

---

**Verbundprojekt: Digitale Fingerprints (Digi-Dak) - Teilvorhaben:  
Datenvorverarbeitung und Datenhaltung - FKZ 13N10817**

**Autoren**

Gunter Saake  
Martin Schäler

**Ausführende Stelle**

Otto-von-Guericke-Universität Magdeburg - Fakultät für Informatik -  
Institut für Technische und Betriebliche Informationssysteme -  
Lehrstuhl Datenbanken  
Prof. Dr. Gunter Saake  
Postfach 41 20  
39016 Magdeburg

---

**Zusammenfassung**

Der Einsatz von Mustererkennungstechniken in der kriminalistischen Forensik durch das Automatisierte Fingerabdruck-Identifizierungssystem AFIS hat in den letzten Jahren zu deutlich verbesserten Fahndungserfolgen geführt. Allerdings hat die derzeitige, ausschließlich manuelle Suche nach daktyloskopischen Spuren (Fingerprints/-abdrücke) diverse Nachteile. Dazu zählt u.a. durch die physikalische oder chemische Behandlung (z.B. Bedampfung) bedingte Veränderung oder gar Zerstörung weiterer Spuren (z.B. DNA) oder der Objekt-Oberfläche. Das Ziel des Verbundprojektes „Digi-Dak“ ist die Erforschung eines rechtskonformen und automatischen Systems aus berührungsloser optischer 3D Oberflächensensortechnik, Bildverarbeitung, Biometrie, Forensik und Mustererkennungstechnik.

In dem vorliegenden Abschlussbericht werden die Kernergebnisse der Teilprojektes „Datenvorverarbeitung und Datenhaltung“ dargestellt. Insgesamt konnten wir, wie nachfolgend erläutert, durch die Erforschung neuartiger Grundlagentechniken, wie der Maßschneiderung der Datenhaltungskomponente, der Optimierung des Zugriffes auf Große Datenmengen zum Zwecke der Vorverarbeitung durch optimierte multi-dimensionale Indexstrukturen, moderne Co-Processing Ansätze und Speicherstrukturen sowie ein generalisiertes modulares Provenance-Framework zur beweisbaren Datenhaltung (chain-of-custody) und Adressierung der verfassungskonformen Gestaltung des Gesamtsystems einen signifikanten Beitrag zur Erforschung des angestrebten *rechtskonformen und automatischen Systems* leisten.

---

## Inhalt

I Kurzdarstellung.....	3
I.1 Zielstellung.....	3
I.2 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde.....	3
I.3 Planung und Verlauf des Verfahrens.....	4
I.4 Anknüpfung an den Forschungsstand zu Beginn des Vorhabens.....	7
I.5 Kooperation mit anderen Stellen.....	9
II Eingehende Darstellung.....	10
II.1 Zuwendung.....	10
Verwendung der Zuwendung.....	10
Erläuterung der Kernergebnisse der einzelnen APs.....	12
Darstellung der Ergebnisse für M3.1 .....	13
Darstellung der Ergebnisse für M3.2.....	25
Darstellung der Ergebnisse für M5.1.....	33
Darstellung der Ergebnisse für M8.2.....	33
Darstellung der Ergebnisse für M8.3.....	42
Gegenüberstellung der Ergebnisse vorgegebenen Zielen.....	42
II.2 Wesentliche Positionen des zahlenmäßigen Nachweises.....	43
II.3 Notwendigkeit und Angemessenheit der geleisteten Arbeit.....	43
II.4 Fortschreibung des Verwertungsplans.....	44
II.5 Bekanntgewordene Ergebnisse dritter Stellen.....	45
II.6 Erfolgte und geplante Veröffentlichungen.....	47
Literaturverzeichnis und sonstige Referenzen.....	50

