat{win}_n\}$	$f_{max1,i}, f_{max2,i}, O_N, O_O$

The two frequencies with the highest and the second highest amplitudes are selected in each window for its corresponding blocks, and contribute to the orientation field $O(p, q, k)$, where p and q denote the position of the block and k the label:

$$f_{max1,i} = \underset{f\in\widehat{win}_i}{\operatorname{argmax}}\left|\widehat{win}_i(f)\right|$$

$$f_{max2,i} = \underset{f\in\widehat{win}_i \wedge \delta(|\arg(f_{max1,i}) - \arg(f)|) > \tau}{\operatorname{argmax}}\left|\widehat{win}_i(f)\right|, \tau = 22°$$

$$O_O, O_N \text{ remain unchanged}$$

5. Perform relaxation labelling on the overlapped region together with the boundaries	O_O, O_N, ε	$O'_O\colon H_{blocks} \times W_{blocks} \times 2$

In this step relaxation labelling is performed on not only the overlapped region but also predefined boundary regions, where the labels already discriminate the origins of both nonoverlapped orientation fields. Therefore, during relaxation labelling the labels of the overlapped regions converge to the respective labels of boundary regions.

$$M_B(p,q) = \begin{cases} 1 & \text{if the block } (p,q) \text{ is inside the boundary with a width of } \varepsilon, \varepsilon = 2 \\ 0 & \text{otherwise} \end{cases}$$

$$O'_O(p,q,k) = \begin{cases} O_O(p,q,k) & M_O(p \cdot b, q \cdot b) = 1 \\ O_N(p,q,k) & M_B(p,q) = 1 \end{cases}$$

$$s_{kk'} = 1 - \frac{\delta(|O'_O(p,q,k) - O'_O(p',q',k')|)}{\frac{\pi}{2}}, k, k' \in \{1,2\}$$

$R_{pq,p'q'} = \begin{cases} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & \text{if } \lvert p-p' \rvert > 6 \text{ or } \lvert q-q' \rvert > 6 \text{ or } (p=p' \text{ and } q=q' \text{ and } k=k') \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \text{if } p=p' \text{ and } q=q' \text{ and } k \neq k' \\ \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} & \text{if } M_B(p,q)=0 \text{ and } M_B(p',q')=0 \\ \begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{11} \end{bmatrix} & \text{if } M_{N1}(p \cdot b, q \cdot b)=1 \text{ and } M_B(p',q')=0 \\ \begin{bmatrix} s_{22} & s_{21} \\ s_{21} & s_{22} \end{bmatrix} & \text{if } M_{N2}(p \cdot b, q \cdot b)=1 \text{ and } M_B(p',q')=0 \end{cases},$ $p_{pq,k1}, p_{pq,k2}, \alpha$ and t_{max} remain unchanged.		
6. Combine the two separated orientation fields with non-overlapped regions with corresponding labels	O_N, O'_O, M_O	O_1, O_2
Due to the involvement of boundary regions in step 5, this step simply combines each nonoverlapped region with the separated one with its corresponding label to from the two complete orientation fields. $O_k(p,q) = \begin{cases} O'_O(p,q,k) & \text{if } M_O(p \cdot b, q \cdot b) = 1 \\ O_N(p,q,k) & \text{otherwise} \end{cases}$		
7. Apply Gabor filter to render the separation results	O_1, O_2	I'_1, I'_2
At last the merged orientation for each block from step 6 are used as the input of a Gabor filter applied on the corresponding block of the fingerprint. $I'_k(p,q) = \text{Gabor}\left(bl_{p \cdot b+q}, O_k(p,q)\right)$		

Table 14. The improved separation algorithm, with its modification w.r.t. the algorithm in Table 5 highlighted in blue and parameters optimised in subsection 5.1.2 marked in red (reproduced from [QSZ+2014])

From the point of view of data level security-oriented context modelling (see Figure 10 on page 49 and Figure 18 on page 74), the algorithm proposed here contributes to the secondary processing context: it is fundamentally a cluster of sorted processing steps that execute a series of operations between the TE (in this case the fingerprint ridges) and its CEs (which are various in different processing steps, including the assigned fingerprint masks, the appointed parameters, the iteratively updated labels, etc.). Such secondary processing context addresses the primary one by introducing two new features, comparing to the previous algorithms in [CFJ+2011], [QSS+2012], and [QSD2013]:

- An angle threshold is introduced to address the orientation extraction errors (e_{sc}^{p0}). One of the complications that come with the high-resolution images is that the ridges are largely magnified to become "stripes" rather than "lines". Such "stripes" come with a certain width and are made up of particles with various sizes, shapes and intensities. Therefore, it is often that such heterogeneity of the ridges yield multiple similar orientations, which would be erroneously extracted as dominant orientations in a block and conceal the actual second dominant orientation. With the assignment of angle threshold, after the first dominant orientation is extracted, the second one could appear in the following three ways:

1. Its angle difference to the first one is larger than the value of the threshold, so it can be correctly located and extracted, while the automatic omission of the candidate orientations that are close, i.e. similar, to the first one.
2. Its angle difference to the first one is smaller than the angle threshold, thus it cannot be extracted. However, in case of properly chosen value for the threshold, this would only happen when the two ridge lines in the block are completely overlaid on each other and appearing as one, so they are essentially not separable in the first place. Therefore, for the sake of further forensic analysis, it is more reasonable to output no result instead of a potential misleading one, so no second orientation is rendered in this case.
3. There exists no second dominant orientation at all, due to either a ridge ending or relatively weak intensity in the block. The frequency representation of this situation would be similar to that of last case, and it would also yield the lack of second orientation rendered in the block, which is consistent with its physical appearance.

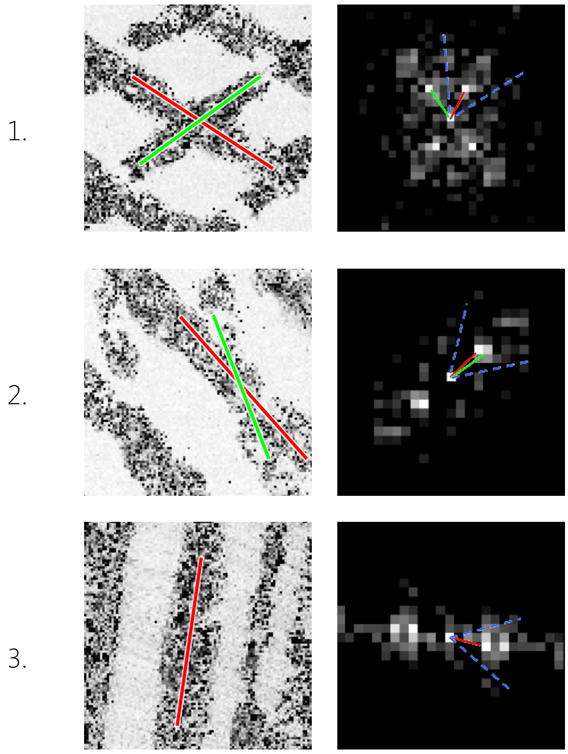


Figure 19. Three overlapping behaviours in DFT windows, with the central parts of their frequency representations on the right side, with red line referring to the first dominant orientation, green the second, and blue dashed lines denoting the angle threshold applied in the frequency domain (re-sketched from [QSZ+2014])

Figure 19 shows the three preceding situations respectively, together with their representations in frequency domain. Notice here that the figure only illustrates the zoomed-in central parts, so the bright pixels are more perceptual, corresponding to the dominant orientations.

- The enhanced algorithm uses the orientation information of the boundaries already at the beginning of the relaxation labelling step, instead of in the merging step (as in [CFJ+2011]), to address labelling and merging errors (e_{sc}^{p1} and e_{sc}^{p2}). In [CFJ+2011], the application of relaxation labelling involves only the overlapped region. The two resulting separated orientation fields are subsequently taken over by the merging step, which is supposed to match them to their corresponding nonoverlapped regions. Despite that this merging step takes boundary information into its consideration, labelling and merging errors still often occur in previous tests. After investigating those testing results, it appears that most of such errors occur because some areas of the overlapped region on the boundaries come with relatively weak quality (due to smeared ridges, dust or fibres on the substrate, etc.), causing labelling errors. Because these errors appear on the boundaries, they directly compromise the correctness of the subsequent merging step, causing merging errors. Addressing this issue, the proposed algorithm takes the boundary information into consideration in the relaxation labelling step (as shown in Table 14), so it yields two orientation fields which already match their corresponding boundaries hence also the corresponding nonoverlapped region. Therefore, no merging step is needed, and this is also beneficial from the point of view of computational complexity. In Figure 20 a typical example is given, showing the improvement of separation performance that this modification introduces.

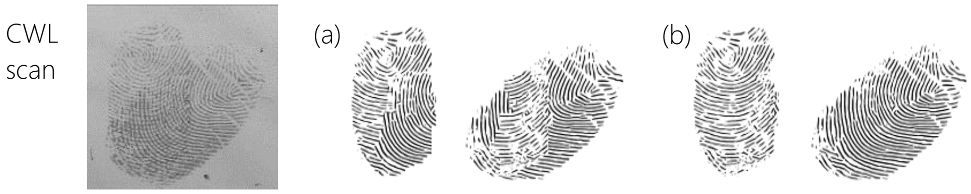


Figure 20. Comparison of separation results using (a) the previous approach from [CFJ+2011], (b) the current approach (re-sketched from [QSZ+2014])

This proposed algorithm differs from the one from [FSZ2012] in the following two aspects:

- [FSZ2012] treats each overlapped block as single object to simplify the relaxation labelling procedure. However, as explained later, in this scenario the high-resolution brings more details, resulting more complications to the overlapping behaviour of the orientations. Therefore, the algorithm here treats each single orientation separately for a more precise separation decision.
- Despite it also utilises the information from the nonoverlapped regions in the relaxation labelling procedure, the algorithm here calculates the compatibilities differently with [FSZ2012].

Furthermore, as already marked in Table 5 and Table 14, the proposed algorithm introduces six parameters, which contributes further to secondary processing context:

- Iteration control factors α and t_{max} are inherited from the approaches in [CFJ+2011]. With regard to the convergence of the iterative relaxation labelling, the parameter α works as the oscillation factor, while t_{max} is the iteration time (see Table 5)
- Angle threshold τ in degrees ensures the second dominant orientation to be correctly identified in each block, as explained earlier (see Table 14);
- DFT window size w (see Table 14);
- Gaussian factor σ controls the Gaussian filter that is applied on each DFT window to filter out the ridges in its centre region (see Table 14);
- Boundary width ε is the width in blocks of the boundaries used in the relaxation labelling step (see Table 14).

5.1.2 Context-based parameter optimisation

As mentioned previously, the parameters involved in the implementation of the proposed algorithm also serve as part of the secondary processing context to its TE, the fingerprint ridges. As shown in Table 14, the assigned value of the parameters directly affect the outcome of the processing steps, hence the correctness of the final separation results. Therefore, this subsection describes a context-based parameter optimisation, yielding optimised parameter values. As described in formula (3.3g) in subsection 3.1.2, it is possible that an entity in a context model serves as context for other entities and at the same time has its own surrounding context. This is revealed here: to perform the optimisation, the parameters themselves are considered in this case as TEs, while their corresponding CEs are further identified and derived with reference to both acquisition and processing context in two classes: *inter-parameter context*, i.e. the context that reflects how one parameter correlates with the rest, and *intra-parameter context*, i.e. the context that constrains one parameter due to its physical nature. Table 15 summarises the functionalities of all the parameters investigated here, together with their inter- and intra-context entities. The clarified CEs are highlighted with different colours according to the sources they are derived from. Based on these CEs, the corresponding parameter optimisation can be further conducted.

Parameter	*Functionalities*	*Intra-parameter context*	*Inter-parameter context*
α	Controls the initial amplitude of the oscillation of the iterative relaxation labelling	> 0 to start an oscillation requires an upper limit for the sake of computational complexity	positively correlated with t_{max}
t_{max}	Controls the duration of the iterative relaxation labelling	> 0 to start an oscillation saturates when growing larger than certain value after the oscillation converges requires an upper limit for the sake of computational complexity	positively correlated with α
τ	Controls the area around the first extracted dominant orientation eliminated from the extraction of the second dominant orientation	in degrees > 0° to make the threshold to function ≤ 45° as the DFT domain is centro-symmetric	-
w	Controls the size of the DFT window processing the block, therefore the number of ridges in each DFT window	in pixels reflects the number of ridges in each DFT window shall enable each DFT window to cover enough number of ridges for the following Gaussian filter requires an upper limit for the sake of computational complexity	resulting DFT window shall exceed the Gaussian filtered area

σ	Controls the area of filtered centre region of DFT window, therefore the number or filtered ridges	reflects the number of ridges being filtered must filter appropriate number of ridges for orientation extraction requires upper limit as the Gaussian filter must filter at least one ridge	resulting filtered area must not exceed the DFT window
ε	Controls the amount of boundary orientation information for relaxation labelling	in blocks ≥ 1 to form a boundary requires an upper limit as orientation bias accumulates requires an upper limit for the sake of computational probability	negatively correlated with ***b***

Table 15. Parameters and their intra- and inter-context entities, those derived from acquisition context are highlighted in red, while those from processing context in blue (reproduced from [QSZ+2014])

Optimisation of the iteration control factors α and t_{max}

As shown in Table 15, parameters α and t_{max} control the iterative relaxation labelling. The values of both parameters show a positive correlation with each other, and both stay in certain ranges. Therefore, here they are investigated together, aiming at a pair of proper value settings. In the optimisation, one of the two parameters is assigned with a constant while adjusting the other, followed by the visual examination of the resulting separation results. This procedure is repeated on 10 images of authentic overlapped latent fingerprints.

Figure 21 shows one group of separation results with various α values with t_{max} being constantly assigned as 2500. Obviously the separation quality is best when $\alpha = 0.2$ or 0.3. As similar observations can be made in all 10 groups of results from the tested images, it is concluded here that 0.2 or 0.3 is the most suitable value for parameter α. For the sake of not only the consistency but also a lower computational complexity, $\underline{\alpha = 0.2}$ is chosen for later tests.

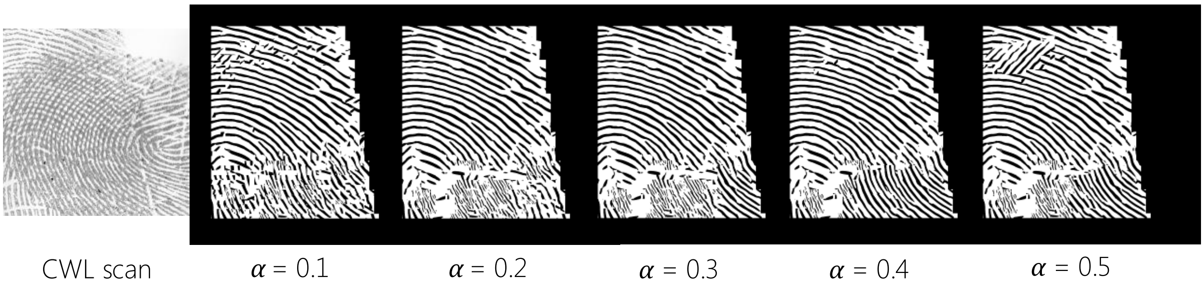


Figure 21. Separation results (left print only) with t_{max} = 2500 for various α values (applying the minimum granularity value 0.1 allowed by the implementation) (re-sketched from [QSZ+2014])

Figure 22 shows one group of separation results with various t_{max} values with α being constantly assigned as 0.2. As shown in the figure, with a constant α value, the increase of t_{max} contributes in a logarithmic behaviour to the accuracy of separation. Considering the similar observations in all results of the 10 different original scans, it is concluded here that the most suitable parameter value is $\underline{t_{max} = 2500}$.

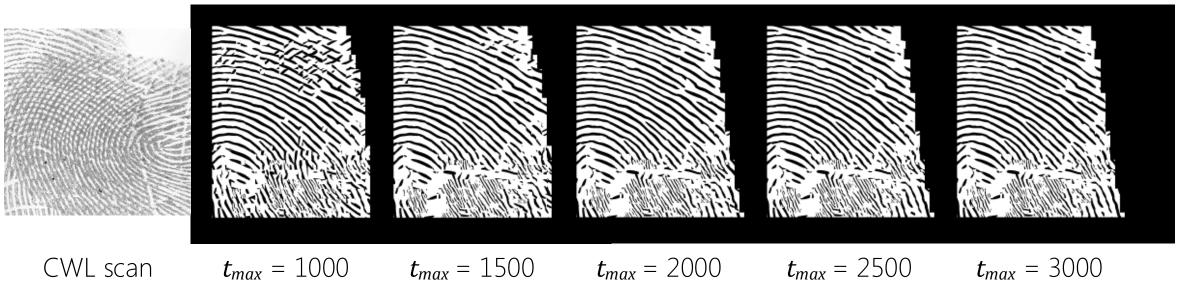


Figure 22. Separation results (left print only) with α = 0.2 for various t_{max} values (applying a granularity value of 500 which yields enough perceptual difference on the results) (re-sketched from [QSZ+2014])

As α and t_{max} are closely correlated, there could exist other combinations of larger values of these two parameters, also yielding satisfactory convergence. However, the convergence would nevertheless consumes longer time, as stronger oscillation that comes with a larger α value would take longer to converge. Therefore, the least time-consuming pair of suitable values is selected here.

Optimisation of the angle threshold τ

The angle threshold τ is introduced with the new features described in subsection 5.1.1. As shown in Table 15, the value of this parameter lies in the range of (0°, 45°], due to the centrosymmetry of the DFT domain in which the threshold functions. As explained, this parameter is set to address the error caused by the appearance of multiple orientations from single ridge, and this phenomenon can be assumed to occur only randomly in the blocks. Considering the sample space (which covers all the blocks in all the fingerprint images in the test sets) being big enough, it is reasonable to assume that the occurrence of such phenomenon follows a normal distribution. Therefore, the mean value of the range (0°, 45°] is the best candidate. In the implementation, the integer $\underline{\tau = 22°}$ is used to avoid additional computational complexity of using floating-point numbers.

Optimisation of the DFT window size w

As explained in Table 15, its CEs constrain the parameter *w* in following aspects:

- A DFT window needs to cover more ridges than the block in its centre does, thus the value of *w* requires a lower limit;
- A DFT window needs to also cover enough number of ridges for the following Gaussian filter;
- Considering computational complexity, the value of *w* also requires an upper limit.

As Table 14 shows, the block size *b* is assigned with the measured ridge distance so that each block covers one ridge. Therefore, here *w* is assigned with four times the block size, so that each DFT window covers around four ridges, addressing all three preceding aspects.

Optimisation of the Gaussian factor σ

The Gaussian factor σ decides the area of the filtered centre region of a DFT window, therefore also decides the number of ridges that directly contribute to the extracted orientations. As described in Table 14, here the σ value is set in a dynamic way that two ridges clearly appear in the centred filtered region.

Optimisation of the boundary width ε

According to Table 15, the value of boundary width *ε* needs to be assigned in a range: the lower limit of such range is trivially 1, as a boundary must be at least one block wide, and the upper limit exists for the following two aspects of its context:

- The orientation bias accumulates with the growth of the distance between the block in the boundary and the overlapped region;
- The value of *ε* should be as small as possible for the sake of computational complexity.

Additionally, as its value is in blocks, it is reasonable that a larger *b* yields a smaller *ε*, and vice versa.

A test is therefore conducted to approach the optimised value of *ε*. Similar to previous tests for *α* and t_{max}, 10 images of authentic overlapped latent fingerprints are processed with fixed parameter setting except for various *ε* values. Figure 23 shows a representative group of test results, where separation errors are highlighted with red ellipses.