---

# I Kurzdarstellung

## *I.1 Zielstellung*

Das Gesamtziel des Teilvorhabens „Datenvorverarbeitung und Datenhaltung“ ist die Aufbereitung und Speicherung von Daten zur Unterstützung des automatisierten Prozesses der Fingerspuren Erfassung mit dem Ziel der besseren, situationsgerechten Informationsgewinnung in präventiven und forensischen Prozessen. Dabei geht es in erster Linie um Grundlagenforschung zu der Problematik, Fingerspuren schneller zu erfassen, zu erkennen und zu analysieren. Mithilfe der gewonnen Erkenntnisse wird eine effizientere Strafverfolgung (im speziellen auch präventive Maßnahmen zur effizienteren Verfolgung) ermöglicht und ein größerer Schutz vor Kriminalität gewährleistet. Weiterhin sollen im Zuge dieses Teilvorhabens sowohl neue bzw. erweiterte Ansätze auf dem Gebiet der (Multimedia)-Datenbanken, wie z. B. Indexstrukturen oder Ähnlichkeitssuche, als auch hinsichtlich der zu klärenden rechtlichen Absicherung (z. B. Revisionsicherheit) durch interdisziplinäre Zusammenarbeit mit den anderen Teilvorhaben erforscht werden.

## *I.2 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde*

Das Vorhaben Digi-Dak ist als Verbundprojekt mit Koordination durch den Verbundpartner UNI-MD-AMSL in Person von Prof. Dr.-Ing. Jana Dittmann. Die Laufzeit des Vorhabens betrug ursprünglich drei Jahre (01.01.2010 bis 31.12.2012) , wurde jedoch konsteneutral um fünf Monate verlängert, so dass das Vorhaben am 31.05.2013 endete. Die beteiligten Verbundpartner und deren Kernexpertise bzw. -ziele sind:

- **Otto-von-Guericke-Universität Magdeburg, Fakultät für Informatik – Arbeitsgruppe Multimedia and Security (AMSL)** - Erforschung von Vorgehensmodellen zur Mustererkennungstechniken von optisch erfassten Fingerspuren zur Verbesserung und Unterstützung der kriminalistischen Forensik (Daktyloskopie). Weitere Schwerpunkte sind die Beschreibung potenzielle Präventivszenarien speziell auch für Spurenüberlagerungen und Alterungsdetektion.
- **Otto-von-Guericke-Universität Magdeburg, Fakultät für Informatik – Arbeitsgruppe Datenbanken (DB)** - Datenvorverarbeitung und Datenhaltung: Erforschung von Techniken zur Aufbereitung und Speicherung von forensischen Daten zur Unterstützung des automatisierten Prozesses der Fingerspuren Erfassung mit dem Ziel der verbesserten, situationsgerechten Informationsgewinnung in präventiven und forensischen Prozessen.
- **Fachhochschule Brandenburg - Fachbereich Informatik und Medien, Angewandte Informatik - Datensicherheit (FH-BRB)** - Erforschung und Evaluierung von Methoden der Mustererkennung und Klassifikation, Erarbeitung von Testszenarien und Aufnahme von Testdaten in ausgewählten Szenarien sowie technischem Datenschutz
- **Fries Research & Technology GmbH (FRT)** Erforschung optischer Sensortechnik und Evaluation der Sensorik im Anwendungsfall der zerstörungsfreien Akquise von latenten daktyloskopischen Spuren.
- **Landeskriminalamt Sachsen-Anhalt (LKA ST)** Sachverständigenäquivalenz: in Bezug Anwenderaspekte, Beweiswert und Beweiskraft. Definition anwenderbezogene Testszenarien und Evaluierung der Möglichkeiten der optischen Trennung von übergriffenen bzw. überlagerten Fingerspuren, Bewertung von Alterungsmodelle von Fingerspuren

- **Universität Kassel - Projektgruppe verfassungsverträgliche Technikgestaltung (PROVET)**  
Definition rechtliche Anforderungen an die einzelnen Komponenten, das Gesamtsystem und den späteren Einsatz in der Praxis.

Neben den eigentlichen Verbundpartnern waren weitere assoziierte Partner am Digi-Dak Vorhaben beteiligt, dazu gehören u.s. das BKA sowie das LKA-Niedersachsen.