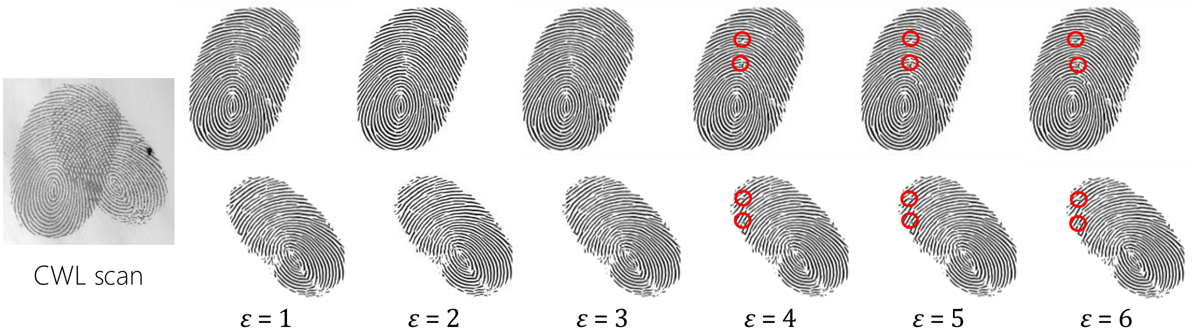
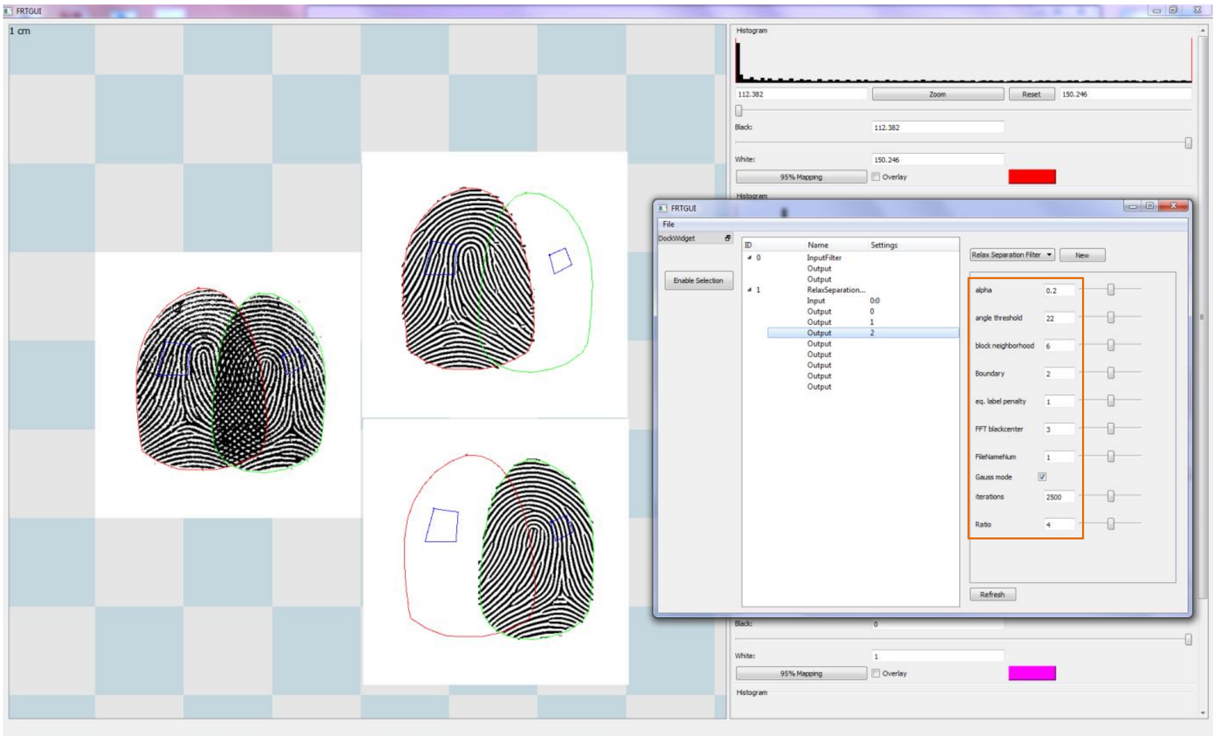


Figure 23. Separation results with various *ε* values, while the rest of parameter settings appointed as optimised previously (*α* = 0.2, t_{max} = 2500, *τ* = 22°, *w* = 4*b*), with separation errors marked with red ellipses (re-sketched from [QSZ+2014])

As shown by the results in Figure 23 together with the rest nine groups of them, better separation results can be observed when *ε* is assigned as 1, 2, or 3. To achieve the balance in the trade-off between the amount of boundary information and the computational complexity, it is concluded here that *ε* = 2 is most suitable.

With the preceding optimised parameters, the developed algorithm is realised with C++ within the scope of the forensic processing toolkit developed in DigiDak research project. Figure 24 shows the user interface of the toolkit that implements the separation algorithm as one of its processing components. As shown in the figure, with the help of the user interface, the user (in this case fingerprint expert), has full access to the whole separation process: he can easily import the overlapped latent fingerprint image, apply the fingerprint masks (rendered in red and green in Figure 24) together with the additional reference masks (rendered in blue in Figure 24), assign the parameters (highlighted in the orange rectangle in Figure 24), initialise the separation process, and observe and store the separation results. It should be noted here, that for the demonstration the exemplary overlapped fingerprints shown in the figure being separated is generated artificially using SFinGe[5] [Cappeli2009], printed using the method introduced in [HSD+2013] and captured using CWL sensor.



112.382 150.246

Black: 112.382

White: 150.246

95% Mapping □ Overlay 🟥

ID	Name	Settings
0	InputFilter	
	Output	
	Output	
1	RelaxSeparation...	
	Input	0:0
	Output	0
	Output	1
	Output	2
	Output	
	Output	
	Output	
	Output	
	Output	

Relax Separation Filter

alpha	0.2
angle threshold	22
block neighborhood	6
Boundary	2
eq. label penalty	1
FFT blackcenter	3
FileNameNum	1
Gauss mode	☑
iterations	2500
Ratio	4

Black: 0

White: 1

95% Mapping □ Overlay 🟪

Figure 24. User interface of the toolkit implementing the proposed separation algorithm, developed within the scope of DigiDak research project[6]

5.2 Evaluation with optimised parameters

There are various ways of evaluating the separation results regarding their correctness and applicability for subsequent forensic testing procedures, e.g. assessing the quality of the fingerprint-like pattern in the separation results as described in [HMQ+2013]. This section assumes for the primary goal of dactyloscopy, which is to examine fingerprints and yield identifications of persons of interest, thus it applies here a biometric-based evaluation mechanism. The evaluation is performed in two stages: A preliminary stage aims to establish the premise that the original CWL scan of overlapped fingerprint is not applicable even for single identification, followed by the main stage that focuses on assessing the correctness of the separated fingerprint ridges.

5.2.1 Test sets generation

Two test sets are generated for the evaluation. Test set 1 (TS1) is generated as follows to address the secondary acquisition CEs:

- *Trace source* e_{sc}^{A2}: TS1 comprises four fingerprints from two different volunteers. FP1 and FP2 are collected from a male volunteer, while FP3 and FP4 from a female one. As a compromise between the highest potential similarity with same finger and highest

[5] "Fingerprint generation", Biometric System Laboratory, DISI, University of Bologna, http://biolab.csr.unibo.it/research.asp?organize=Activities&select=&selObj=12&pathSubj=111%7C%7C12&, (accessed on Jun. 18, 2014).

[6] The author thanks Marcus Leich, Wolfgang Büschel, Maik Schott, and Wenju Zheng for their contribution to the implementation.

potential dissimilarity with different fingers from different persons, the test set employs different fingers from same person, i.e. FP1 overlaid on FP2, and FP3 on FP4.

- *Substrate* e_{sc}^{A4}: three substrates are used, namely hard disk platter (Sub1), white paint-coated metal surface commonly seen on kitchen furniture (Sub2), and aluminium foil (Sub3).
- *Overlapping behaviour* e_{sc}^{A1}: The volunteers are asked to press their appointed fingers on the substrates with random pressure, overlapping angle and percentage, but not to smear. However, the occurrence of the resulting overlapped fingerprints exhibit nevertheless slight random behaviour of smearing.
- *Sensor settings* e_{sc}^{A6}: The sample images are acquired by scanning the overlapped latent fingerprints using a CWL sensor with the settings of scanning frequency at 2000 Hz and dot distance at 10 μm, to achieve a high resolution (in this case 2540 ppi) and image quality. Depending on the actual fingerprint area, a single scan typically takes 1–2 hours.

TS1 consists in total of 60 sample images, among those 28 on Sub1, 8 on Sub2, and 24 on Sub3.

Furthermore, the evaluation also employs the "Tsinghua Overlapped Latent Fingerprint Database (Tsinghua OLF)" published in [FSZ2012] as a second test set (TS2). TS2 consists of 100 latent overlapped fingerprint images at 500 ppi, developed conventionally with fingerprint powder. Despite that the separation approach introduced in this chapter is not developed or tuned for these samples, it is nevertheless worth investigating if it can still be used for lower resolution samples that come with different context.

5.2.2 Evaluation methods

The evaluation utilises NBIS (NIST Biometric Image Software) for fingerprint verification. NBIS applies the MINDTCT (minutiae detector) algorithm to locate and identify the minutiae in fingerprint images, and the Bozorth3 matcher to determine a matching score between two fingerprint images [GWM+2004]. The higher the matching score is, the more similar the two fingerprints are. As shown in Figure 25, two kinds of templates are used in the evaluation: one is the original CWL scan image (TempO), the other is the black and white representation of TempO (TempBW), which is binarised by processing them only using the processing steps 1 – 4 and 7 described in Table 14 on page 55.

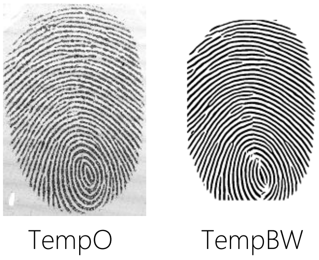


Figure 25. Examples of the two types of templates used for verification (re-sketched from [QSZ+2014])

The evaluation starts with the preliminary stage. In this stage, a group of matching scores are run using NBIS on 20 original overlapped scans from TS1 with their corresponding TempO images. This stage aims to explore for the applicability of overlapped fingerprint without separation as forensic evidence.

After that, the evaluation proceeds with the main stage. In this stage, separation is performed on both TS1 and TS2, with the optimised parameter setting derived in subsection 5.1.2, i.e. $\alpha = 0.2$, t_{max}

= 2500, $\tau = 22°$, $w = 4b$. Therefore, it achieves 120 separation results from TS1 and 200 from TS2. Subsequently, NBIS (NIST Biometric Image Software) is applied cross-examine the separation results and the templates, i.e. matching scores are achieved between each separation result and all templates of two types. After that, these scores are compared with the previous ones to observe performance differences. FAR (False Acceptance Rate) and FRR (False Rejection Rate), together with the ROC (Receiver Operating Characteristic) curves are plotted to reach the proper threshold score.

5.2.3 Evaluation results

The matching scores achieved in the preliminary stage are illustrated in Figure 26. As clearly shown by the figure, the overlapped fingerprint, even acquired with high resolution hence high richness of detail, is not applicable in forensics, as neither the minutiae from the complete scan nor those from nonoverlapped parts are capable of yielding a convincing identification result.

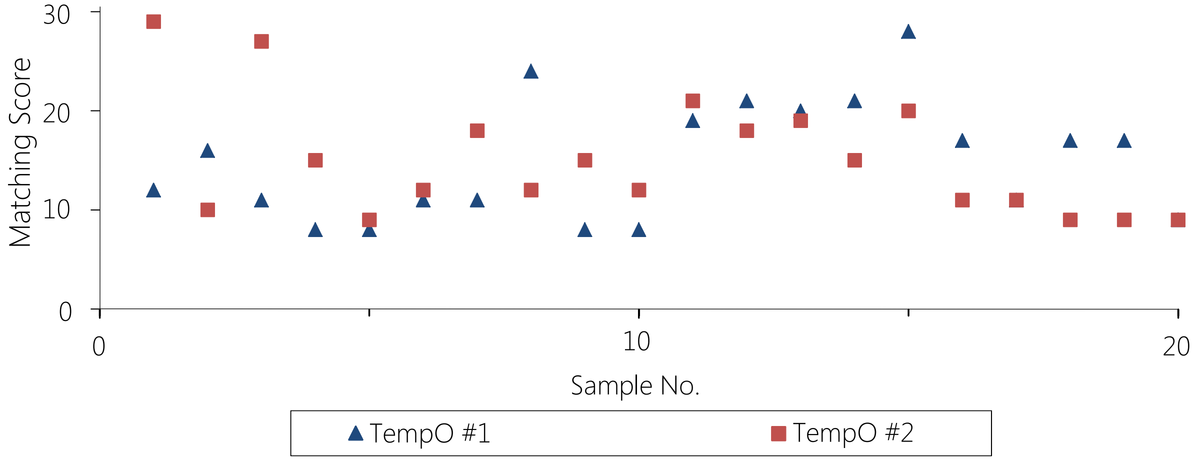


Figure 26. Matching scores between 20 samples from TS1 and their corresponding TempO images (re-sketched from [QSD2013])

Figure 27 on page 89 illustrates the matching scores of each separation result on TS1 with all four TempO templates in (a) and all four TempBW templates in (b). In both subfigures, series 1 refers to the scores with the corresponding template, whereas the remaining three the non-corresponding ones. Among all 120 separation results, no. 1-56 are on Sub1, no. 57-72 on Sub2, and no. 72-120 on Sub3. Observing both (a) and (b), the separation results on all three substrates show clear tendency of yielding correct matching, whereas the type of template poses no significant influence on the overall performance. Figure 28 on page 90 illustrates the ascending FRR curve based on the 240 matching scores from series 1 in Figure 27 (a) and (b), and the descending FAR curve based on the 720 matching scores from the remaining three series in them. The intersection of two curves indicates an EER of 5.7% with the corresponding threshold value of matching score of 14. This is also confirmed with the ROC curve illustrated in Figure 29 on page 90. Similar to Figure 28 and Figure 29, in Figure 30 (page 91) and Figure 31 (page 91) FAR, FRR and ROC curves are also plotted for the total 4800 matching scores for TS2, and an EER of 17.9% is indicated, with the corresponding threshold of 8.

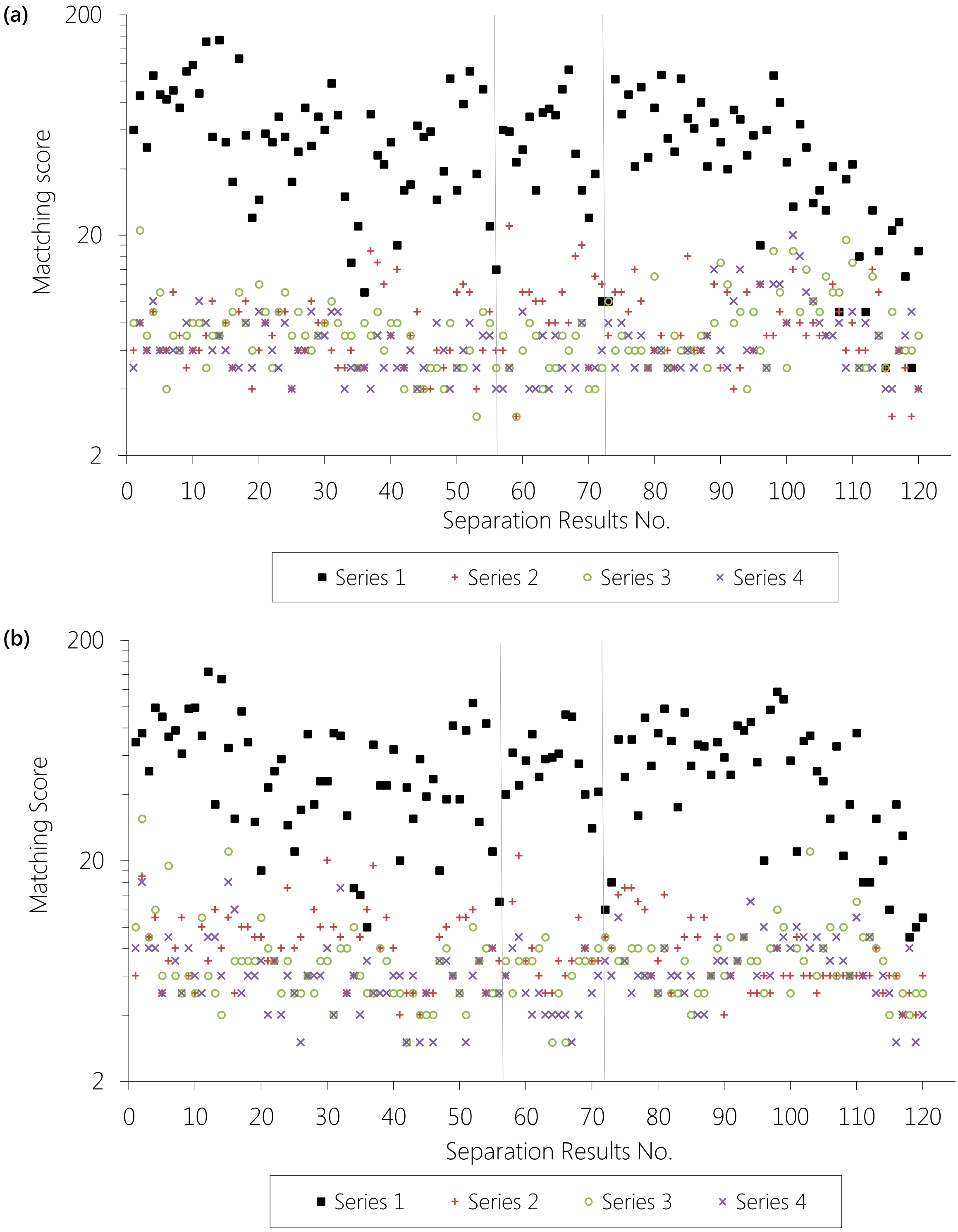


Figure 27. Matching scores (log scale) on TS1 with (a) TempO and (b) TempBW

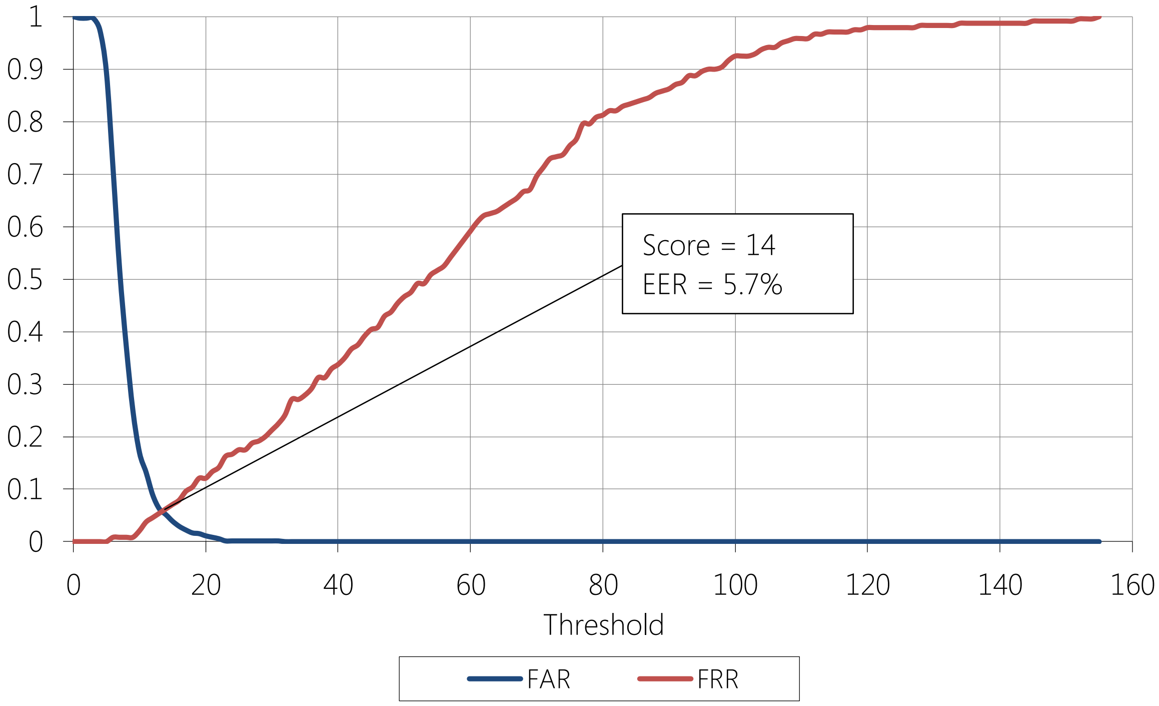


Figure 28. FAR/FRR curves of TS1 (re-sketched from [QSZ+2014])

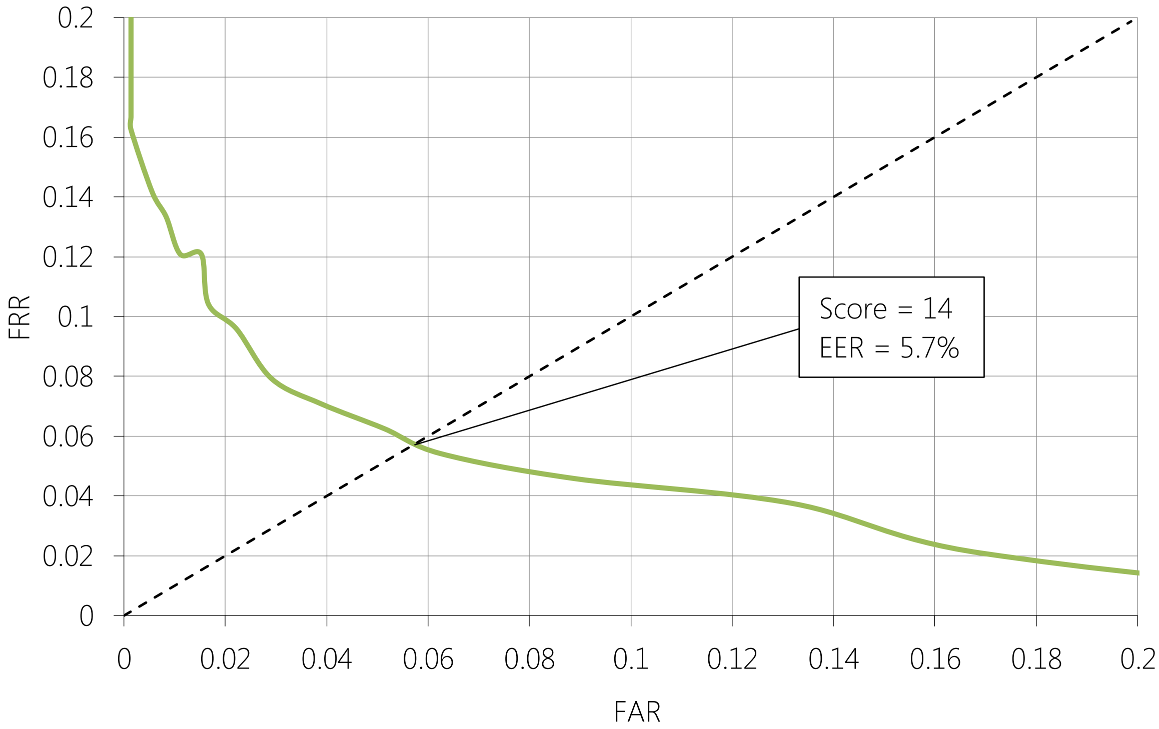


Figure 29. ROC curve of TS1 (re-sketched from [QSZ+2014])

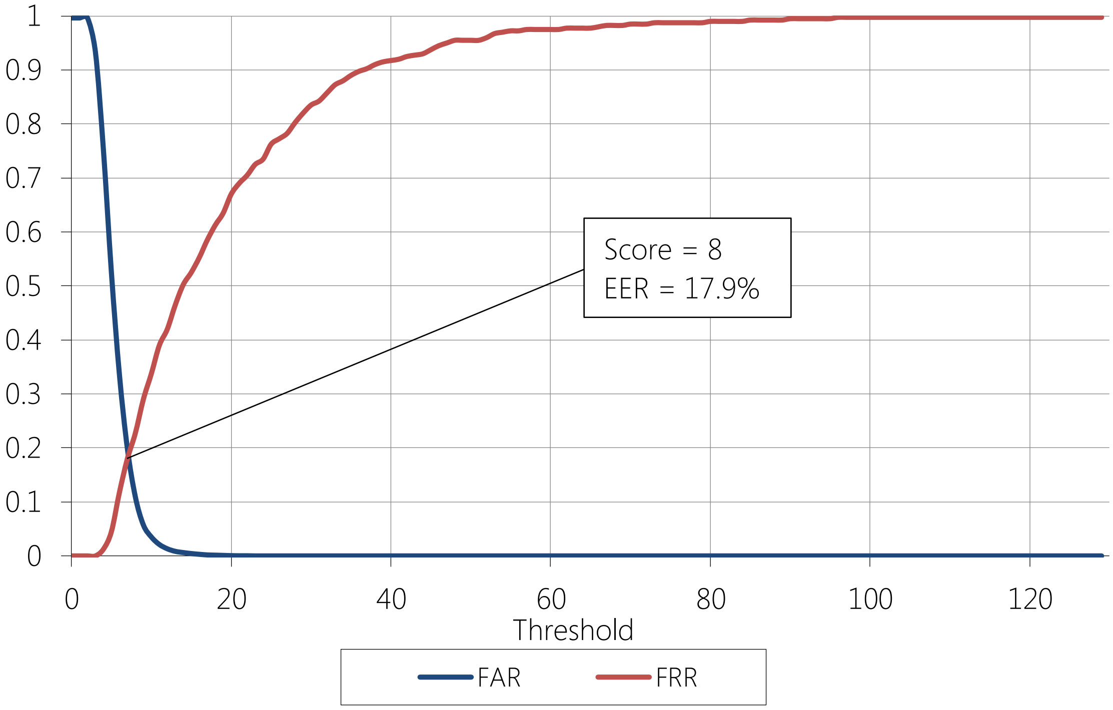


Figure 30. FAR/FRR curves of TS2 (re-sketched from [QSZ+2014])

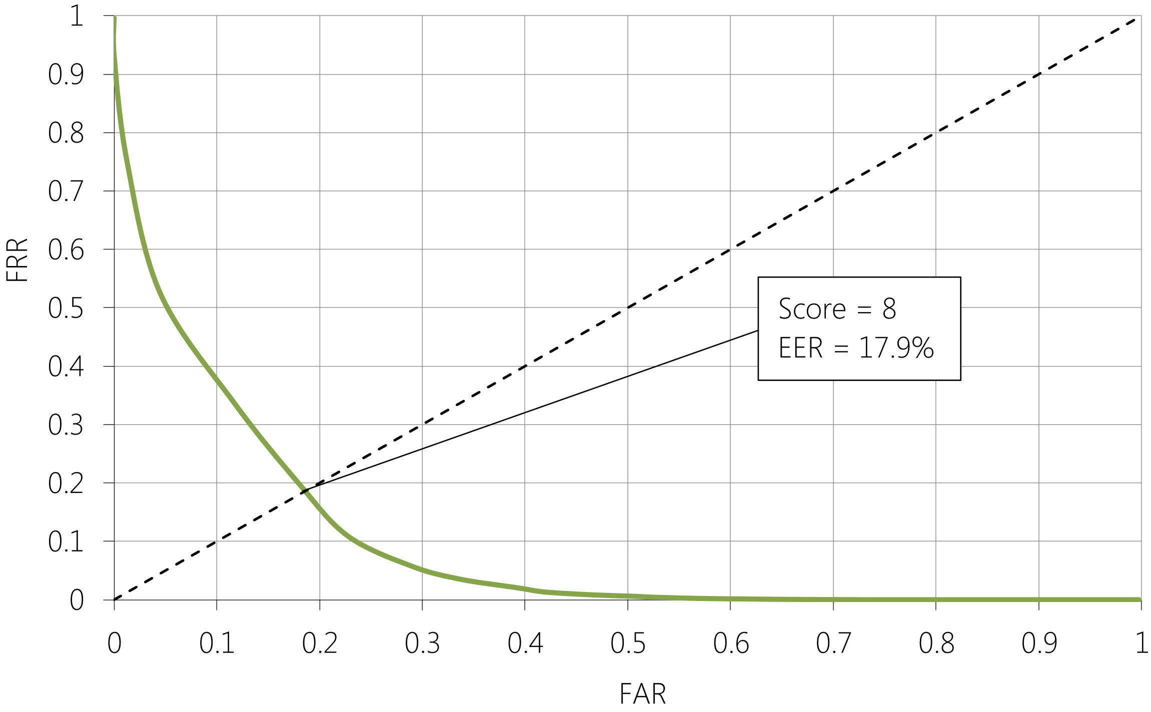


Figure 31. ROC curve of TS2 (re-sketched from [QSZ+2014])

The EER of 5.7% on TS1 shows the improvement in performance, comparing to the EER of 8.3% reported in [QSD2013], which applies an earlier version of relaxation labelling based separation algorithm without context awareness, tested with a smaller test set of 20 authentic samples. This justifies the advantages of applying data level context modelling into various aspects of the data processing in the separation approach. As the state-of-the-art literature (including [CFJ+2011] and [SFZ2011]) a) did not conduct imposter tests, b) applied a more sophisticated fingerprint matcher (VeriFinger 6.2 SDK), and c) did not specified that on what basis the TARs are calculated on the FAR range of 10^{-8} to 10^{-4} (see subsection 2.5.2), the evaluation results achieved in this chapter cannot be directly used for comparison, as the current statistical evaluation method bases on matching/non-matching cases would have needed a much larger test set to yield a similar FAR range.

Comparing to TS1, the evaluation results show that the performance of the approach evidently degrades on TS2, showing no obvious enhancement comparing to the results reported in [FSZ2012]. Besides the same three factors as explained before regarding [CFJ+2011] and [SFZ2011] making it difficult for the direct comparison of evaluation results, this could also be explained, as all the modifications made to the approach are derived from the context of TS1, and the context of TS2 differs already in the phase of acquisition, thus would require different corresponding adjustment on the algorithm.

Nevertheless, observing the results on both test sets, the samples with higher richness of detail and clarity (TS1) yield better separation results comparing to those with lower ones (TS2), implying the advantage of applying advanced sensing technology for fingerprint acquisition. As analysed in Table 13, the secondary acquisition CEs “overlapping behaviour” (e_{sc}^{A1}) and “fingerprint development technique” (e_{sc}^{A5}) are the most dominant factors with regard to such richness and clarity, they introduce the largest impact on the separation performance.

It should also be mentioned that the topic of separation of overlapped latent fingerprint is further addressed in a Master thesis that the author co-supervised (see [Zheng2014]). The thesis applies the same separation algorithm as proposed in this chapter and published in [QSZ+2014], and evaluates its implementation with an extended test set of overlapped latent fingerprints acquired using CWL sensor, as described in Table 16.

Substrate / *Resolution*	*500 ppi*	*1000 ppi*	*2540 ppi*
Hard disk platter	16	16	28
Smooth white paint-coated metal surface	-	-	8
Aluminium foil	18	18	24

Table 16. The extended test set of overlapped latent fingerprints generated in [Zheng2014] (re-sketched from [Zheng2014])

As shown in the table, the extended test set comprises overlapped latent fingerprint samples acquired using CWL sensor with three different resolutions (500 ppi, 1000 ppi, and 2540 ppi) on the same three types of substrates as used in this section. Similar statistical analysis is conducted respectively on the separation results of samples with the same resolution from the same type of substrate to the same two types of fingerprint templates as used also in this chapter. The evaluation results are shown in Table 17.

Template type	*Substrate*	*EER (500 ppi)*	*EER (1000 ppi)*	*EER (2540 ppi)*
TempO	Hard disk platter	8.4%	8.1%	1.8%
	Smooth white paint-coated metal surface	-	-	6.2%
	Aluminium foil	17.9%	15.3%	8.0%
TempBW	Hard disk platter	13.0%	14.7%	3.6%
	Smooth white paint-coated metal surface	-	-	6.2%
	Aluminium foil	25.6%	20.5%	9.2%

Table 17. The achieved EERs on the extended test set w.r.t. the two types of fingerprint templates reported in [Zheng2014], with best/worst performances highlighted (re-sketched from [Zheng2014])

Table 17 clarifies several factors that influence the separation: First, the acquisition resolution of the overlapped latent fingerprints shows a positive correlation with the separation performance. Second, the hard disk platter shows the highest cooperativeness, while the aluminium foil shows the lowest, presumably with regard the amount of noise the substrate introduces to the fingerprint images. Third, the TempO yields better separation performance than TempBW. Besides the results reported in Table 17, the test set in [Zheng2014] also includes the 100 conventionally acquired overlapped latent fingerprints from TsinghuaOLF, on which the testing yields EERs of 11.9% and 20.6%, respectively with TempO and TempBW.

To sum up, the evaluation reported in [Zheng2014] further verifies the applicability of the propose separation approach, especially on fingerprint samples acquired in various ways and with different resolutions.

5.3 ASSESSMENT REGARDING THE SECURITY RULES

At the beginning of this chapter, several exemplary SRs are derived as Example 8, describing the general working environment and requirements for latent fingerprint separation in the forensic scenario. Therefore, besides the objective evaluation of the performance of the developed separation approach described in last section, it is also necessary to subjectively apply a scenario-specific assessment on the approach with regard to those SRs to see if it addresses them well. It should be noted here that the following assessment is conducted by the author as a demonstration from the point of view of a developer of a forensic toolkit. Therefore, the assessment results will need to be verified through a joint work of both developer and expected end-user, in this case forensic experts. This falls out of the scope of this dissertation, yet it is worth mentioning that at the time of writing this chapter related work has already been planned in DigiDak research project

Based on the nature of the regulation that the SR represents, the eleven SRs listed in Example 8 can be divided into two classes: 1) the SRs that directly regulates what the separation approach shall achieve, including SR_3-SR_5, SR_7, and SR_9, 2) the SRs that regulates how the separation approach shall be used, including SR_0-SR_2, SR_6, SR_8, and SR_{10}. Therefore, for the first class, the developed approach is directly assessed to see if the requirements regulated by the SRs are either "sufficiently addressed", "partially addressed", or "not addressed". For the second class of SRs, the developed approach is assessed from the angle that if it is compatible or not with the ways of usage regulated by the SRs. The assessment results are given and further elucidated in Table 18.

SRs	*SR Class*	*Assessment results*	*Reasoning*
SR_0	2	Compatible	The execution of the separation can be easily accessed and overseen.
SR_1	2	Compatible	With the toolkit the fingerprint image can be easily imported and examined by the fingerprint expert.
SR_2	2	Compatible	With the toolkit the fingerprint expert can easily select proper processing method for the image.
SR_3	1	Sufficiently addressed	As established by the evaluation testing, the approach can be used on overlapped latent fingerprint samples developed in various ways with various resolutions.
SR_4	1	Partially addressed	While the global flow of friction ridges can be well preserved on the separated fingerprints, the evaluation conducted based on minutiae shows that the separation cannot always preserve all the minutiae information. However, it should be noted that in case the third level feature is shown on the original sample (like the ones acquired using CWL sensor) the developed approach has a good potential to preserve it as well.
SR_5	1	Partially addressed	Endeavour towards *Daubert* compliance is put into the development of the approach: it has been tested for its reliability and a series of publications, in which the approach is clearly elucidated, have been published and subject to peer review. However, its error rate is established on limited test sets, and it needs further improvement to form its corresponding standards, and eventually towards its general acceptance.
SR_6	2	Compatible	The user interface enables easy operation on the separation procedures and proper interference from the fingerprint experts.
SR_7	1	Sufficiently addressed	The fingerprint masks and parameters applied by the first fingerprint expert can be easily reused to reproduce the separation results for the verification by the second expert.
SR_8	2	Compatible	The separation results can be easily stored for further usage.
SR_9	1	Sufficiently addressed	The whole acquisition and processing flow does not employ any invasive procedure on the original latent

			fingerprint samples.
SR_{10}	2	Compatible	With the user interface, all the actions (e.g. the assignment of fingerprint masks, the appointment of parameters) can easily be accessed and documented for the completion of chain-of-custody information.

Table 18. Subjective assessment of the developed separation approach with regard to SRs in Example 8

As explained in Table 18, despite its performance limitation shown in by the evaluation, the developed approach manages to support all the derived SRs. To better compare this approach with the state-of-the-art ones introduced in subsection 2.5.2, the first class of SRs are also used to similarly assess these separation approaches, as the second class of SRs are not applicable here due to the lack of necessary information in the state-of-the-art literatures, especially regarding the implementation environments of these approaches. The assessment is summarised in Table 19.

Table 19 on page 95 clearly shows the advantage of the developed context-aware separation comparing to the state-of-the-art ones. Besides their common lack of consideration of *Daubert* criteria or any standard regarding the admissibility of their separation results to be accepted as court evidence (required by SR_5), a few additional remarks should be mentioned here regarding the assessment: The fact that the ICA based approach requires the source image refrains its application in the forensic scenario (reflected by SR_3, SR_4, and SR_7), yet it indeed applies invasive processing (required by SR_9). The infrared spectroscopic imaging and mass spectrometry based approaches both require the application of chemical agents as pre-processing, compromising the intactness of the original fingerprint sample (SR_9) and the reproducibility of the examination procedures (SR_7). The blind source separation approach applies non-invasive processing (SR_9), yet its testing reveals that its performance is severely limited regarding the evidence type (SR_3) and there is no mechanism to assure the reproducibility (SR_7). The previous relaxation labelling based approaches suit better to the forensic scenario comparing to the rest, yet still show limitations regarding to the evidence type that they are able to process (SR_3) and the reproducibility (SR_9). Furthermore, despite no perfect separation correctness has been reported (SR_4), it should be noted that the size of the test set is more convincing with the relaxing labelling based approaches other than the rest.

	SR_3	*SR_4*	*SR_5*	*SR_7*	*SR_9*
ICA based approach [STK2006]	NA	PA	NA	PA	SA
Infrared spectroscopic imaging based approach [BSF+2009]	PA	PA	NA	NA	NA
Mass spectrometry based approach [TLC+2010]	PA	PA	NA	NA	NA
Blind source separation approach [KGL+2011]	PA	PA	NA	PA	SA
Previous relaxation labelling based approaches [CFJ+2011] [SFJ2011] [ZJ2012] [FSZ2012]	PA	PA	NA	PA	SA
The developed context-aware approach	**SA**	**PA**	**PA**	**SA**	**SA**

Labels used in this table SA: sufficiently addressed, PA: partially addressed, NA: not addressed

Table 19. Subjective assessment of the state-of-the-art separation approach with regard to the first class SRs in Example 8; comparing to the context-aware approach developed in this chapter

5.4 SUMMARY

This chapter provides an instantiation of the theory of data level security-oriented context modelling proposed in Chapter 3. Within the scope of digitised forensics, it works in a "bottom-up" way, i.e. extracts and interprets useful context from the acquisition and processing environments, and eventually derives a context-based approach of separating non-invasively captured high-resolution overlapped latent fingerprints. The approach shows obvious improvement to previous approaches in both objective and subjective evaluations. Therefore, for the first time context awareness is introduced for the development and enhancement of a specific algorithm and its realisation.

Following the proposed conception (see subsection 3.2.3), this chapter specifies those defined context types:

- *Primary acquisition context* comprises the information describing the property of the acquired data, namely the overlapped latent fingerprints, and the information introduced to it by the acquisition environment.
- *Secondary acquisition context* is achieved by interpreting the primary one, resulting specific context entities along with the information that comes with them (see Table 13 on page 76).
- *Primary processing context* is derived by processing the secondary acquisition context using the chosen previous processing approach with no context awareness, which nevertheless potentially conforms with the processing environment regulated by SRs. In this case it is described in the form of separation error classes.
- *Secondary processing context* is achieved by interpreting the primary one. In this case, it comprises several parts: 1) the new features introduced to the algorithm to tackle the identified separation errors, 2) the parameters that come with the enhanced algorithm, 3) the intra- and inter-context of the parameters, leading to their optimised values.

This chapter clearly demonstrates the advantages gained by introducing security-oriented context modelling on the data level to derive specific data processing procedures, e.g. an algorithm and its implementation. The SRs with formalisation depicts clearly the expectations for the approach to achieve from the angle of application scenario, thus can be directly used for its scenario-specific evaluation. The development and interpretation of various types of context guides the design and realisation of the separation algorithm, and the enhancement that has been brought to it is also reflected by the performance improvement shown by the evaluation. More importantly, by applying the conception, all the influential factors that the data processing approach introduces to the processed data object, in this case potential fingerprint evidence, are clearly organised and described by these various context types on different levels. Therefore, this provides a further advantage that contributes to non-repudiation, which is an essential security aspect that expected for general forensic approaches.

6 MODELLING SECURITY WITH CONTEXT AWARENESS

A theoretical framework for security-oriented context modelling has been established in Chapter 3, and two instantiations based on this framework have been presented in Chapter 4 and 5, respectively on system and data level. However, as first raised in section 1.2, then pointed out again in section 2.6, further explained in section 3.3, and reflected in section 4.3 and 5.3, one issue remains unsolved: currently, there is a lack of proper systematic means for the assessment of such context models. Therefore, in this chapter, the first goal is to close this gap between context model generation and its evaluation. Section 6.1 derives and applies a series of applicable scenario-unspecific evaluation metrics, and also discusses the basic principles for the scenario-specific ones. After that, based on the previous established theoretical framework and its instantiations together with the evaluation metrics, section 6.2 induces a general methodology, which is meant for guiding the application of security-oriented context modelling in future security-related application scenarios.