### ***1.3 Planung und Verlauf des Verfahrens***

#### **Planung des Vorhabens laut Antragstext**

Die Arbeitsplanung orientiert sich an der Verbundbeschreibung, wobei UNI-MD-DB an vier Arbeitspaketen beteiligt ist. In AP3 liegt der Fokus auf der Definition der DB-Schnittstelle und der Realisierung der Referenzdatenbank. In AP5 ist die Erforschung von Datenvorverarbeitungsmethoden der Schwerpunkt. AP8 behandelt die Zusammenführung der Schnittstellendefinitionen und erforscht Möglichkeiten zur revisionssicheren Datenhaltung.

Die professoralen Eigenleistungen, die Unterstützung durch die Haushaltsmitarbeiter sowie die Unterstützung von Sekretariats- und Verwaltungsarbeiten und Technikerarbeiten sind nicht explizit angegeben. Im Detail sind im ersten Jahr eine nationale und eine europäische Teilnahme geplant. Im zweiten Jahr sind eine nationale, eine europäische und eine internationale Teilnahme geplant, während im dritten Jahr zwei europäische und zwei internationale Teilnahmen geplant sind.

#### ***Vergleich Personalplanung zu tatsächlichem Aufwand***

<b>Arbeitspaket</b>	<b>geplante Personalkapazität</b>	<b>benötigte Personalkapazität</b>
AP3 „Testszenarien, -daten und Evaluierung“	17,4 PM	17,4 PM
AP5 „Datenvorverarbeitung“	21,3 PM	<b>22,3 PM</b>
AP8 „Beweissichere Datenhaltung“	27,3 PM	27,3 PM
AP10 „Verfassungs- und Datenschutzrechtskonforme Gestaltung“	6,0 PM	6,0 PM
Summe	72,0 PM	<b>73,0 PM</b>

---

## ***Arbeitspaketbeschreibungen***

Nachfolgend werden die Inhalte der oben aufgeführten Arbeitspakete kurz dargestellt.

### *AP3 „Testszenarien, -daten und Evaluierung“:*

Im AP3 wird Uni-MD-DB auf Basis von AP3.1 als AP3.5 „Definition und Erstellung der Referenzdatenbank“ sich mit einer ersten Definition der Datenbankschnittstelle (DB-Schnittstelle) und der Datenbank für die Referenzdaten aus den Testszenarien beschäftigen. Zur Definition der DB-Schnittstelle erfolgt eine Abstimmung mit dem Partner FRT, welcher in AP1.1 die zu verarbeitenden Sensordaten spezifiziert. Darauf aufbauend soll die DB-Schnittstelle so aufgebaut werden, dass die Daten des Grob- und Feinscans der Testszenarien erfasst werden können und eine erste Informationsgewinnung auf den Rohdaten möglich ist. Neben der Datenerfassung muss die DB-Schnittstelle auch so konzipiert werden, dass eine Bereitstellung der Daten in Abstimmung mit den jeweiligen Anforderungen der Partner LKA-ST, FH-BRB und UNI-MD-AMSL möglich ist. Konkret geht es dabei um die Zusammenführung, Speicherung und Verwaltung der Referenzdaten aus AP3.2, AP3.3 und AP3.4 in verschiedenen Versionsständen. Entsprechend ist die Erweiterbarkeit der Schnittstelle ein Schwerpunkt bei deren Definition.

Neben der DB-Schnittstelle liegt der Fokus auf der Referenzdatenbank, welche die Daten aus den Testszenarien erfasst. Dafür soll ein Metadatenmodell geschaffen werden, welches die (Daten-)Fusion zwischen den Umgebungsdaten und den eigentlichen Fingerspuren realisiert. Weiterhin werden die bei den Testscans benutzten Parameter der Sensortechnik hinterlegt, welche ebenfalls als Metadaten einfließen. Für die Suche nach Fingerspuren in der Referenzdatenbank müssen die Daten entsprechend einem abstrakten Fingerspurenmodell aufbereitet werden. Des Weiteren wird eine Evaluierung des Digi-Dak-Software-Demonstrators unter Nutzung verschiedener Referenzdatenversionsstände durchgeführt. Den Abschluss des Arbeitspaketes bildet dann eine Gesamtevaluierung und ein Bericht zu den Evaluierungsergebnissen aus der Gesamtevaluierung und den Teilevaluierungen mit den Partnern LKA-ST, FH-BRB und Uni-MD-AMSL.