6.1 EVALUATION METRICS FOR SECURITY-ORIENTED CONTEXT MODELS

This section addresses the lack of means for evaluation in the field of context modelling with regard to the two phases of security-oriented context modelling. First, after a context model is generated, the model itself needs to be assessed to estimate its general *quality*. Despite the assessment in this phase needs to refer to the expected application scenario of the model, the applied criteria must nevertheless be *scenario-unspecific*, i.e. the criteria shall stay consistent even if the nature of the application scenario changes. This is elucidated in subsection 6.1.1. Second, after the generated context model is applied to yield a data processing system or a series of data processing procedures, the *performance* of such application also needs to be assessed. In this case, the applied criteria shall reflect the actual requirements of the application scenario, hence be *scenario-specific*. This is explained in subsection 6.1.2.

6.1.1 Evaluation of the model's quality

What properties constitute a "good" context model? To answer this question, Bettini et al. summarise several criteria to evaluate context models applied in pervasive computing in [BBH+2010] (see subsection 2.2.3), namely *heterogeneity and mobility*, *relationship and dependency*, *timeliness*, *imperfection*, *reasoning*, *usability of modelling formalism*, *efficient context previsioning*. These criteria manage to cover several important properties expected for context models, but there are also limitations: first, they are derived specifically with regard to pervasive computing; second, they do not take consideration from the perspective of security.

Therefore, based on the criteria enumerated above, the evaluation metrics derived here need to meet following requirements:

- The metrics need to be used to assess the pros and cons in the properties of the context model itself.
- The metrics need to be applied on a more general level, i.e. not specifically only for pervasive computing or any other single application scenario, hence scenario-unspecific.
- The metrics must reveal the requirements from a security point of view.

- Regarding the different nature of the context models on system and data levels, the metrics need to be able to cover such difference and be applicable on both levels.
- The metrics shall also be able to reveal the dynamics of the context, as it might further develop over time and space.

As the first step towards the quality evaluation metrics of security-oriented context models, addressing the above five requirements, ten criteria are developed here as listed in Table 20.

Criteria in the evaluation metrics	*Potential security contribution*
Heterogeneity	
Relationships	
Imperfection	Integrity
Reasoning	Integrity
Provenance	Authenticity, non-repudiation
Evolvability	
Automatability	Confidentiality
Completeness	Integrity
Complexity and clarity	Non-repudiation
Trustiness	Authenticity, confidentiality, integrity, non-repudiation, availability

Table 20. Criteria defined here for evaluation metrics of security-oriented context models

These ten criteria are defined in the form of core questions and further elucidated as follows:

- *Heterogeneity – "How well can the context model process various types of context?"*: Similar to the definition in [BBH+2010], security-oriented context models must be able to handle various types of context entities collected from various sources, with various forms, evolving rates and qualities, and on various levels of granularity. *Mobility* defined in [BBH+2010] is no longer emphasised here in particular, as it is a feature that pervasive computing specifically requires and in fact already covered by the nature of heterogeneity. For example, on the system level, this could mean that the context model needs to synchronise various system components in different locations, while on the data level, it needs to be able to combine various data types.
- *Relationships – "How well can the context model represent the relationships in the context?"*: Similar to the definition in [BBH+2010], security-oriented context models must be able to interpret and express the relationships among various context entities. According to the theoretical framework proposed in this dissertation, *dependencies* defined in [BBH+2010] can be considered as one type of relationships (see subsection 3.1.2). For instance, a system level context model is required to assign various functionalities to various agents in the

system, while a data level context model needs to reflect various influential factors on the collection of data objects.

- *Imperfection – "How well can the context model tolerant with incomplete or erroneous information?"*: Inherited from [BBH+2010], it is necessary for security-oriented context models to be able to tolerate incomplete or even erroneous context information in the processing. For example, such information could be some object or concept that fall out of the required formalism on the system level, or data biased by noise on the data level. Therefore, the ability of handling such imperfection also contributes to *integrity* among the five security aspects (see subsection 2.3.1).

- *Reasoning – "How well can the context model apply reasoning to interpret and derive context entities?"*: Also inherited from [BBH+2010], one important benefit of using context modelling is its ability of reasoning, i.e. security-oriented context models shall be able to interpret context with reasoning techniques for necessary adaptations to be deployed and new context to be derived, and they shall also be able to verify the consistency of the context regarding its accuracy and completeness before and after the interpretation (hence constitute *integrity*). On the system level, this could mean that the model needs to deduct specific functionalities from more general ones, while on the data level it could mean that specific operations can be derived from certain features of data objects.

- *Provenance – "How available is the provenance information of the context entities in the context model?"*: Security-oriented context model shall preserve the provenance information of context entities, so the previous status of entities can be traceable. This contributes directly to *authenticity* and to *non-repudiation*, if the well preserved records of the previous status can be subject to auditing. A system level example would be, in case that a combination is deployed on two system components to form a new one, the status of the original two together with their functionalities must be documented in advance. On the data level, the source and all following processing procedures applied on a data object must go to its meta-data and the audit trail.

- *Evolvability – "How well can the context model absorb new context and evolve accordingly?"*: Within the scope of context modelling, it is expected that the context entities might evolve through time and space. Correspondingly, it is also expected that security-oriented context models are able to evolve themselves to better process its changing context. On the system level, when the security requirements change due to stricter protocols, new functionalities need to be realised accordingly by including new system components. On the data level, when some old data format becomes obsolete and gets replaced by a new one, the operations designed to process it also need to be adjusted accordingly.

- *Automatability – "To what degree is the context model able to function automatically / how much human intervention does it need?"*: This assesses the ability of security-oriented context models to execute their functionalities automatically, reflecting the level of required human intervention. The involvement of human agents is considered as a major dynamic parameter, as human behaviour is usually considered as hard to predict and it could yield potential security leaks. Therefore, this criterion also contributes to *confidentiality* and *integrity*. A system level model could be designed to derive specific policies from its functionalities, yet it might need human administrators to specify the protocols beforehand. A data level model could be designed to process data objects automatically, yet it might need a specialist to adjust its parameters.

- *Completeness – "How complete is the context model regarding the coverage of its representation of what is modelled?"*: As the progress of modelling is fundamentally speaking a progress to selectively represent the physical reality in a systematically formalised way, this criterion is set to assess security-oriented context models for their completeness of their coverage of the physical reality being modelled, i.e. if there is relevant information failed to be contextualised. Therefore, this criterion also contributes to the *integrity* from the security perspective of view. On the system level, if a security-oriented context model extracts its global security context from use-cases, those use-cases shall cover all aspects of the expected usage of resulting system. On the data level, the completeness of security-oriented context models could mean that they shall cover all the environmental factors that related to the acquisition of data objects to be processed.
- *Complexity and clarity – "How complex is the representation of the context in the context model? And how clear is it to the user? "*: Both of them contribute to the usability of the security-oriented context models. *Complexity* refers to the amount of context entities and the degree of interactions between them, while *clarity* refers to the ability of the models to represent, express and manage the entities and their interactions. From the security point of view, these can eventually contribute to *non-repudiation*. A system level example would be that the model shall be able to handle large amount of security policies using clear hierarchy and formalism, while on the data level this could mean that the operations which are deployed on the data objects shall be well clarified.
- *Trustiness – "How well is the context model designed to meet the security-related requirement from the application scenario?"*: The overall security performance of security-oriented context models is assessed by this criterion, which requires that the models come with security-compliant design and operate only with context entities that are trustable, i.e. the context entities shall be subject to necessary security requirements in various aspects (*authenticity*, *confidentiality*, *integrity*, *non-repudiation*, and *availability*). On the system level, this means that some system components must be able to realise the functionality of check and enforce corresponding security enhancement measures, e.g. central audit service for *non-repudiation*. On the data level the data processing procedures must be derived with security compliance, e.g. meta-data update for *authenticity*.

The universality of the above criteria determines that they are applicable for subjective assessment of any context model. With these criteria, the next step towards the metrics is the development of the evaluation scale. In this section, the following three-level evaluation scale is proposed:

- *Sufficiently addressed* is assigned, if the context model being evaluated meets all the important aspects of the requirement regarding a certain criterion in the evaluation metrics;
- *Partially addressed* is assigned, if the context model being evaluated meets some aspects of the requirement regarding a certain criterion, yet missing at least one important one;
- *Not addressed* is assigned, if the context model being evaluated is not able to address any of the aspects of the requirement regarding a certain criterion at all.

To be able to conduct a subjective quality assessment using the metrics, a series of regulations are developed to clarify that how the evaluation scale should be utilised, with regard to each criterion in specific circumstances. These regulations are described in the Table 21.

	Sufficiently addressed	***Partially addressed***	***Not addressed***
Heterogeneity	The context model is designed to process a diversity of context (e.g. data objects with various types, appearances, forms, formats, etc.), as expected in the application scenario.	The context model is designed to process context with a certain level of diversity (e.g. certain type of data object with various formats), yet not completely covering the expectation of the application scenario.	The context model does not consider the need of processing the diversity of the context from the application scenario at all. Typically, the model is designed to process only fixed types of them.
Relationships	The context model is designed with effective means of expression to represent all kinds relationships expected to appear in the application scenario in the context.	The context model is designed with adequate means of expression to represent parts of the relationships expected to appear in the application scenario in the context.	The context model is lacking of means of expression to represent the relationships expected to appear in the application scenario in the context.
Imperfection	The context model is designed with effective mechanisms to handle all kinds of flawed context information during the processing expected to appear in the application scenario.	The context model is designed with adequate mechanisms to handle certain kinds of flawed context information during the processing expected to appear in the application scenario.	The context model possesses no mechanism to handle flawed context information during the processing expected to appear in the application scenario at all.
Reasoning	The context model is designed with effective reasoning mechanisms/techniques to derive/interpret/organise all kinds of context collected from the application scenario, and to maintain their consistency during the processing.	The context model is designed with adequate reasoning mechanism/technique to derive/interpret/organise certain, but not all kinds of context collected from the application scenario, and optionally to maintain their consistency during the process.	The context model is lacking proper means to derive/interpret/organise context collected from the application scenario.
Provenance	The context model is designed with effective mechanism to preserve the provenance information of the context and reuse it as expected	The context model is designed with limited mechanism to preserve the provenance information of the context, yet not	The context model is lacking mechanism to preserve the provenance information of the context or reuse it as expected in

	in the application scenario.	necessarily support its reuse, as expected in the application scenario.	the application scenario.
Evolvability	The context model is designed with effective adaptability for evolving context, within the changing range expected in the application scenario.	The context model is designed with adequate adaptability for evolving context, yet not completely covering the changing range expected in the application scenario.	The context model possesses no adaptability for evolving context, albeit the expectation of the application scenario.
Automatability	The context model is designed with proper balance between automatic processing and human intervention as expected in the application scenario.	The design of the context model considers the issue, yet it introduces either too much or too human intervention, comparing to the expectation in the application scenario.	The design of the context model does not consider this issue at all, despite that human intervention is expected in the application scenario.
Completeness	The context model is designed to thoroughly cover all context information on each level of granularities from the application scenario.	The context model is designed to cover most context information on all the important levels of granularities from the application scenario.	The design of the context model fails to contextualise important information from the application scenario.
Complexity & clarity	The context model is designed with effective mechanisms to represent large amount of context information and sophisticated relationships, while its infrastructure is clear and understandable enough, both as expected in the application scenario.	The context model is designed with adequate mechanisms to represent certain amount of context information and sophisticated relationships, yet its infrastructure is not as clear as expected in the application scenario.	The context model is lacking of mechanism to represent large amount of context information and sophisticated relationships, and/or its infrastructure is not clear at all, according to the expectations of the application scenario.
Trustiness	The context model is designed with mechanisms to assure all aspects of the security requirements from the application scenario.	The context model is designed with mechanisms to assure some of the aspects of the security requirements from the application scenario.	The context model is lacking of mechanisms to assure security requirements from the application scenario.

Table 21. Regulations on the application of quality evaluation metrics of security-oriented context models

As shown in Table 21, although the criteria in the evaluation metrics are not specific regarding to the nature of the application scenario, the application of such metrics nevertheless needs to refer to its expectations. To demonstrate how to apply the metrics to assess the quality of context models, they are further utilised on the two context models described in Chapter 4 and 5.

Table 22 on page 104 shows the evaluation results for the system level security-oriented context model derived in the instantiation in Chapter 4. The context model uses a hierarchical structure to process different types of context entities that come on different levels of granularities, thus *heterogeneity* is considered as sufficiently addressed. As the form of security policies is used in the hierarchical structure, enabling the convenient expression of relationships between different context entities, the criterion regarding *relationships* is also sufficiently met. Within the scope of the context model, there is no strictly enforced formalism for the context entities, therefore the lack of information can be tolerated to a certain degree. Additionally, the existence of conflict handling mechanisms is able to resolve the conflicts triggered by the potential erroneous information. Therefore, it is also considered that *imperfection* is sufficiently addressed. As the carrier of context information, security policies are derived from higher level to lower level. This progress of interpretation of context information contributes to the ability of *reasoning* of the context model. The assurance of *provenance* is realised by the derivation of security policies for certain specialised functions in the context model, e.g. "provenance info copy and enrichment" in the *ingest* functional entity, "provenance info maintenance" in the *archival storage* functional entity, as well as "provenance info update" in the *access* functional entity (see subsection 4.1.1). The sufficient addressing of *evolvability* is reflected by the ability of the context model to handle evolving context entities, and this is the task of the *control* block of the developed security framework (see subsection 4.1.4): when the context changes, new security policies will be derived accordingly to replace obsolete ones. *Automatability* is not within the main focus of the context model. Despite the processing of archived data objects is designed to be automatic, human agents are still heavily involved in various roles, e.g. administrators, auditors, or content managers. Despite the fact that they are needed, there is not further investigation conducted on e.g. the potential security risk brought by such involvement. Therefore, this criterion is only partially addressed. As the whole context model is developed to cover the raw context in the form of use-cases and possess the ability to evolve in case of newly emerged context, it is assumed that the *completeness* of coverage is sufficiently addressed. With the large amount of security policies representing complicated relationships both within and in between the functional entities, the *complexity* of the context model is considered as high. However, as the boundaries between different hierarchies and the reasoning for the derivation of lower level security policies from higher level ones is well formulated, the *clarity* of the model is also high. Therefore, the criterion of *complexity and clarity* is sufficiently addressed, yielding satisfactory usability. The last criterion *trustiness* is also sufficiently addressed, as security is the most essential focus of the development of the context model in the first place: the security requirements in all five aspects are considered at first during the context extraction from the use-cases, then in the derivation of security relevant functional entities and their functions, subsequently throughout the development of security policies on different levels, and at last the construction of the final security framework.

Criteria	***Assessments***
Heterogeneity	Sufficiently addressed
Relationships	Sufficiently addressed
Imperfection	Sufficiently addressed

Reasoning	Sufficiently addressed
Provenance	Sufficiently addressed
Evolvability	Sufficiently addressed
Automatability	Partially addressed
Completeness	Sufficiently addressed
Complexity and clarity	Sufficiently addressed
Trustiness	Sufficiently addressed

Table 22. The application of the evaluation metrics on the system-level security-oriented context model for digital long-term preservation described in Chapter 4

Table 23 on page 105 summarises the evaluation results for the data level security-oriented context model derived in the instantiation in Chapter 5. The context model is designed to process various types of context entities that emerge in different phases in the data processing (see the summaries in section 5.4). However, as the processed data object is restricted to certain format (i.e. the fingerprint image acquired with CWL sensor), its ability of handling the *heterogeneity* of the data objects is to some degree compromised. Within the scope of the context model, the *relationships* between various context entities are fully analysed and eventually reflected by the developed processing procedures (see section 5.1), thus this criterion is sufficiently addressed. The criterion of *imperfection* is partially addressed, as the context model has considered the context that could influence the quality of the data object and handled some of these influential factors, yet the overlapping behaviour of the processed fingerprint sample still exists as the main cause of separation failure. The criterion of *reasoning* is sufficiently reflected by the progress of context derivation, i.e. secondary context from primary context, and processing context from acquisition context, and also by how the specific operations are derived from the interpretation of the derived context. *Provenance* is partially addressed by the context model in two aspects: a) It is one of the primary objectives of the forensic approach yielded by context model, as the successful separation results are expected to be subject to follow-up authentication/identification procedures to confirm/identify the source of the fingerprints. b) On a finer level of granularity, each single processing step and its input and output are clarified (see Table 14 on page 55) together with the context histories, so the whole separation process is deterministic and described with enough details to be reproduced on the same type of data objects. However, the current approach saves no copies of them as back-ups and applies no audit trail on the processing procedures. The *evolvability* of the context model is addressed by the self-adjustability it contributes to the resulting separation approach, e.g. the self-adjustable parameter setting according to the context of input data (see subsection 5.1.2). However, as all the context entities collected and analysed is referred to only fingerprint samples acquired by CWL sensor, the approach shows a limited *evolvability* to other data sources. Regarding *automatability*, the context model is designed for an automatic separation approach with an appropriate degree of human intervention from forensic specialists (e.g. the assignment of fingerprint masks, adjustment of the parameters, etc.), thus this criterion is partially addressed. The *completeness* of the context model is addressed by the contextualisation of both the acquisition and processing environment (see subsection 5.1.1), which is considered as sufficient here, as it covers not only the context directly relevant in the discussed scenario, but also alternatives in the state of the art. Consisting of four types of context entities with various forms of expression, the *complexity* of the context model is considered as high. However, as the reasoning from one type of context to another is clearly explained and demonstrated, the *clarity* of the model is also considered as high. Therefore, it can be assumed that the context model achieves high usability. The *trustiness* of a forensic approach can be generally assessed regarding to its *Daubert* compliance: comparing to the state-of-the-art approaches, the introduction of context awareness makes its contribution to the admissibility of the approach. Regarding the required *non-repudiation*, the context model provides a systematic method to reveal explicitly all the influential factors on the processed potential evidence piece from both the acquisition and the follow-up processing, and this at the same time assures *integrity*. However, the investigation regarding the known error rate is still to conduct, due to the relatively small sample place. Therefore, the criterion of *trustiness* is considered as partially addressed.

Criteria	*Assessments*
Heterogeneity	Partially addressed
Relationships	Sufficiently addressed
Imperfection	Partially addressed
Reasoning	Sufficiently addressed
Provenance	Partially addressed
Evolvability	Partially addressed
Automatability	Partially addressed
Completeness	Sufficiently addressed
Complexity and clarity	Sufficiently addressed
Trustiness	Partially addressed

Table 23. The application of the evaluation metrics on the data-level security-oriented context model for digital dactyloscopy described in Chapter 5

6.1.2 Evaluation of the model's performance

With the evaluation metrics derived and demonstrated in subsection 6.1.1 and 6.1.2, it can be established that if a context model should be considered as a "good" one. However, without further evaluation which takes into consideration the actual situation of the application scenario, it would be preconceived to conclude that the "good" context model would at the same time be a "well-performing" one. Therefore, it is also necessary to assess the performance of the context model, i.e. how well the application of the context model functions in its application scenario, so the achieved improvement by introducing the context awareness can be evaluated according to the specific requirements of the scenario. Therefore, unlike the evaluation depicted in the last two sections, the one discussed here focuses on the outcome of the application of context modelling, i.e. data processing system/proceduress, of which the corresponding evaluation methods have been in fact widely developed and applied, as summarised in the state of the art (see section 2.4.2 and 2.5.2). Similar evaluation has also been conducted on the instantiations derived in previous chapters: in section 4.3, the assessment of the performance of the context model for secure long-term preservation is conducted subjectively with regard to a series criteria assembled together working as a checklist, whereas in section 5.2 and 5.3 objective and subjective evaluations are respectively performed to reveal the performance of the context model for a overlapped latent fingerprint separation approach with regard to the error rates in its processing results and it security coverage. As reflected by these two cases, the nature of the application scenarios of security-oriented context modelling varies dramatically, so the suitable assessing means should also vary accordingly.

Therefore, instead of presenting universally applicable general metrics (which in this case do not exist), this section summarises some basic principles for the development of metrics to evaluate the performance of security-oriented context models:

- The metrics must be **scenario-specific**, i.e. the metrics need to be derived based on the specific requirements from the application scenario of the context model. For example, in section 4.3, the selected part of TRAC that used as the evaluation metrics reflects the expectations from the scenario of digital long-term preservation.
- The metrics must be able to reveal the **security compliance**. Different from the general criteria of trustiness derived in section 6.1.1, the scenario-specific metrics should focus on if the particular data object is processed to meet its expected security requirement in the scenario. For example, section 5.2 conducts an evaluation that assesses the integrity of the separation results with regard to its expected quality as potential evidence.
- The metrics must come with **understandable expression**, especially for those who do not develop them. It is reasonable to assume that in most cases the developer of the evaluation metrics is not the one who applies them. Therefore, the metrics need to be presented in a way that can be relatively easy to understand and utilise. This ensures not only the correctness of the evaluation, but also its reproducibility by third parties if necessary. For example, the subjective assessment in section 4.3 employs the metrics consisting of evaluation criteria expressed in natural language, which eases the understanding and application of the criteria.
- Besides subjective assessment, the metrics should also consider employing **objective indicators**, if applicable. Objective indicator has its advantage as it shows the benefit of introducing context modelling in a more quantifiable way. For example, section 5.2 applies EER as such indicator to show the improvement brought by context modelling.

The following table summarises the performance assessments conducted on the two context models derived in Chapter 4 and 5 (see section 4.3, 5.2, and 5.3): they are respectively analysed with regard to the above four principles, clarifying if and how the principles are addressed.

Principles	**Assessment on the context model for digital long-term preservation**		**Assessment on the context model for separating overlapped latent fingerprints**	
	Principle addressed or not?	*Reasoning*	*Principle addressed or not?*	*Reasoning*
Scenario-specific	Addressed	The applied assessment framework represents the expectations of the application scenario of digital long-term archiving.	Addressed	The subjective assessment is based on the security rules that describe the specific forensic scenario, and the objective assessment is based on the investigation of error rates, which directly reflect the reliability

				of forensic approaches.
Security compliance	Addressed	The selected and applied assessing criteria from the assessment framework all reflect various security aspects.	Addressed	Both subjective and objective assessments reflect the requirements regarding various security aspects from the application scenario (e.g. *Daubert* criteria).
Understandable expression	Addressed	The assessing criteria are described in natural language.	Addressed	The security rules that the subjective assessment is based on are described in natural language, while the methods of applying the objective assessment is also explicitly described (see subsection 5.2.1 and 5.2.2)
Objective indicators	Not addressed	Only subjective assessment is applied at this stage.	Addressed	Error rate is used as the objective indicator in the assessment.

Table 24 Summary of the performance assessments of the two context models developed in this dissertation regarding the four principles

To sum up, the development of the scenario-specific evaluation metrics is of little difference with those already existing in academia and industry for data processing systems or procedures. Nevertheless, the above summarised principles shall be fulfilled, to assess what is brought to the scenario by the introduction of the context awareness.

6.2 A GENERALISED MODELLING METHODOLOGY TOWARDS CONTEXT AWARENESS

Till now, this dissertation has developed a theoretical framework of security-oriented context modelling, applied the framework in two selected application scenarios that cover respectively the system and data level, and discussed the evaluation of the quality as well as the performance of the resulting context models. However, it should be noted that despite that it has already been mentioned that context modelling tends to be an iterative rather than one-time process due to the fact that the context evolves, this is not specifically addressed by the previous chapters, as they rather focus on addressing the context from a relatively static status of a potentially dynamic scenario. In other words, the focus has been how one *version* of a context model should be developed, regarding the current status of the context from the application scenario, despite that the model possesses the potential to evolve to future *versions* (here *version* as part of the descriptive scheme presented in section 3.3). Therefore, in this section, a general methodology is further developed for the security-oriented context modelling, summarising the previous chapters and emphasising the iterative nature of the modelling process.

Extending the descriptive scheme of security-oriented context model introduced in section 3.3, the general methodology is proposed here as Figure 32 on page 108 illustrates. As the figure shows, the *identifier* links the prospective application scenario to the model, the *version* indicates the status of the model in its iteration of modelling process, the two types of evaluation metrics both reflect the application scenario in different ways as depicted in last section, and the *security-oriented context model* itself is the output of modelling process either on system or data models.

The security-oriented context modelling starts with *context collection*, where relevant information is collected from the application scenario. On the system level, such information addresses the question "what are expected for the data processing system to be developed to achieve?" As for the data level, it focuses on the question "what are expected to happen on the data during its acquisition and processing?" The thorough coverage and correctness of the collected information regarding the requirements from the application scenario, especially the security-related ones, is vital for the context model to be developed. As the context collected in this step usually contains a large amount of information, a follow-up step is *context interpretation*, which aims to the categorisation and formalisation of the context. On the system level, this is achieved by the gradual derivation of security policies with various levels of granularities corresponding to their roles in a policy hierarchy, so the context entities (including. corresponding system component, human agent, actions, etc.) are clarified and formalised, as described in section 4.1. On the data level, this is conducted first in the acquisition environment by specifying its impact on the data object (i.e. deriving secondary acquisition context from the primary one), and then in the processing environment by analysing the introduced impact in this environment, developing mechanisms to address it, and forming specific data processing operations (i.e. deriving secondary processing context from the primary one), as described in section 5.1.

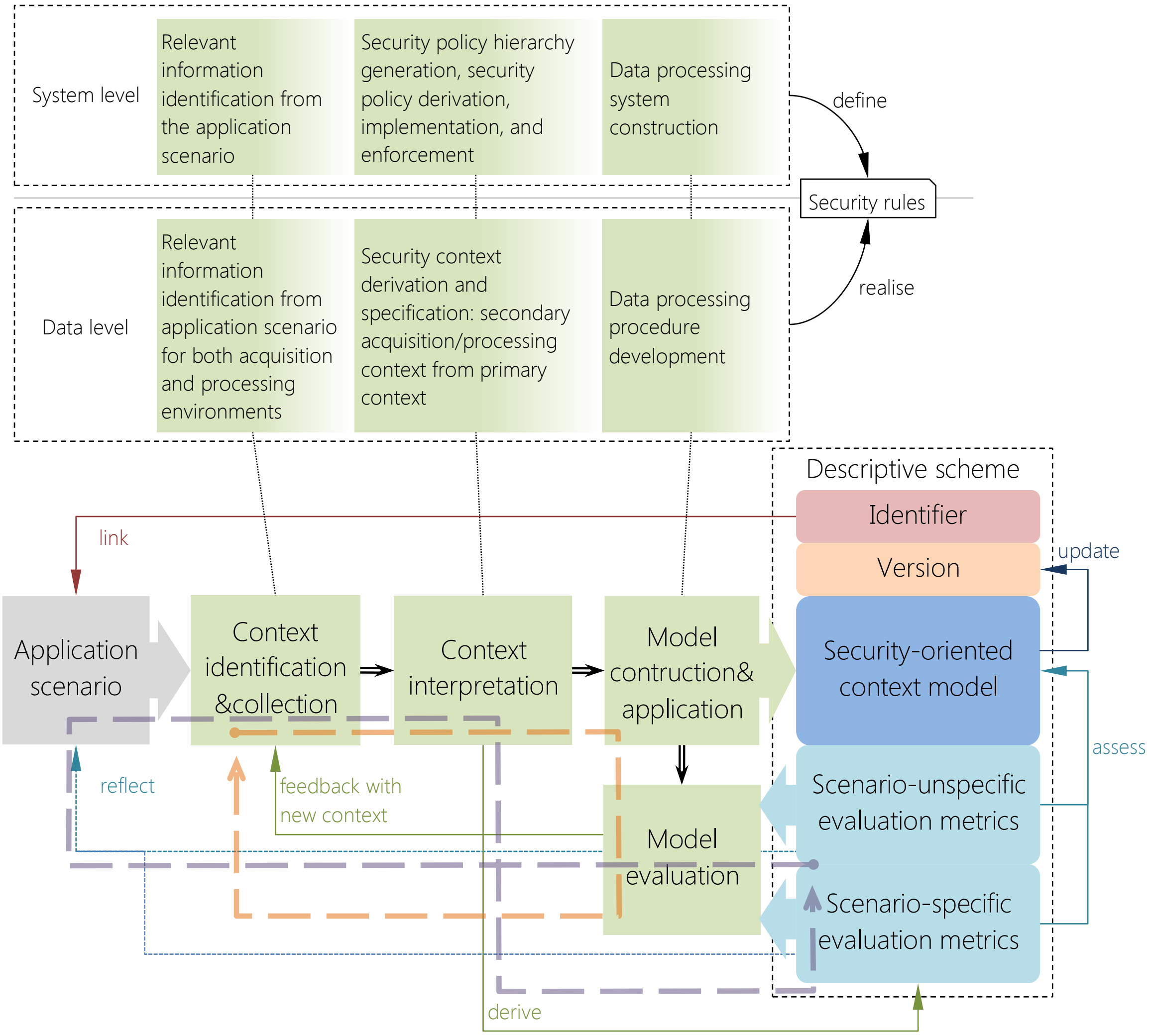

System level
Relevant information identification from the application scenario
Security policy hierarchy generation, security policy derivation, implementation, and enforcement
Data processing system construction
define
Security rules
realise
Data level
Relevant information identification from application scenario for both acquisition and processing environments
Security context derivation and specification: secondary acquisition/processing context from primary context
Data processing procedure development
Descriptive scheme
Identifier
Version
update
link
Application scenario
Context identification &collection
Context interpretation
Model contruction& application
Security-oriented context model
assess
Scenario-unspecific evaluation metrics
Scenario-specific evaluation metrics
reflect
feedback with new context
Model evaluation
derive

Figure 32. General methodology to introduce context awareness into modelling process

Afterwards, the clarified security context can directly leads to *model construction and application*, resulting in context models being applied in the designated application scenario. On the system level, the constructed model yields a data processing system that enforces its security rules (see section 4.1 and 4.2), while on the data level a series of data processing procedures that meet expectations, typically described by security rules (see section 5.1). Therefore, the security rules connect the system and data levels by serving as different roles: on the system level, it is one of the outcomes of the modelling process, while on the data level it contributes to the description of processing environment. After a security-oriented context model is constructed, it enters *model evaluation*, where it needs to be assessed in two aspects: its quality needs to be assessed using the universally applicable *scenario-unspecific metrics* (see subsections 6.1.1 and 6.1.2), and its performance using *scenario-specific metrics* (subsection 6.1.3), which should be derived based on the context from the application scenario (see section 4.3 and 5.3). The result of the assessment is then fed back to the start of the modelling process as new context, leading to necessary modification on the context model.

The iterative nature of the proposed security-oriented context modelling process is clearly illustrated in Figure 32 (see the two bold dashed arrows respectively in purple and orange). In case of an evolving application scenario, the corresponding security-oriented context modelling consists of two iterative processes, instead of being static: First, as the purple dashed arrow shows, when the application scenario changes, the evaluation on the context model correspondingly changes accordingly, i.e. this can lead to the necessary change of the scenario-specific metrics themselves as well as the assessment results from them, and despite that scenario-unspecific metrics stay the same, their assessment results can also vary. Second, as the orange one shows, by regularly applying the evolving evaluation metrics, the current version of the context model is assessed for its quality and performance, and new context is meanwhile generated according to the application scenario, so a new version of the context model can therefore be correspondingly developed. Figure 33 further illustrates how these two iterative processes interplay over time and space.

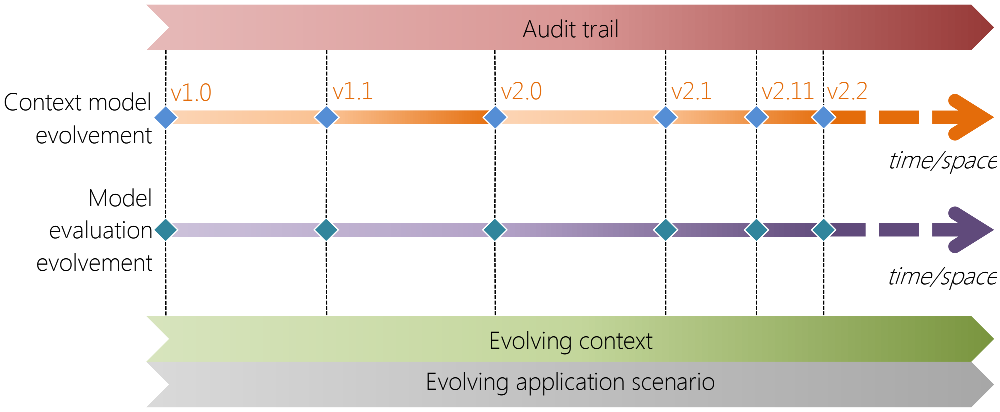


Figure 33. The two iterative processes within security-oriented context modelling

As shown in the figure, the evolvement of the context model progresses simultaneously with its evaluation, as the context contributed by the application scenario evolves over time and space. The application of the evaluation can either be employed routinely as means of regular quality/performance check (e.g. from the model version v1.0 to v1.1 then to v2.0 in the figure), or triggered by certain incident, typically an unexpected/unplanned change from the application scenario that requires a new version of the model (e.g. the appearance of model version v2.11). Between different versions of the context model, the degree of evolvement can either indicate a minor update (e.g. from model version v1.0 to v1.1) or a major one, which can even be regarded as the replacement of an obsolete model with a completely new one (e.g. from model version v1.1 to v2.0), if necessary. It should be noted here the “x.y” or “x.yz” forms or model version serve only as demonstrations, as in practice its sophistication should be directly related to how often it is expected to be updated, which further depends on the nature of the application scenario. Some application scenario can be relatively static, e.g. forensic application scenario, as its security protocols and requirements tend to change only over months or even years, during which they remain consistent in large geographical areas (e.g. always within the same country). In this case, the resulting context model can also maintain the same status for relatively longer period of time, so the form of its version can also be relatively simple. Some other application scenario can however be relatively dynamic, e.g. digital long-term archiving, as new requirements can emerge over days, if not hours, for e.g. archiving new types of data, including new roles of human agents, etc., and they can also be different geographically. In this case, the update of the context model can be expected to be quite frequent, and its form of version is expected to be sophisticated enough to comprise time and space information. At last, as also shown in the figure, for the sake of non-repudiation over time and space, an audit trail also needs to be constructed accompanying the evolvement of the context model and its evaluation, so the specific changes made during both processes can be documented and revisited if needed.

7 SUMMARY, CONCLUSIONS, ONGOING AND FUTURE WORK

In this chapter, section 7.1 summarises the whole dissertation and give conclusions, section 7.2 suggests several research paths towards the future work.

7.1 SUMMARY AND CONCLUSIONS

Motivated by the demand for the solution to the security compromise caused by applying modelling with little context awareness, this dissertation aims at closing the identified research gaps in the current study of context modelling: the state-of-the-art theories are scattered in their foci thus hard to compare or reproduce, the applications are limited in certain scenarios, and there is little focus on security. It proposes a theoretical framework of security-oriented context modelling with a series of formalised concepts, applies the framework on both system and data levels in the selected application scenarios of digital long-term preservation and digital dactyloscopy, regulates the evaluation of both quality and performance of security-oriented context models, and eventually induces a general methodology which describes the security-oriented context modelling process and identifies its nature of being iterative.

Regarding the research challenges raised at the beginning of this dissertation in section 1.1, this dissertation addresses them in the following ways:

1) *"How to collect and organise useful information that constitutes security related context?"* The nature of the application scenario of the resulting model decides what and how information from the scenario should be handled in contribution to the context, following the proposed conception of security-oriented context modelling. In case of a system level scenario where a data processing system is expected, context information describes the security-related expectations for the system (e.g. in the form of use-cases), which need to be formalised and interpreted to reveal system infrastructure and functionalities. In case of a data level scenario where a series of data processing procedures is expected, context information describe the influential factors on the data objects from both the acquisition and processing environments, and needs to be interpreted to derive a series of corresponding processing steps to meet the security expectations.

2) *"What is the general context modelling methodology that can be used to address various security issues in sophisticated data processing systems?"* On both system and data levels, the security-oriented context modelling process consists of steps including context identification and collection, context interpretation, model construction and application, and model evaluation. More importantly, such modelling process is designed to be iterative, addressing the evolving security context from the changing application scenario, which further results in an evolving evaluation progress.

3) *"How to evaluate the context models, especially from the security point of view?"* The context models should be evaluated with regard to their qualities and performances using different metrics. As for the quality, subjective assessment should be conducted using the

presented scenario-unspecific criteria, while for the performance scenario-specific metrics should be designed for subjective and/or objective assessments.

Therefore, summarising the contributions that this dissertation has made, the following conclusions can be drawn:

1) **Context modelling is able to serve as an effective solution for the security assurance in various application scenarios.** A sound theory has been formed in this dissertation. It conforms with the state-of-the-art theories on context modelling, yet for the first time it takes security requirements into systematic consideration and gives formalised descriptions of all concepts related to security context, providing clear guidance and regulation for its application. Furthermore, the correctness and applicability have been established by the two instantiations in two different application scenarios, where context models are developed, yielding a security framework for long-term archiving system and a separation approach of overlapped latent fingerprints for forensic dactyloscopy, both have gained improvement regarding their security assurance, in contrast to their state-of-the-art achievements.

2) **The competence of a context model should be estimated by assessing it using both scenario-unspecific and scenario-specific evaluation metrics.** As mentioned previously, the competence of a context model is reflected by its quality, i.e. how well its design is able to address the scenario-unspecific criteria, and its performance, i.e. how well the data processing system/procedures that it yields perform regarding the expectations from the application scenario.

3) **A generalised modelling process can be followed for the introduction of context awareness.** As a first answer to the general problem identified at the beginning of this dissertation, the presented generalised modelling methodology summaries the modelling steps, which can be further specified according to the nature of the application scenario.

4) **In practice, the security-oriented context modelling is an iterative process.** With evolving application scenarios, the security-oriented context modelling is an iterative process, as not only the context model evolves with the changing context from the application scenario, it evaluation also evolves correspondingly.

5) **The proposed security framework poses a potential solution for its application scenario of digital long-term archiving.** Comparing to stat-of-the-art achievements which lack effective solution for security issues, it provides a prototype based on which a secure digital long-term archive can be constructed, to handle large amounts of data with more reasonable storage solution, effectively manage and utilise large number of security policies for security requirements, and possess proper mechanism to derive and process evolving context. Therefore, it can be used to avoid similar incidents like NASA's "digital dark age" from happening, and also to serve as the foundation of a digital evidence archive, which would have solved the problems that DEA had to face (see Chapter 1).

6) **The proposed overlapped latent fingerprint separation approach achieves improvements in various aspects regarding its application scenario of digital dactyloscopy.** It takes into consideration the security requirements from the forensic scenario by introducing context awareness into the entire progress of its development. The reported error rate shows an obvious improvement on the high-resolution samples acquired by CWL sensor, and its applicability is also further extended to the conventionally acquired samples.

7.2 FUTURE WORK

There is only so much that can be addressed in one dissertation. Therefore, this section discusses the future work that should be considered conducting.

Regarding the modelling theory introducing context-awareness for security as well as its application in general, the following aspects can be explored in the future:

- In the current modelling theory, the security context is expressed to a large extend by using natural language, which has its advantage regarding the understandability of the resulting model. However, it can at the same time cause complications, if the model is expected to employ automatic mechanism to process the context. Therefore, it is necessary to explore a further formalised expression for the security context, to ease such burden. This could be essential in case of large amount of context.
- The applicability of the presented modelling theory can be further justified. In this dissertation, the system and data level modelling is conducted respectively in two different application scenarios, mainly due to the time limit of the SHAMAN project that restrains the modelling work from proceeding further into the data level. Therefore, it is necessary to seek for the opportunity to apply the presented modelling theory on both system and data levels in one application scenario, resulting in a complete data processing system with implementations of its functionalities with data procedures. As a matter of fact, such work has already been planned and is currently undergoing within the scope of research project DigiDak+[1], where a forensic processing infrastructure is planned to be developed to manage and realise the de-personalisation of the biometric data being processed, so the confidentiality/privacy of such data can be further enhanced in a general forensic scenario.

For the extension of the security framework derived in this dissertation for digital long-term archiving, the future work can be explored in the following aspects:

- The current security framework focuses on the contextualisation in the archiving module, thus it ensures the security requirements only when the digital data is being archived. However, no security assurance has been integrated into the progress before it enters or after it leaves the archive. Therefore, it is necessary to carry out the contextualisation to introduce further security assurance to the complete information lifecycle as described by the ILM.
- As mentioned earlier, the contextualisation described in this dissertation covers only the system level, so the resulting secure framework still needs to be further implemented on the data level. Therefore, the next step of work should be the derivation of a processing environment described by the security rules for each single functional entity, and apply data level context modelling to develop specific processing procedures of the IPs.
- The developed security framework poses a solution for digital evidence archiving. Therefore, in the future it is also necessary to verify the use cases used in the development with regard to corresponding legal regulations and protocols. As such regulations and protocols tend to evolve over time, new context should be introduced into the modelling process, resulting updates to the security framework.