### *AP5 „Datenvorverarbeitung“:*

In AP5 wird UNI-MD-DB als AP5.1 „Datenvorverarbeitungsmethoden“ sich mit der Definition und Implementierung von Datenvorverarbeitungsmethoden unter Nutzung der in AP3.5 definierten DB-Schnittstelle befassen. Dazu wird es zunächst theoretische Überlegungen zu den in AP3.1 definierten Testszenarien geben, wobei ein besonderes Augenmerk auf die unterschiedliche Handhabung von Grob- und Feinscandaten gelegt wird. Der eigentliche Schwerpunkt liegt dann in den darauf aufbauenden, zu erforschenden und zu implementierenden Datenvorverarbeitungsmethoden. Eine zentrale Forschungsfrage ist dabei, welche möglichen merkmalsbasierten Repräsentationsformen jeweils für Grob- und Feinscandaten möglich sind, d. h. wie können diese Daten effizient und nachvollziehbar in der Datenbank gespeichert werden. Die dafür zu erforschenden Vorverarbeitungsmethoden müssen dabei die Daten so aufbereiten, dass die in AP6 und AP7 erforschten Methoden zur Klassifizierung und Mustererkennung darauf angewendet werden können. Deshalb erfolgt in diesem Zuge eine Abstimmung mit FH-BRB bezüglich der Identifizierung von inhaltstragenden Eigenschaften der Testszenarien (bzw. der daraus resultierenden Daten für Grob- und Feinscan). Ein weiterer Schwerpunkt ist in diesem Zusammenhang die Erforschung von effizienten Datenstrukturen zur Unterstützung der Ergebnisberechnung von Ähnlichkeitsanfragen unter

---

Berücksichtigung der speziellen Eigenschaften von Grob- und Feinscandaten. Der Fokus liegt hier insbesondere auf Verfahren für hoch-dimensionale Indexstrukturen und Datenfusion unter Berücksichtigung der Datenqualität. Basierend auf den vorhergehenden Forschungsergebnissen der Datenvorverarbeitung werden dann, innerhalb einer Testumgebung, versuchsweise vordefinierte Anfragen an die aufbereiteten Daten gestellt. Durch diese rudimentären Versuchsanfragen sollen erste Erkenntnisse über die Anwendbarkeit der Repräsentationsformen der Daten für Mustererkennung und Klassifizierung gewonnen werden. Dabei wird eng mit dem Partner FH-BRB zusammengearbeitet, um möglichst die optimalen Repräsentationsformen für Grob- und Feinscandaten im Hinblick auf die in AP6 und AP7 zu erforschenden Methoden der Mustererkennung und Klassifizierung zu erzielen. Abschließend werden die erzielten Ergebnisse (d. h. die Methoden) mit den Partnern UNI-MD-AMSL, FH-BRB und FRT abgestimmt und zusammengeführt. Im Anschluss werden die Evaluierungsergebnisse aus AP3 ausgewertet und die verfolgten Ansätze und Methoden entsprechend angepasst. Die implementierten Vorverarbeitungsmethoden werden dann in AP6, AP7, AP8 und AP11 in zwei Softwarebibliotheksversionen genutzt.