[1] Funded in part by the German Federal Ministry of Education and Science (BMBF) through the research program under the contract no. FKZ: 13N10816.

As for the improvement of the developed separation approach of overlapped latent fingerprint, the following directions should be considered for the planning of future work:

- As a further pursuit regarding the integrity of the potential evidence, the local reliability of the separation results should be investigated. As the separation is conducted in a block-wise way, the reliability of the separation result in each of the blocks should be estimated (e.g. based on the compatibility probability value computed in the relaxation labelling step) and visualised, so the forensic expert can use such estimation as a reference to decide which part of the separated fingerprint, if not all of it, can be considered as reliable enough to contribute to the evidence chain.
- As a forensic approach needs to eventually be subject to the testing with corresponding standards, e.g. *Daubert* criteria, future work should also be planned for the separation approach regarding its *Daubert* compliance. To approach to a more statistically significant error rate, the objective assessment needs be extended. The test set should be expanded with larger number of latent fingerprint samples, and the diversity of the involved substrate types should also be increased. Alternative fingerprint matching software (e.g. VeriFinger 6.2 SDK mentioned in the state of the art) can also be considered, for the sake of better comparability of the testing results with those reported in the state-of-the-art literature. Furthermore, the current objective evaluation is established in conformity with the forensic scenario, i.e. it establishes the reliability of the separation results only based on the examination of the minutiae points (which is referred to as the second level of detail of fingerprint features), similar to what forensic experts do. However, in addition to such method, it is also necessary to assess the integrity of the separation results with regard to the first level of detail of fingerprint features, i.e. investigate on the global flow of resulting friction ridges. This could be done subjectively, by involving trained eyes, and/or using objective indicator that is automatically computed.
- The subjective assessment presented in this dissertation is conducted by the author from the point of view of the developer of the approach. However, as already mentioned earlier, such assessment makes better sense if conduced in a form of joint work of both the developer and the prospective user of the approach. Therefore, there is this gap to be closed in the future work: the approach needs to be presented in an understandable way, not only for the academia (e.g. as presented using series of formulas in [QSZ+2014]), but also for the forensic experts. As such the approach can be assessed more thoroughly, regarding the needs in the field of the forensic investigation, and ultimately be utilised there.

APPENDIX A

This appendix comprises all the security-related use-cases provided by SHAMAN project partners, and these use-cases contribute to the global context for the modelling.

Use Case No.	***Use Case Description***	***Roles***	***Objects***	***Rights***
UC-DOF1-510b	Registration of the Content Requester	Access System	certification list	read, write
UC-DOF1-620b	Adhering access rights to digital objects	Access System	archived object and its metadata	read, write
UC-DOF1-630b/UC-DOF3-531b	Performing the search and returning corresponding digital objects	Access System	archived object	read
UC-DOF1-630c/UC-DOF1-640/UC-DOF1-670b/UC-DOF1-680b/UC-DOF3-560b	Providing access to digital objects	Access System	archived object	read
UC-DOF1-631b	Performing the search and returning corresponding metadata entries	Access System	metadata of archived object, search index	read
UC-DOF1-632b	Performing the full-text search and returning corresponding digital objects	Access System	archived object, search index	read
UC-DOF3-530b	Execution of query and displaying the metadata	Access System	metadata of archived object	read
UC-DOF3-561b	Dissemination of the source code	Access System	archived object	read
UC-DOF3-562b	Dissemination of the publication	Access System	archived object	read

UC-DOF3-563b	Dissemination of the set of raw data	Access System	non-archived object	read
UC-DOF3-564b	Dissemination of the data product	Access System	archived object	read
UC-DOF3-565b	Execution of query and dissemination of the resulting data product	Access System	archived object	read
UC-DOF3-566b	Computation of the data product out of the raw data and dissemination afterwards	Access System	non-archived object	read
UC-DOF3-567b	Dissemination of the digital object in selected representation	Access System	archived object	read
UC-DOF3-570b	Performing automatic format migration	Access System	archived object and its metadata	read, write
UC-DOF1-340b	Manual error resolving	Administrator	non-archived object and its metadata	read
UC-DOF1-410	System administration	Administrator	archived objects and its metadata	read, write
UC-DOF1-441b	Dealing with the objects failed in the consistency check	Administrator	archived object	read
UC-DOF1-460	Execution of preservation measure	Administrator	archived objects and its metadata	read, write
UC-DOF3-310	Prepare data for ingestion	Administrator	archived object and its metadata	read
UC-DOF3-330	Correction of errors during ingesting	Administrator	archived object and its metadata	read
UC-DOF3-410	Move data between storage levels	Administrator	archived objects and its metadata	read, write
UC-DOF3-420	Refreshment of media	Administrator	archived object and its metadata	read
UC-DOF3-440	Recovery of the data after	Administrator	archived object	read

	disaster		and its metadata	
UC-DOF3-450b	Identification of digital objects stored in an obsolete format	Administrator	metadata of archived object	read
UC-DOF3-470b	Implementation of updating usage rights and dissemination restrictions	Administrator	certification list	read, write
UC-DOF3-580	Registration of the user	Administrator	certification list	read, write
UC-DOF3-540	Move selected data to fast access location	Administrator	archived object and its metadata	read
UC-DOF3-590	Audit the archive	Auditor	audit-trail	read
UCE2	Verification of the archival system, the archived objects and their metadata	Auditor	archived objects and its metadata	read
UC-DOF1-320	Import of the data into the system and indexing	Collection Manager	non-archived object and its metadata/index	read/ write
UC-DOF1-322	Separation of digital objects from delivery carrier media	Collection Manager	non-archived object	read
UC-DOF1-323	Creation of a catalogue entry from a new object	Collection Manager	catalogue	create
UC-DOF1-330	Generation and consignment of SIPs to the long-term archival system	Collection Manager	non-archived object and its metadata	read, write
UC-DOF1-340a	Monitoring the import and ingest process	Collection Manager	non-archived objects and its metadata	read
UC-DOF1-440	Collection management	Collection Manager	archived objects and its metadata	read, write, delete
UC-DOF1-441a	Consistency check to avoid corruption of the objects	Collection Manager	archived object and its metadata	read

UC-DOF1-442	Deleting objects from the long-term archive	Collection Manager	archived object and its metadata	read, write
UC-DOF1-443	Changing objects using update instead of replacement	Collection Manager	archived object and its metadata	read, write
UC-DOF1-470	Import and export of metadata	Collection Manager	metadata of archived object	read
UC-DOF1-620a	Access rights management	Collection Manager	certification list	read, write
UC-DOF1-680a	Access to long-term archive to retrieve digital objects and metadata	Collection Manager	archived object and its metadata	read
UC-DOF3-460	Update metadata	Collection Manager	metadata of archived object	read, write
UC-DOF3-571b	Performing manual format migration	Collection Manager	archived object and its metadata	read, write
UC-DOF1-310	Delivery of new material for collection	Content Provider	non-archived object and its metadata	read
UC-DOF1-311	Grant of usage rights for individual digital objects	Content Provider	certification list	write
UC-DOF1-312	Assignment of the object as a subsequent version of a previous one	Content Provider	metadata of non-archived object	read, write
UC-DOF1-641a	Providing access to a digital object	Content Provider	certification list	create
UC-DOF1-650	Collection sharing	Content Provider	certification list	write
UC-DOF1-660	Access to own original data	Content Provider	non-archived object and its metadata	read, write
UC-DOF1-670a	Access to changed version of own data	Content Provider	archived object and its metadata	read, write

UC-DOF3-130a	Providing necessary information	Content Provider	non-archived object and its metadata	read, write
UC-DOF3-320a	Ingestion of data and metadata into the archive	Content Provider	non-archived object and its metadata	read
UC-DOF3-470a	Requesting to update rights and dissemination restrictions	Content Provider	system metadata	read, write
UC-DOF1-510a	Providing registration information	Content Requester	certification list	write
UC-DOF1-630a/UC-DOF3-531a	Formulating a query to start a search for digital objects	Content Requester	archived object	read
UC-DOF1-631a	Formulating a query to start a search for metadata catalogue	Content Requester	catalogue	read
UC-DOF1-632a	Formulating a query to start a full-text search	Content Requester	archived object	read
UC-DOF1-641b	Access to digital object from content provider	Content Requester	archived object	read
UC-DOF1-642	Access to an alternative representation of a selected digital object	Content Requester	archived object	read
UC-DOF3-510	Access to metadata of the digital objects	Content Requester	metadata of archived object	read
UC-DOF3-520	Browse the metadata catalogue	Content Requester	catalogue	read
UC-DOF3-530a	Query for metadata	Content Requester	metadata of archived object	read
UC-DOF3-550	Access to metadata via OAI-PMH for harvesting system	Content Requester	metadata of archived object	read
UC-DOF3-560a	Selection of a specific digital object	Content Requester	archived object	read

UC-DOF3-561a	Selection of a specific source code	Content Requester	archived object	read
UC-DOF3-562a	Selection of a specific publication	Content Requester	archived object	read
UC-DOF3-563a	Selection of a set of raw data	Content Requester	archived object	read
UC-DOF3-564a	Selection of a specific data product	Content Requester	archived object	read
UC-DOF3-565a	Query for data	Content Requester	archived object	read
UC-DOF3-566a	Selection of a data product	Content Requester	archived object	read
UC-DOF3-567a	Selection of a digital object and an alternative representation	Content Requester	archived object	read
UC-DOF3-570a/UC-DOF3-571a	Selection of a digital object and its representation	Content Requester	archived object	read
WP1-TDUC012	Viewing data in obsolete formats	Content Requester	archived object	read
UC-DOF1-211b/UC-DOF3-210b	Registration of the content provider	Import System	certification list	read, write
UC-DOF1-315	Harvesting of publications	Import System	non-archived object and its metadata	read, write
UC-DOF3-320b	Verification of the ingestion and Backup	Import System	archived objects and its metadata	read
WP1-TDUC002	Indexing of ingested data	Index System	archived object and its metadata/index	read/write
WP1-TDUC003	Re-index Data	Index System	archived object and its	read/write

			metadata/index	
UC-DOF3-450a	Identification of an obsolete format	Preservation Manager	metadata of archived object	read
UC-DOF1-450	Preservation planning for objects whose file formats are in danger of technical obsolescence	Preservation Manager	archived objects and its metadata	read
UC-DOF3-461	Addition of erratum	Preservation Manager	metadata of archived object	write
UCE1	Creation of the Clark-Wilson certification lists	Preservation Manager	certification list	create
WP1-TDUC005	Semantic modelling relevant to indexing	Semantic Modeller	archived objects and its metadata	read, write
WP1-TDUC006	Illustrated semantic modelling relevant to indexing	Semantic Modeller	metadata of archived object	read

APPENDIX B

This appendix enumerates the audit and certification criteria from TRAC (Trustworthy Repositories Audit & Certification: Criteria and Checklist) [TRAC2007].

A Organisational infrastructure

A1 Governance & organizational viability

A1.1 Repository has a mission statement that reflects a commitment to the long-term retention of, management of, and access to digital information.

A1.2 Repository has an appropriate, formal succession plan, contingency plans, and/or escrow arrangements in place in case the repository ceases to operate or the governing or funding institution substantially changes its scope.

A2 Organizational structure & staffing

A2.1 Repository has identified and established the duties that it needs to perform and has appointed staff with adequate skills and experience to fulfil these duties.

A2.2 Repository has the appropriate number of staff to support all functions and services.

A2.3 Repository has an active professional development program in place that provides staff with skills and expertise development opportunities.

A3 Procedural accountability & policy framework

A3.1 Repository has defined its designated community(ies) and associated knowledge base(s) and has publicly accessible definitions and policies in place to dictate how its preservation service requirements will be met.

A3.2 Repository has procedures and policies in place, and mechanisms for their review, update, and development as the repository grows and as technology and community practice evolve.

A3.3 Repository maintains written policies that specify the nature of any legal permissions required to preserve digital content over time, and repository can demonstrate that these permissions have been acquired when needed.

A3.4 Repository is committed to formal, periodic review and assessment to ensure responsiveness to technological developments and evolving requirements.

A3.5 Repository has policies and procedures to ensure that feedback from producers and users is sought and addressed over time.

A3.6 Repository has a documented history of the changes to its operations, procedures, software, and hardware that, where appropriate, is linked to relevant preservation strategies and describes potential effects on preserving digital content.

A3.7 Repository commits to transparency and accountability in all actions supporting the operation and management of the repository, especially those that affect the preservation of digital content over time.

A3.8 Repository commits to defining, collecting, tracking, and providing, on demand, its information integrity measurements.

A3.9 Repository commits to a regular schedule of self-assessment and certification and, if certified, commits to notifying certifying bodies of operational changes that will change or nullify its certification status.

A4 Financial sustainability

A4.1 Repository has short- and long-term business planning processes in place to sustain the repository over time.

A4.2 Repository has in place processes to review and adjust business plans at least annually.

A4.3 Repository's financial practices and procedures are transparent, compliant with relevant accounting standards and practices, and audited by third parties in accordance with territorial legal requirements.

A4.4 Repository has ongoing commitment to analyse and report on risk, benefit, investment, and expenditure (including assets, licenses, and liabilities).

A4.5 Repository commits to monitoring for and bridging gaps in funding.

A5 Contracts, licenses, & liabilities

A5.1 If repository manages, preserves, and/or provides access to digital materials on behalf of another organization, it has and maintains appropriate contracts or deposit agreements.

A5.2 Repository contracts or deposit agreements must specify and transfer all necessary preservation rights, and those rights transferred must be documented.

A5.3 Repository has specified all appropriate aspects of acquisition, maintenance, access, and withdrawal in written agreements with depositors and other relevant parties.

A5.4 Repository tracks and manages intellectual property rights and restrictions on use of repository content as required by deposit agreement, contract, or license.

A5.5 If repository ingests digital content with unclear ownership/rights, policies are in place to address liability and challenges to those rights.

B Digital Object Management

B1 Ingest: acquisition of content

B1.1 Repository identifies properties it will preserve for digital objects.

B1.2 Repository clearly specifies the information that needs to be associated with digital material at the time of its deposit (i.e., SIP).

B1.3 Repository has mechanisms to authenticate the source of all materials.

B1.4 Repository's ingest process verifies each submitted object (i.e., SIP) for completeness and correctness as specified in B1.2.

B1.5 Repository obtains sufficient physical control over the digital objects to preserve them.

B1.6 Repository provides producer/depositor with appropriate responses at predefined points during the ingest processes.

B1.7 Repository can demonstrate when preservation responsibility is formally accepted for the contents of the submitted data objects (i.e., SIPs).

B1.8 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (Ingest: content acquisition).

B2 Ingest: creation of the archivable package

B2.1 Repository has an identifiable, written definition for each AIP or class of information preserved by the repository.

B2.2 Repository has a definition of each AIP (or class) that is adequate to fit long-term preservation needs.

B2.3 Repository has a description of how AIPs are constructed from SIPs.

B2.4 Repository can demonstrate that all submitted objects (i.e., SIPs) are either accepted as whole or part of an eventual archival object (i.e., AIP), or otherwise disposed of in a recorded fashion.

B2.5 Repository has and uses a naming convention that generates visible, persistent, unique identifiers for all archived objects (i.e., AIPs).

B2.6 If unique identifiers are associated with SIPs before ingest, the repository preserves the identifiers in a way that maintains a persistent association with the resultant archived object (e.g., AIP).

B2.7 Repository demonstrates that it has access to necessary tools and resources to establish authoritative semantic or technical context of the digital objects it contains (i.e., access to appropriate international Representation Information and format registries).

B2.8 Repository records/registers Representation Information (including formats) ingested.

B2.9 Repository acquires preservation metadata (i.e., PDI) for its associated Content Information.

B2.10 Repository has a documented process for testing understandability of the information content and bringing the information content up to the agreed level of understandability.

B2.11 Repository verifies each AIP for completeness and correctness at the point it is generated.

B2.12 Repository provides an independent mechanism for audit of the integrity of the repository collection/content.

B2.13 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (AIP creation).

B3 Preservation planning

B3.1 Repository has documented preservation strategies.

B3.2 Repository has mechanisms in place for monitoring and notification when Representation Information (including formats) approaches obsolescence or is no longer viable.

B3.3 Repository has mechanisms to change its preservation plans as a result of its monitoring activities.

B3.4 Repository can provide evidence of the effectiveness of its preservation planning.

B4 Archival storage & preservation/maintenance of AIPs

B4.1 Repository employs documented preservation strategies.

B4.2 Repository implements/responds to strategies for archival object (i.e., AIP) storage and migration.

B4.3 Repository preserves the Content Information of archival objects (i.e., AIPs).

B4.4 Repository actively monitors integrity of archival objects (i.e., AIPs).

B4.5 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (Archival Storage).

B5 Information management

B5.1 Repository articulates minimum metadata requirements to enable the designated community(ies) to discover and identify material of interest.

B5.2 Repository captures or creates minimum descriptive metadata and ensures that it is associated with the archived object (i.e., AIP).

B5.3 Repository can demonstrate that referential integrity is created between all archived objects (i.e., AIPs) and associated descriptive information.

B5.4 Repository can demonstrate that referential integrity is maintained between all archived objects (i.e., AIPs) and associated descriptive information.

B6 Access management

B6.1 Repository documents and communicates to its designated community(ies) what access and delivery options are available.

B6.2 Repository has implemented a policy for recording all access actions (includes requests, orders etc.) that meet the requirements of the repository and information producers/depositors.

B6.3 Repository ensures that agreements applicable to access conditions are adhered to.

B6.4 Repository has documented and implemented access policies (authorization rules, authentication requirements) consistent with deposit agreements for stored objects.

B6.5 Repository access management system fully implements access policy.

B6.6 Repository logs all access management failures, and staff review inappropriate "access denial" incidents.

B6.7 Repository can demonstrate that the process that generates the requested digital object(s) (i.e., DIP) is completed in relation to the request.

B6.8 Repository can demonstrate that the process that generates the requested digital object(s) (i.e., DIP) is correct in relation to the request.

B6.9 Repository demonstrates that all access requests result in a response of acceptance or rejection.

B6.10 Repository enables the dissemination of authentic copies of the original or objects traceable to originals.

C Technologies, Technical Infrastructure, & Security

C1 System infrastructure

C1.1 Repository functions on well-supported operating systems and other core infrastructural software.

C1.2 Repository ensures that it has adequate hardware and software support for backup functionality sufficient for the repository's services and for the data held, e.g., metadata associated with access controls, repository main content.

C1.3 Repository manages the number and location of copies of all digital objects.

C1.4 Repository has mechanisms in place to ensure any/multiple copies of digital objects are synchronized.

C1.5 Repository has effective mechanisms to detect bit corruption or loss.

C1.6 Repository reports to its administration all incidents of data corruption or loss, and steps taken to repair/replace corrupt or lost data.

C1.7 Repository has defined processes for storage media and/or hardware change (e.g., refreshing, migration).

C1.8 Repository has a documented change management process that identifies changes to critical processes that potentially affect the repository's ability to comply with its mandatory responsibilities.

C1.9 Repository has a process for testing the effect of critical changes to the system.

C1.10 Repository has a process to react to the availability of new software security updates based on a risk-benefit assessment.

C2 Appropriate technologies

C2.1 Repository has hardware technologies appropriate to the services it provides to its designated community(ies) and has procedures in place to receive and monitor notifications, and evaluate when hardware technology changes are needed.

C2.2 Repository has software technologies appropriate to the services it provides to its designated community(ies) and has procedures in place to receive and monitor notifications, and evaluate when software technology changes are needed.

C3 Security

C3.1 Repository maintains a systematic analysis of such factors as data, systems, personnel, physical plant, and security needs.

C3.2 Repository has implemented controls to adequately address each of the defined security needs.

C3.3 Repository staff have delineated roles, responsibilities, and authorizations related to implementing changes within the system.

C3.4 Repository has suitable written disaster preparedness and recovery plan(s), including at least one off-site backup of all preserved information together with an off-site copy of the recovery plan(s).

BIBLIOGRAPHY

[ABR2009] A. Agostini, C. Bettini, D. Riboni, "Hybrid reasoning in the CARE middleware for context-awareness," *International Journal of Web Engineering and Technology*, vol. 5, no. 1, pp. 3-23, 2009.

[Achinstein1968] P. Achinstein, *Concepts of Science. A Philosophical Analysis.* Baltimore: Johns Hopkins Press, 1968.

[Ackerlof1970] G. A. Ackerlof, "The market for 'lemons': quality uncertainty and the market mechanism," *Quarterly Journal of Economics*, vol. 84, pp. 488-500, 1970.

[AH2011] D. Allemang, J. Hendler, *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. 2nd ed., Morgen Kaufmann, 2011.

[ALC2012] D. Arend, M. Lange, C. Colmsee, S. Flemming, J. Chen, U. Scholz, "The e!DAL JAVA-API: store, share and cite primary data in life sciences," in *Proc. of 2012 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Philadephia, PA, USA, Oct. 2012, pp. 1-5.

[Ashbaugh1999] D. R. Ashbaugh, *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. CRC Press, 1999, ISBN: 978-0-8493-7007-6.

[Bailer-Jones2003] D. M. Bailer-Jones, "When scientific models represent," *International Studies in the Philosophy of Science*, vol. 17, pp. 59-74, 2003.

[Bauer2003] J. Bauer, "Identification and modeling of contexts for different information scenarios in air traffic," Diplomarbeit, Institut für Computergestützte Informationssysteme, Fakultät IV – Elektrotechnik und Informatik, TU Berlin, Germany, 2003.

[BB2002] D. M. Bailer-Jones, C. A. L. Beiler-Jones, "Modelling data: analogies in neural networks, simulated annealing and genetic algorithms," in *Model-Based Reasoning*, L. Magnani, N. Nersesssian, Eds. Springer US, 2002, pp. 147-165.

[BBH+2010] C. Bettini, O. Brdiczka, K. Henricksen, J. Indulska, D. Nicklas, A. Ranganathan, D. Riboni, "A survey of context modelling and reasoning techniques," *Pervasive and Mobile Computing*, vol. 6, pp. 161-180, 2010.

[BBG2004] R. Bhatti, E. Bertino, A. Ghafoor, "A trust-based context-aware access control model for web-services," in *Proc. of the IEEE Int. Conf. on Web Services*, Jul. 2004, pp. 184-191.

[BC2004] H. E. Byun, K. Cheverst, "Utilizing context history to provide dynamic adaptations," *Applied Artificial Intelligence*, vol. 18, no. 6, pp. 533-548, 2004.

[BCH+2000] J. Boyle, R. Cohen, S. Herzog, R. Rajan, A. Sastry, "The COPS (Common Open Policy Service) protocol", *RFC2748*, 2000.

[BCS+2006] C. Bolchini, C. Curino, F. A. Schreiber, L .Tanca, "Context integration for mobile data tailoring," in *Proc. 7th IEEE/ACM Int. Conf. on Mobile Data Management*, pp. 5, 2006.

[BCQ+2007] C. Bolchini, C. A. Curino, E. Quintarelli, F. A. Schreiber, L. Tanca, "A data-oriented survey on context models," *SIGMOD Record*, vol. 36, no. 4, pp. 19-26, Dec. 2007.

[BFA2005] A. van Bunningen, L. Feng, P. Apers, "Context for ubiquitous data management," in *Proc. of Int. Workshop on Ubiquitous Data Management*, Apr. 2005, pp. 17-24.

[Biba1977] K. J. Biba, *Integrity Considerations for Secure Computer Systems*. MTR-3153, the Mitre Corporation, Apr. 1977.

[Bishop2002] M. Bishop, *Computer Security: Art and Science*. Addison-Wesley, 2002, ISBN 0-201-44099-7.

[BKG+2009] C. Becker, H. Kulovits, M. Guttenbrunner, S. Strodl, A. Rauber, H. Hofman, "Systematic planning for digital preservation", *Int. Journal on Digital Libraries*, vol. 10, no. 10, pp.133–157, 2009.

[BKJ+2010] H. Brocks, A. Kranstedt, G. Jäschke, M. Hemmje, "Modeling context for digital preservation," in *Smart Information and Knowledge Management: Advances, Challenges, and Critical Issues*, E. Szczerbicki, N. T. Nguyen, Ed. Springer-Verlag Berlin Heidelberg, 2010, pp. 197-226.

[Black1962] M. Black, *Models and Metaphors. Studies in Language and Philosophy*. Ithaca, New York: Cornell University Press, 1962.

[Blakeslee1990] S. Blakeslee, "Lost on earth: wealth of data found in space," *The New York Times*, Mar. 20, 1990. Available: <http://www.nytimes.com/1990/03/20/science/lost-on-earth-wealth-of-data-found-in-space.html>

[BM1977] J. Bell, M. Machover, *A Course in Mathematical Logic*. Amsterdam, Noth-Holland, 1977.

[BMW+2011] J. G. Barnes, A. V. Marceo, K. Wertheim, et al., *The Fingerprint Sourcebook*. U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, 2011.

[Britannica2013a] Encyclopædia Britannica Inc. (2013). "scientific modelling," *Encyclopædia Britannica Online Academic Edition* [Online]. Available: <http://www.britannica.com/EBchecked/topic/387006/scientific-modeling>

[Britannica2013b] Encyclopædia Britannica Inc. (2013). "computer science," *Encyclopædia Britannica Online Academic Edition* [Online]. Available: <http://www.britannica.com/EBchecked/topic/130675/computer-science>

[Britannica2013c] Encyclopædia Britannica Inc. (2013). "computer security," *Encyclopædia Britannica Online Academic Edition* [Online]. Available: <http://www.britannica.com/EBchecked/topic/130682/computer-security>

[BS1989] W. H. Boshoff, S. H. von Solm, "A path context model for addressing security in potentially non-secure environments," *Computers & Security*, vol. 8, no. 5, pp. 417-425, 1989.

[BS2002] R. Baskerville, M. Siponen, "An information security meta-policy for emergent organizations," *Logistics Information Management*, vol. 15, no. 5/6, pp. 337-346, 2002.

[BSF2009] R. Bhargava, R. Schwarz Perlman, D. C. Fernandez, I. W. Levin, E. G. Bartick, "Noninvasive detection of superimposed latent fingerprints and inter-ridge trace evidence by infrared spectroscopic imaging," *Analytical and Bioanalytical Chemistry*, vol. 294, no. 8, pp. 2069-2075, 2009.

[Cameron2011] S. Cameron, "Digital evidence," *FBI Law Enforcement Bulletin*, vol. 80, no. 8. Federal Bureau of Investigation, U.S. Department of Justice, 2011.

[Cappelli2009] R. Cappelli, "Synthetic fingerprint generation" in *Handbook of Fingerprint Recognition*, 2nd ed., D. Maltoni, D. Maio, A. K. Jain, S. Prabhakar, Eds. London, UK: Springer, 2009, pp. 271-302.

[Cartwright1989] N. Cartwright, *Nature's Capacities and their Measurement*. Oxford: Oxford University Press, 1989.

[CASPAR2007] CASPAR Consortium, *CASPAR Guidelines*, CASPAR-D1202-TN-0101-1_0, May 2007.

[CBP+2007] N. J. Crane, E. G. Bartick, R. S. Perlman, S. Huffman, "Infrared spectroscopic imaging for noninvasive detection of latent fingerprints," *Journal of Forensic Science*, vol. 52, no. 1, pp. 48-53, 2007.

[CFJ+2011] F. Chen, J. Feng, A. K. Jain, J. Zhou, J. Zhang, "Separating overlapped Fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 2, pp. 346-359, 2011.

[Congress1923] U.S. Congress, *Frye v. United States, 293 F. 1013, 1014 (D.C. Cir.)*, 1923.

[Congress2011] U.S. Congress, *Federal Rules of Evidence* (amended by U.S. Supreme Court Apr. 26, 2011, effective since Dec. 1, 2011), 2011.

[Contessa2007] G. Contessa, "Scientific representation, interpretation and surrogative reasoning," *Philosophy of Science*, vol. 74, no. 1, pp. 48-68, 2007.

[CLM+2004] C. Champod, C. Lennard, P. Margot, M. Stoilovic, *Fingerprints and Other Ridge Skin Impressions*. CRC Press, 2004, ISBN: 978-0-415-27175-2.

[CMD1999] K. Cheverst, K. Mitschell, N. Davies, "Design of an object model for a context sensitive tourist GUIDE," *Computers and Graphics*, vol. 23, no. 6, pp. 883-891, 1999.

[CPG+2010] R. M. Connatser, S. M. Prokes, O. J. Glembocki, R. L. Schuler, C. W. Gardner, S. A. Lewis, "Toward surface-enhanced Raman imaging of latent fingerprints," *Journal of Forensic Sciences*, vol. 55, pp. 1462-1470, 2010.

[CW1987] D. D. Clark, D. R. Wilson, "A comparison of commercial and military computer security policies," *IEEE Symposium on Security and Privacy*, 1987.

[Dey2001] A. K. Dey, "Understanding and using context," *Personal Ubiquitous Computing*, vol. 5, issue 1, pp. 4-7, Feb. 2001.

[DF2003] N. Da Costa, S. French, *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. Oxford: Oxford University Press, 2003.

[DRAMBORA2014] DigitalPreservationEurope and Digital Curation Centre (2014, Jun 05). *DRAMBORA Interactive: Digital Repository Audit Method Based on Risk Assessment* [Online]. Available: <http://www.repositoryaudit.eu/>

[DSS+2006] A. K. Dey, T. Sohn, S. Streng, J. Kodama, "iCAP: interactive prototyping of context-aware applications," in *Proc. 4th Int. Conf. on Pervasive Computing*, pp. 254-271, 2006.

[DT2009] K. Dunkerley, G. Tejay, "Developing an information systems security success model for e-gevernment context," in Proc. of Americas Conf. on Information Systems, San Francisco, CA, USA, Aug. 2009.

[EKB+2009] F. Engel, C. Klas, H. Brocks, A. Kranstedt, G. Jäschke, M. Hemmje, "Towards supporting context-oriented information retrieval in a scientific-archive based information lifecycle," in *Proc. of Cultural Heritage on line. Empowering users: an active role for user communities*, Florence, Italy, Dec. 2009.

[Elgin2010] C. Elgin, "Telling instances," in *Beyond Mimesis and Nominalism: Representation in Art and Science*, R. Frigg, M. Hunter, Eds. Berlin and New York: Springer, pp. 1-17, 2010

[FC2004] M. Folk, V. Choi, "Scientific formats for geospatial data preservation – a study of suitability and performance," NCSA/NARA Technical Report, Jan. 8, 2004.

[FH2012] R. Frigg, S. Hartmann, "Models in Science," *The Stanford Encyclopedia of Philosophy (Fall 2012 Edition)*, Edward N. Zalta (ed.), 2012. Available: <http://plato.stanford.edu/archives/fall2012/entries/models-science/>

[FKW+2014] M. Filax, A. Kenner, S. Wabnitz, B. Landmesser, M. Ulrich, T. Leich, "Fabige Daktyloskopie Verifizierte Separation überlagerter Fingerspuren," (in German), *Kriminalistik*, issue 5/2014, pp. 297-302, 2014.

[FM2008] J. B. Filho, H, Martin, "QACBAC: an owner-centric QoC-aware context-based access control model for pervasive environments," in *Proc. of ACM SPRINGL '08*, Irvine, CA, USA, 2008, pp. 30-38.

[FN2006] K. Falkovych, F. Nack, "Context aware guidance for multimedia authoring: harmonizing domain and discourse knowledge," *Multimedia Systems*, vol. 11, issue 3, pp. 226-235, Mar. 2006.

[Frigg2006] R. Frigg, "Scientific representation and the semantic view of theories," *Theoria*, vol. 55, pp. 37-53, 2006.

[Frigg2010] R. Frigg, "Fiction and scientific representation," in *Beyond Mimesis and Nominalism: Representation in Art and Science*, R. Frigg, M. Hunter, Eds. Berlin and New York: Springer, pp. 97-138, 2010.

[FS1994] M. Forster, E. Sober, "How to tell when simple, more unified, or less ad hoc theories will provide more accurate predictions," *British Journal for the Philosophy of Science*, vol 45, pp. 1-35, 1994.

[FSZ2012] J. Feng, Y Shi, J. Zhou, "Robust and efficient algorithms for separating latent overlapped fingerprints", *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1498-1510, 2012.

[Galison1997] P. Galison, *Image and Logic. A Material Culture of Microphysics*. Chicago: University of Chicago Press, 1997.

[Garfinkel2010] S. L. Garfinkel, "Digital forensics research: the next 10 years," in *Proc. of the 10th Annual DFRWS Conf.*, vol. 7, 2010.

[GBB2012] A. P. Gibson, M. Bannister, S. M. Bleay, "A comparison of three ultraviolet searching and imaging systems for the recovery of fingerprints," *Journal of Forensic Identification*, vol. 62, pp. 349-367, 2012.

[GFV2012] S. Gruhn, R. Fischer, C. Vielhauer, "Surface classification and detection of latent fingerprints based on 3D surface texture parameters," in *SPIE – Optics, Photonics, and Digital Technologies for Multimedia Application II*, Brussels, Belgium, 2012.

[Giaretta2011] D. Giaretta, *Advanced Digital Preservation*. Springer-Verlag Berlin Heidelberg, 2011, ISBN 978-3-642-16809-3.

[Giere1988] R. Giere, *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press, 1988.

[Giere2004] R. Giere, "How models are used to represent reality," *Journal of Philosophy*, vol. 71, no. 5, pp. 742-752, 2004.

[GS2001] P. Gray, D. Salber, "Modelling and using sensed context information in the design of interactive applications," in *Proc. of 8th IFIP Int. Conf. on Engineering for Human-Computer Interaction (EHCI 2001)*, Toronto, Canada, May 2001, pp. 317-336.

[GST+2013] J. A. Guicheteau, H. Swofford, A Tripathi, P. G. Wilcox, E. D. Emmons, S. D. Christesen, J. Wood, A. W. Fountain III, "Sequential Raman chemical imaging and biometric analysis on fingerprints for rapid identification of threat materials and individuals," *Journal of Forensic Identification*, vol. 63, pp. 90-101, 2013.

[GV1978] A. Gibbard, H. Varian, "Economic models," *Journal of Philosophy*, vol 75, pp 664-677, 1978.

[GV2011] S. Gruhn, C. Vielhauer, "Surface classification and detection of latent fingerprints: novel approach based on surface texture parameters," in *Proc. of 7th International Symposium on Image and Signal Processing and Analysis (ISPA)*, Dubrovnik, Croatia, 2011.

[GWM+2004] M. D. Garris, C. I. Watson, R. M. McCabe, C. L. Wilson, *User's guide to NIST Fingerprint Image Software (NFIS)*, NISTIR 6813, NIST, U.S. Dept. of Commerce: Gaithersburg, MD, 2004

[GWP+2004] T. Gu, X. H. Wang, H. K. Pung, D. Q. Zhang, "An ontology-based context model in intelligent environments," in *Proc. of Communication Networks and Distributed Systems Modeling and Simulation Conf.*, San Diego, California, USA, 2004.

[HAC+2004] R. Hill, J. Al-Muhtadi, R. Campbell, A. Kapadia, P. Naldurg, A. Ranganathan, "A middleware architecture for securing ubiquitous computing cyber infrastructures," *IEEE Distributed Systems Online*, vol. 5, issue 9, pp. 1-14, Sep. 2004.

[Harris2003] T. Harris, "Data models and the acquisition and manipulation of data," *Philosophy of Science*, vol. 70, pp. 1508-1517, 2003.

[Hedstrom1997] M. Hedstrom, "Digital preservation: a time bomb for digital libraries," *Computers and the Humanities*, vol. 31, issue 3, pp. 189-202, 1997.

[Henricksen2003] K. Henricksen, "A framework for context-aware pervasive computing applications," Ph. D. dissertation, School of Information Technology and Electrical Engineering, University of Queensland, Australia, Sep. 2003.

[Hesse1963] M. Hesse, *Models and Analogies in Science*. London: Sheed and Ward, 1963.

[Hesse1974] M. Hesse, *The Structure of Scientific Inference*. London: Macmillan, 1974.

[HDP2011] M. Hildebrandt, J. Dittmann, M. Pocs, M. Ulrich, R. Merkel, T. Fries, "Privacy preserving challenges: new design aspects for latent fingerprint detection systems with contact-less sensors for future preventive applications in airport luggage handling," in *BioID 2011*, Brandenburg an der Havel, Germany, 2011, pp. 286-298.

[HIR2003] K. Henricksen, J. Indulska, A. Rakotonirainy, "Generating context management infrastructure for high-level context models," in *Industrial Track Proc. of the 4th Int. Conf. on Mobile Data Management*, Melbourne, Australia, Jan. 2003, pp. 1-6.

[HMQ+2013] M. Hildebrandt, A. Makrushin, K Qian, J Dittmann, "Visibility assessment of latent fingerprints on challenging substrates in spectroscopic scans," in *Proc. of Communications and Multimedia Security*, Magdeburg, Germany, Sep. 2013, pp. 200-203.

[Hodges1997] W. Hodges, *A Short Model Theory*. Cambridge: Cambridge University Press, 1997.

[HRL2011] E. H. Holder, L. O. Robinson, J. H. Laub, *The Fingerprint Sourcebook*. US Department of Justice, Office of Justice Programs, National Institute of Justice, 2011.

[HS2010] M. M. Houck, J. A. Siegel, *Fundamentals of Forensic Science* (second edition). Academic Press, 2010. ISBN: 978-0-12-374989-5.

[HSD+2013] M. Hildebrandt, J. Sturm, J. Dittmann, C. Vielhauer, "Creation of a public corpus of contact-less acquired latent fingerprints without privacy implications", in *Proc. of Communications and Multimedia Security*, Magdeburg, Germany, Sep. 2013, pp. 204-206.

[HSK2009] J. Hong, E. Suh, S. Kim, "Context-aware systems: a literature review and classification," *Expert Systems with Applications*, vol. 36, issue 4, pp. 8509-8522, May 2009.

[HT1995] K. Holyoak, P. Thagard, *Mental Leaps. Analogy in Creative Thought*. Cambridge, Mass.: Bradford, 1995.

[HWD2013] B. Hu, Z. X. Wang, Q. C. Dong, "A novel context-aware modeling and reasoning method based on OWL," *Journal of Computers*, vol. 8, no. 4, pp. 943-950, 2013.

[IRM+2009] P. Innocenti, S. Ross, E. Maceviciute, T. Wilson, J. Ludwig, W. Pempe, "Assessing digital preservation frameworks: the approach of the SHAMAN project," in *Proc. of International ACM Conference on Management of Emergent Digital EcoSystems (MEDES'09)*, Lyon, France, Oct 2009.

[ISO1989] *Information processing systems – Open Systems Interconnection – Basic Reference Model – Part 2: Security Architecture*, ISO 7498-2:1989, 1989. / *Security architecture for Open Systems Interconnection for CCITT applications*, ITU-T X.800, 1991.

[ISO2012] *Information technology – Security techniques – Information security management systems – Overview and vocabulary*, ISO/IEC 27000:2012, 2012.

[JHS+2012] M. Jankow, M. Hildebrandt, J. Sturm, S. Kiltz, C. Vielhauer, "Performance analysis of digital cameras versus chromatic white light (CWL) sensors for the localization of latent fingerprints in crime scenes," in *SPIE – Optics, Photonics, and Digital Technologies for Multimedia Application II*, Brussels, Belgium, 2012.

[JL2002] X. Jiang, J. A. Landay, "Modeling privacy control in context-aware systems," *IEEE Pervasive Computing*, vol. 1, issue 3, pp. 59-63, Jul.-Sep. 2002.

[Johnson2009] G. Johnson, "Towards shrink-wrapped security: a taxonomy of security-relevant context," in *Proc. of IEEE Int. Conf. on Pervasive Computing and Communication*, Galveston, TX, USA, 2009, pp. 1-2.

[JSB2008] J. Jürjens, J. Schreck, P. Bartmann, "Model-based security analysis for mobile communications," in *Proc. of ICSE '08*, Leipzig, Germany, May. 2008, pp. 683-692.

[JSG+2011] G. Johnson, P. Shakarian, N. Gupta, A. Agrawala, "Towards shrink-wrapped security: practically incorporating context into security services," Procedia Computer Science, vol. 5, pp. 782-787, 2011.

[KGL+2011] R. Kärgel, S. Giebel, M. Leich, J. Dittmann, "Separation and sequence detection of overlapped fingerprints: experiments and first results," in *Security+Defense Proc. SPIE 8189*, 2011

[KLD2007] S. Kiltz, A. Lang, J. Dittmann, "Taxonomy for Computer Security Incidents," in *Cyber Warfare and Cyber Terrorism*, L. J. Janczewski, A. M. Colarik, Ed. Information Science Reference (IGI Global), ISBN 978-1-59140-991-5, 2007.