#### *AP8 „Beweissichere Datenhaltung“:*

In AP8 wird UNI-MD-DB als AP8.1 „Zusammenfassung Schnittstellendefinition und Revisionsicherheit“ sich mit der Zusammenfassung der Schnittstellendefinitionen aus AP1, AP3, AP4, AP5, AP6 und AP7 für AP11 und der Definition und Spezifikation der Datenfusion befassen. Bei der Zusammenfassung der DB-Schnittstellendefinitionen liegt der Schwerpunkt auf der Integration der Anforderungen der einzelnen Partner an die DB-Schnittstelle. Bei der Datenfusion liegt der Fokus auf der Spezifikation von (Daten-)Modellen, die die Erstellung eines Datenstammbaums unter Berücksichtigung von forensischen Vorgehensweisen, wie zum Beispiel der Definition von Zugriffsrechten und dem Management von Schlüsseln und Rollen sowie Rechten, ermöglichen. Insbesondere soll hierbei auch die Gerichtsverwertbarkeit und Revisionsicherheit der fusionierten Daten durch Zusammenarbeit mit dem Partner PROVET sichergestellt werden sowie durch Berücksichtigung der Anforderungen aus AP2, AP8.3 und AP10. In diesem Kontext sollen auch verschiedene Aspekte der IT-Sicherheit sowohl betrachtet als auch deren Einfluss auf die Beweissicherheit und notwendige Gegenmaßnahmen bewertet werden. Basierend auf den fusionierten Daten bzw. deren Repräsentation sollen dann Grundlagen erarbeitet werden, die eine Anbindung an Vergleichsfingerabdruck-Datenbestände wie AFIS ermöglichen.

In einem weiteren AP8.2 „Implementierung der Infrastruktur“ befasst sich UNI-MD-DB, in Abstimmung mit METOP, mit der Umsetzung der gesamten Datenbankinfrastruktur für den Digi-Dak-Demonstrator. Dabei sollen auch die in AP8.1 erarbeiteten Grundlagen zur Anbindung an Vergleichsfingerabdruck-Datenbestände realisiert werden. Neben der Umsetzung in zwei verschiedenen Versionsständen sollen mithilfe der Implementierung auch die Evaluierungsergebnisse von AP3 ausgewertet werden.

#### *AP10 „Verfassungs- und Datenschutzrechtskonforme Gestaltung“:*

In AP10 wird UNI-MD-DB als AP10.2 „Technische Bewertung der Digi-Dak-Infrastruktur“ sich mit der Bewertung der Datenbankaspekte in der Digi-Dak-Infrastruktur als auch bei der Datenvorverarbeitung beschäftigen. Bei der Bewertung der Infrastruktur wird basierend auf den Ergebnissen von AP8 eine Bewertung hinsichtlich Performance und Effizienz der DB-Komponenten durchgeführt. Ein besonderes Augenmerk liegt bei der Bewertung außerdem auf der

datenschutzrechtlich konformen Gestaltung, d. h., inwieweit wurde dem Datenschutz bei der Infrastruktur (aus DB-Sicht) entsprochen. Dies erfolgt in Zusammenarbeit und Abstimmung mit dem Partner PROVET. In einem ähnlichen Rahmen wie die Digi-Dak-Infrastruktur sollen dann auch die Datenvorverarbeitungsmethoden aus AP5 sowohl aus technischer als auch rechtlicher Sicht (in Abstimmung mit dem Partner PROVET) bewertet werden.

### ***Planung von Zwischenergebnissen***

Die Planung der einzelnen Zwischenergebnisse bei denen wir beteiligt sind, ist in der nachfolgenden Tabelle aufgelistet.

<b>AP</b>	<b>Name des Meilensteins</b>	<b>Grad der Verantwortlichkeit</b>
AP 3	<b>M3.1</b> Schnittstellendefinition Datenbank	verantwortlich
	<b>M3.2</b> Zusammenführung und Einpflegen der Referenzdaten	verantwortlich
AP 5	<b>M5.1</b> Schnittstellendefinition Datenvorverarbeitung	verantwortlich
AP 8	<b>M8.2</b> Schnittstellenzusammenfassung	verantwortlich
	<b>M8.3</b> Datenstammbaum	verantwortlich
	<b>M8.4</b> Definition Nutzerschnittstelle	Zuarbeit
	<b>M8.5</b> Bericht zur Anbindung an Vergleichssysteme	Zuarbeit
<b>Teilprojekt-übergreifende Meilensteine</b>		
<b>PM</b>	Projekt-Meilenstein	Zuarbeit
<b>M1</b>	Software-Demonstrator Version 1	Zuarbeit
<b>M2</b>	Erster Evaluierungsbericht	Zuarbeit
<b>M3</b>	Software-Demonstrator Version 2	Zuarbeit
<b>M4</b>	Zweiter Evaluierungsbericht	verantwortlich
<b>M5</b>	Abgestimmte Endfassung Evaluierungsbericht	Zuarbeit