[Knuuttila2009] T. Knuuttila, "Representation, idealisation and fiction in economics: from the assumptions issue to the epistemology of modelling," in *Fictions in Science. Philosophical Essays on Modelling and Idealisation*, M. Suárez, Ed. London: Routledge, pp. 205-233, 2009.

[KRW+2004] G. Klyne, F. Reynolds, C. Woodrow, H. Ohto, J. Hjelm, M. H. Butler, L. Tran, "Composition capability/preference profiles (CC/PP): structure and vocabularies 1.0," W3C Recommendation, Tech. Rep., W3C, Jan. 2004.

[Kroes1989] P. Kroes, "Structural analogies between physical systems," *British Journal for the Philosophy of Science*, vol. 40, pp. 145-154, 1989.

[KS2012] M. F. F. Khan, K. Sakamura, "Context-awareness: exploring the imperative shared context of security and ubiquitous computing," in *Proc. of 14th Int. Conf. on Inform. Integration and Web-based Applications & Services*, Bali, Indonesia, Dec. 2012, pp. 101-110.

[Landesman2009] M. Landersman, "What is a virus signature?" *About.com*, 2009. Available: <http://antivirus.about.com/od/whatisavirus/a/virussignature.htm>

[Laymon1982] R. Laymon, "Scientific realism and the hierarchical counterfactual path from data to theory," in *Proc. of the Biennial Meeing of the Philosophy of Science Association*, vol. 1, pp. 107-121, 1982.

[Laymon1991] R. Laymon, "Thought experiments by Stevin, Mach and Gouy: thought experiments as ideal limits and semantic domains," in *Horowitz and Massey*, 1991, pp. 167-192.

[Lee2011] C. A. Lee, "A framework for contextual information in digital collections," *Journal of Documentation*, vol. 67, no. 1, pp. 95-143, 2011.

[LII2013] Legal Information Institute (LII), Cornell University Law School. (2013, Aug 6). *Federal Rules of Evidence – Notes on FRE 702 [Online]*. Available: <http://www.law.cornell.edu/rules/fre/rule_702>

[Mäki1994] U. Mäki, "Isolation, idealization and truth in economics," in *Idealization VI: Idealization in Economics*, B. Hammings, N. B. De Marchi, Eds. Poznan Studies in the Philosophy of the Sciences and the Humanities, Amsterdam: Rodopi, vol. 38, pp. 147-168, 1994.

[Mayo1996] D. Mayo, *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press, 1996.

[McMullin1968] E. McMullin, "What do physical models tell us?" in *Logic, Methodology and Science III*, B. van Rootselaar, J. F. Staal, Eds. Amsterdam: North Holland, pp. 385-396, 1968.

[McMullin1985] E. McMullin, "Galilean idealization," *Studies in the History and Philosophy of Science*, vol. 16, pp. 247-273, 1985.

[MGD+2012] R. Merkel, S. Gruhn, J. Dittmann, C. Vielhauer, A. Bräutigam, "On non-invasive 2D and 3D chromatic white light image sensors for age determination of latent fingerprints," *Forensic Science International*, vol. 222, pp. 52-70, 2012.

[MHF+2012] A. Makrushin, M. Hildebrandt, R. Fischer, T. Kietscher, J. Dittmann, C. Vielhauer, "Advanced techniques for latent fingerprint detection and validation using a CWL device," in *SPIE – Optics, Photonics, and Digital Technologies for Multimedia Application II*, Brussels, Belgium, 2012.

[MK2005] K. Minami, D. Kotz, "Secure context-sensitive authorization," *Pervasive and Mobile Computing*, vol. 1, issue 1, pp. 123-156, Mar. 2005.

[Moore1989] R. T. Moore, "Analysis of ridge-to-ridge distance on fingerprints," *Journal of Forensic Identification*, vol. 39, no. 4, pp. 231-238, 1989.

[Morrison1998] M. Morrison, "Modelling nature: between physics and the physical world," *Philosophia Naturalis*, vol. 35, pp. 65-85, 1998.

[Morrison2009] M. Morrison, "Fictions, representations and reality," in *Fictions in Science. Philosophical Essays on Modelling and Idealisation*, M. Suárez, Ed. London: Routledge, pp. 110-135, 2009.

[MPD+2013] R. Merkel, M. Pocs, J. Dittmann, C. Vielhauer, "Proposal of non-invasive fingerprint age determination to improve data privacy management in police work from a legal perspective using the example of Germany," in *Data Privacy Management and Autonomous Spontaneous Security*, Pisa, Italy, 2013, pp. 61-74.

[Mundy1986] B. Mundy, "On the general theory of meaningful representation," *Syntheses*, vol. 67, pp. 391-437, 1986.

[Musgrave1981] A. Musgrave, "'Unreal assumptions' in economic theory: the F-twist untwisted," *Kyklos*, vol. 34, pp. 377-387, 1981.

[NPC1996] *Archives Law of People's Republic of China*. Standing Committee of the Eighth National People's Congress, P.R. China, 1996.

[Nye2012] J. Nye, "Charges against fugitive internet prescription drugs doctor dropped because of too much evidence," *Mail Online*, Aug 16, 2012. Available: <http://www.dailymail.co.uk/news/article-2189320/Charges-fugitive-internet-prescription-drugs-doctor-dropped-MUCH-evidence.html>

[OAIS2002] Consultative Committee for Space Data Systems (CCSDS), *Reference model for an Open Archival Information System (OAIS), Recommendation for Space Data System Standards*, CCSDS 650.0-B-1, Blue Book (ISO 14721:2003), 2002.

[OAIS2009] Consultative Committee for Space Data Systems (CCSDS), *Reference model for an Open Archival Information System (OAIS), draft recommended standard, Recommendation for Space Data System Standards*, CCSDS 650.0-P-1.1, Pink Book, 2009.

[Oxford2013a] Oxford University Press. (2013, Apr 13). "modelling." *Oxford Dictionary. Oxford Dictionaries, British and world version* [Online]. Available: <http://oxforddictionaries.com/definition/english/modelling>

[Oxford2013b] Oxford University Press. (2013, Jun 27). "context." *Oxford Dictionary. Oxford Dictionaries, British and world version* [Online]. Available: <http://oxforddictionaries.com/definition/english/context>

[Oxford2013c] Oxford University Press. (2013, Aug 1). "forensic." *Oxford Dictionary. Oxford Dictionaries, British and world version* [Online]. Available: <http://oxforddictionaries.com/definition/english/forensic>

[Pascoe1998] J. Pascoe, "Adding generic contextual capabilities to wearable computers," in *Proc. of 2nd International Symposium on Wearable Computers*, pp. 92-99, 1998.

[PCJ+2008] M. D. Preda, M. Christodorescu, S. Jha, S. Debray, "A semantic-based approach to malware detection," *ACM Trans. on Programming Languages and Systems*, vol. 30, issue 5, pp. 25-54, Aug. 2008.

[Peirce1931-1958] C. S. Peirce, *Collected Papers of Charles Sanders Peirce. Vol. 3.* Harvard University Press, Cambridge, Mass, 1931-1958.

[PES2010] C. A. Plese, D. L. Exline, S. D. Stewart, "Improved methods of visible hyperspectral imaging provide enhanced visualization of untreated latent fingerprints," *Journal of Forensic Identification*, vol. 60, pp. 603-618, 2010.

[Pfeiffer2012] E. Pfeiffer, "Charges dropped against fugitive doctor, because evidence is using too much space on federal servers," *Yahoo! News*, Aug. 16, 2012. Available: <http://news.yahoo.com/blogs/sideshow/federal-charges-dropped-against-fugitive-doctor-because-too-184830159.html>

[PGM1990] C. A. Pounds, R. Grigg, T. Mongkolaussavaratana, "The use of 1,8-diazafluoren-9-one (DFO) for the fluorescent detection of latent fingerprints on paper: a preliminary evaluation," *J. Forensic Sci.*, vol. 35, no. 1, pp. 169-175, 1990.

[PNS+2000] D. Petrelli, E. Not, C. Strapparava, O. Stock, M. Zancanaro, "Modeling context is like taking pictures," in *Proc. of the Workshop "The What, Who, Where, When, Why and How of Context-Awareness" in CHI2000*, 2000.

[Psillos1995] S. Psillos, "The cognitive interplay between theories and models: the case of 19th century physics," *Poznan Studies in the Philosophy of the Sciences and Humanities*, vol. 44, pp. 105-133, 1995.

[PZC+2014] C. Perera, A. Zaslavsky, P. Christen, D. Georgakopoulos, "Context aware computing for the Internet of Things: a survey," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 414-454, 2014.

[QSD2013] K. Qian, M. Schott, J. Dittmann, "Separation of contactless captured high-resolution overlapped latent fingerprints: parameter optimisation and evaluation," in *Proc. of 2013 International Workshop on Biometrics and Forensics IWBF 2013*, Lisbon, Portagal, 2013.

[QSK+2011a] K. Qian, M. Schott, C. Kraetzer, M. Hemmje, "Contextualizing security for digital long-term preservation," in *Proc. of the 13th ACM Workshop on Multimedia and Security (MM&Sec'11)*, Niagara Falls, NY, USA, 2011.

[QSK+2011b] K. Qian, M. Schott, C. Kraetzer, M. Hemmje, J. Dittmann, "A security contextualisation framework for digital long-term preservation," in *Proc. of the International Workshop on Semantic Digital Archive, (part of the 15th International Conference on Theory and Practice of Digital Libraries (TPDL))*, Berlin, Germany, 2011.

[QSS+2012] K. Qian, M. Schott, W. Schöne, M. Hildebrandt, "Separation of high-resolution samples of overlapping latent fingerprints using relaxation labelling," in *Proc. SPIE 8436, Optics, Photonics, and Digital Technologies for Multimedia Applications II, 84361A*, Brussels, Belgium, 2012.

[QSZ+2014] K. Qian, M. Schott, W. Zheng, J. Dittmann, "A context-based approach of separating contactless captured high-resolution overlapped fingerprints," *IET Biometrics*, vol. 3, no. 2, pp. 101-112, Jun 2014. doi:10.1049/iet-bmt.2013.0057

[RCD+1998] T. Rodden, K. Cheverst, K. Davies, A. Dix, "Exploiting context in HCI design for mobile systems," in *Workshop on Human Computer Interaction with Mobile Devices*, 1998.

[Redhead1980] M. Redhead, "Models in physics," *British Journal for the Philosophy of Science*, vol. 31, pp. 145-163, 1980.

[RFJ2007] A. Roßnagel, S. Fischer-Dieskau, S. Jandt, "Handlungsleitfaden zur Aufbewahrung elektronischer und elektronisch signierter Dokumente,"(in German), Federal Ministry of Economics and Technology, Aug. 2007.

[RL2013] A. Richards, R. Leintz, "Forensic reflected ultraviolet imaging," *Journal of Forensic Identification*, vol. 63, pp. 46-69, 2013.

[RRC+2006] P.-G. Raverdy, O. Riva, A. de La Chapelle, R. Chibout, V. Issarny, "Efficient context-aware service discovery in multi-protocol pervasive environments," in *Proc. 7th IEEE/ACM Int. Conf. on Mobile Data Management*, pp. 3, 2006.

[SAA1999] *Measures for the implementation of Archives Law of People's Republic of China* [中华人民共和国档案法实施办法]. State Archives Administration, P.R. China, 1999.

[SAA2012] *Measures for the transfer and archiving of electronic documents* [电子档案移交与接收办法]. State Archives Administration, P.R. China, 2012.

[SAW1994] B. N. Schilit, N. L. Adams, R. Want, "Context-aware computing applications," in *IEEE Workshop on Mobile Computing Systems and Applications*, Santa Cruz, CA, USA, 1994.

[Schilling2006] C. Schilling, "Digital evidence storage and preservation," *Microgram Bulletin*, vol. XXXIX, no. 1, the Drug Enforcement Administration, U.S. Department of Justice, 2006.

[SDA1999] D. Salber, A.K. Dey, G.D. Abowd, "The context toolkit: aiding the development of context-enabled applications," in *Proc. of CHI'99*, pp. 434-441, 1999.

[SDV+2008] M. Schott, J. Dittmann, C. Vielhauer, C. Kraetzer, A. Lang, "Integrity and authenticity for digital long-term preservation in iRods grid infrastructure," in *Proc. of the 6th International Workshop for Technical, Economic and Legal Aspects of Business Models for Virtual Goods incorporating the 4th International ODRL Workshop*, Poznań, Poland, Oct. 2008.

[Serrano2012] J. M. Serrano Orozco, *Applied Ontology Engineering in Cloud Services, Networks and Management Systems*. Springer Science+Business Media, 2012.

[SFZ2011] Y. Shi, J. Feng, J. Zhou, "Separating overlapped fingerprints using constrained relaxation labelling," in *Proc. of International Joint Conference on Biometrics*, 2011.

[SHAMAN2009] P. Innocenti, B. Aitken, A. Hasan, "SHAMAN-WP1-D1.2: SHAMAN requirements analysis report (public version) and specification of the SHAMAN assessment framework and protocol," SHAMAN, Mar. 2009.

[SK2001] G. S. Sodhi, J. Kaur, "Powder method for detecting latent fingerprints: a review," *Forensic Sci. Int.*, vol. 120, no. 3, pp. 172-176, 2001.

[SKD+2010] M. Schott, C. Kraetzer, J. Dittmann, C. Vielhauer, "Extending the Clark-Wilson security model for digital long-term preservation use-cases," in *Proc. of Multimedia on Mobile Devices, 2010, SPIE Electronic Imaging Conference 7542*, San Jose, CA, USA, Jan. 2010.

[SL2004] T. Strang, C. Linnhoff-Popien, "A context modeling survey," in *Proc. of the First International Workshop on Advanced Context Modelling, Reasoning and Management, in conjunction with UbiComp 2004*, Nottingham, England, 2004.

[SLC2009] J. Seifert, A. De Luca, B. Conradi, "A context-sensitive security model for privacy protection on mobile phones," in *Proc. of MobileHCI'09*, Bonn, Germany, Sep. 2009.

[SMD2012] M. Schott, R Merkel, J. Dittmann, "Sequence detection of overlapping latent fingerprints using a short-term aging feature," in *Proc. of IEEE International Workshop on Information Forensics and Security*, Tenerife, Spain, Dec 2012, pp. 85-90.

[SQK+2011] M. Schott, K. Qian, C. Krätzer, J. Dittmann, "Kontextmodellierung und Policies für die Langzeitarchivierung,"(in German), in Proc. of D•A•CH Security 2011, Oldenburg, Germany, 2011.

[SS2006] M. Suárez, A. Solé, "On the analogy between cognitive representation and truth," *Theoria*, vol. 55, pp. 27-36, 2006.

[SSE+2010] J. J. Simpson, M. J. Simpson, B. Endicott-Popovsky, V. Popovsky, "Secure software education – a contextual model-based approach," *Int. J. of Secure Software Eng.*, vol. 1, issue 4, pp. 35-61, Oct. 2010.

[ST1994] B. Schilit, M. Theimer; "Disseminating active map information to mobile hosts," *IEEE Network*, vol. 8, no. 5, pp. 22-32, 1994.

[Staley2004] K. W. Staley, *The Evidence for the Top Quark: Objectivity and Bias in Collaborative Experimentation*. Cambridge: Cambridge University Press, 2004.

[STK2006] D. K. Singh, S. Tripathi, P. K. Kalra, "Separation of image mixture using complex ICA," in Proc. of *the 9th Asian Symposium on Information Display ASID06*, 2006, pp. 314-317.

[STZ2005] X. Shen, B. Tan, C. Zhai, "Context-sensitive information retrieval using implicit feedback," in *Proc. 28th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 43-50, 2005.

[Suárez2003] M. Suárez, "Scientific representations: against similarity and isomorphism," International Studies in the Philosophy of Science, vol. 17, pp. 225-244, 2003

[Suárez2004] M. Suárez, "An inferential conception of scientific representation," Philosophy of Science, vol 71, pp. 767-779, 2004.

[Suárez2009] M. Suárez, "Scientific fictions as rules of inference," in *Fictions in Science. Philosophical Essays on Modelling and Idealisation*, M. Suárez, Ed. London: Routledge, pp. 158-178, 2009.

[Suppes1962] P. Suppes, "Models of data," in *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*. Stanford: Stanford University, pp. 252-261, 1962.

[Suppes2002] P. Suppes, *Representation and Invariance of Scientific Structures*. Stanford: CSLI Publications, 2012.

[Swoyer1991] C. Swoyer, "Structural Representation and surrogative reasoning," *Synthese*, vol. 87, pp. 449-508, 1991.

[TBS2006] A. Tonazzini, L. Bedini, E. Salerno, "A Markov model for blind image separation by a mean-field EM algorithm," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 473-482, 2006.

[Teller2001] P. Teller, "Twilight of the perfect model," *Erkenntnis*, vol. 55, pp. 393-415, 2001.

[Thomson-Jones2010] M. Thomson-Jones, "Missing systems and the face value practice," *Synthese*, vol. 172, no. 2, pp. 283-299, 2010.

[TLC+2010] H. Tang, W. Lu, C. Che, K. Ng, "Gold nanoparticles and imaging mass spectrometry: double imaging of latent fingerprints," *Anal. Chem.*, vol. 82, no. 5, pp. 1589-1593, 2010.

[Toon2010] A. Toon, "Models as make-believe," in *Beyond Mimesis and Nominalism: Representation in Art and Science*, R. Frigg, M. Hunter, Eds. Berlin and New York: Springer, pp. 71-96, 2010.

[Toon2011] A. Toon, "Playing with molecules," *Studies in History and Philosophy of Science*, vol. 42, pp. 580-589, 2011.

[Toon2012] A Toon, *Models as Make-Believe: Imagination, Fiction and Scientific Representation*, Palgrave Macmillan, 2012.

[TRAC2007] CRL and OCLC, *Trustworthy Repositories Audit & Certification, Criteria and Checklist (TRAC)*, Feb. 2007.

[USC1993] USC, *United States Court (USC) 509 U.S. 579, Dauber v. Merrell Dow Pharmaceuticals*, 1993.

[vanFraassen1980] B. C. van Fraassen, *The Scientific Image*. Oxford: Oxford University Press, 1980.

[vanFraassen2004] B. C. van Fraassen, "Scientific as representation: flouting the criteria," *Philosophy of Science*, vol 71, pp. 794-804, 2004.

[VTH2006] R. De Virgilio, R. Torlone, G.-J. Houben, "A rule-based approach to context delivery adaptation in web information systems," in *Proc. 7th IEEE/ACM Int. Conf. on Mobile Data Management*, pp. 21, 2006.

[WJH1997] A. Ward, A. Jones, A. Hopper, "A new location technique for the active office," *IEEE Personal Communications*, vol. 4, no. 5, pp 42-47, 1997.

[WL1993] T. Woo, S. Lam, "A semantic model for authentication protocols," in *Proc. of IEEE Comput. Society Symp. on Research in Security and Privacy*, May 1993, pp. 178-194.

[WRB+2005] D. Wilkinson, D. Rumsby, B. Babin, M. Merrit, J. Marsh, "The results from a Canadian national field trial comparing 1,8-diazafluoren-9-one (DFO) with ninhydrin and the sequence DFO followed by ninhydrin," Technical Report TR-03-2005, Canadian Police Research Centre: Ontario, 2005.

[WSS+2001] S. Wiesner, E. Springer, Y. Sasson, J. Almog, "Chemical development of latent fingerprints: 1,2-indanedione has come of age," *J. Forensic Sci.*, vol. 46, no. 5, pp.1082-1084, 2001.

[YPG2000] R. Yavatkar, D. Pendarakis, R. Guerin, "A framework for policy-based admission control," *RFC2753*, 2000.

[YWL+2007] Y. Ye, D. Wang, T. Li, D. Ye, "IMDS: intelligent malware detection system," in *Proc. of the 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, New York, NY, USA, 2007, pp. 1043-1047.

[Zheng2014] Wenju Zheng, "Separation of overlapping fingerprints with self-adaptive dynamic parameterization," M.Sc. thesis, Faculty of Computer Science, Otto von Guericke University Magdeburg, Magdeburg, Germany, 2014.

[ZJ2012] Q. Zhao, A. K. Jain, "Model based separation of overlapping latent fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 904-918, 2012.

[ZQZ+2007] Q. Zhang, Y. Qi, J. Zhao, D. Hou, Y. Niu, "Fuzzy privacy decision for context-aware access personal information," *Wuhan University Journal of Natural Sciences*, vol. 12, no. 5, pp. 941-945, 2007.