### ***I.4 Anknüpfung an den Forschungsstand zu Beginn des Vorhabens***

Im Folgenden werden bestehende Arbeiten aufgeführt aus den Gebieten Multimediadatenbanken und beweissichere Datenhaltung, welche die Grundlage für die in diesem Teilvorhaben geplanten Forschungsarbeiten bilden. Die Darstellung entspricht dabei der des Originalantrags, da zwischen Antragstellung und Beginn des Vorhabens keine wesentlichen neuen Forschungsergebnisse bekannt wurden.



Da das Gebiet der Multimediatdatenbanken sehr viele forschungsrelevante Themen umfasst, werden hier nur die aufgeführt, die für die im Teilvorhaben behandelten Themen Datenvorverarbeitung und Datenhaltung unmittelbar von Bedeutung sind. Das sind im Einzelnen Merkmalstransformation, hochdimensionale Indexstrukturen und Anfragebehandlung.

Im Rahmen der Detektion und Auswertung von Fingerspuren im Gesamtverbund geht es zunächst um die situationsgerechte Informationsgewinnung aus den unstrukturierten Rohdaten des Grobscans (im Speziellen 3D-Daten) als auch aus den Daten des Feinscans. Eine wesentliche Rolle spielt hierbei die Datenvorverarbeitung, bei der es gilt, merkmalsbasierte Repräsentationsformen (Feature) für die Scandaten zu erforschen.

Ein Schwerpunkt im Teilvorhaben ist deshalb die Transformation der Rohdaten zur Generierung von inhaltstragenden Daten, d. h. aus den Rohdaten sollen inhaltsbezogene Metadaten gewonnen werden, die die ursprünglichen Daten repräsentieren. Da die einzelnen Verfahren zur Transformation sehr weit verbreitet sind, z. B. in der Signal- und Bildverarbeitung, gibt es sehr viele Vorarbeiten auf diesem Gebiet, allen voran zum wohl populärsten Verfahren, der Diskreten Fourier-Transformation (DFT). Gute Einführungen zur ebenfalls weit verbreiteten Diskreten Wavelet-Transformation (DWT) sind in [Bla98, SDS95a, SDS95b] zu finden. Informationen zu weiteren, relevanten Verfahren, wie der Karhunen-Loeve-Transformation (KLT) oder des Latent Semantic Indexing (LSI) können in [BR99, Dum93, GKL01, Jac03] gefunden werden. Neben diesen überwiegend aus der Signalverarbeitung stammenden Verfahren, gibt es auch neuere Ansätze, die explizit zur Auswahl von repräsentativen Eigenschaften dienen, wie z. B. in [EP+03, DK+09]. Trotz der Vielzahl an Vorarbeiten zu diesem Thema ergeben sich für die im Verbundprojekt behandelten Szenarien Problemstellungen, die bisher noch nicht beantwortet wurden, z. B. Auswahl geeigneter Verfahren für Grob- und Feinscan bzw. für Oberflächensensordaten im Allgemeinen, und deshalb auch Gegenstand in diesem Teilvorhaben sind.

Für eine effiziente und qualitative Ergebnisberechnung müssen Datenstrukturen existieren, die die entsprechenden Ähnlichkeitsanfragen unterstützen. Eine zentrale Rolle spielen hierbei hochdimensionale Indexstrukturen. Einen Überblick und Vergleich zu weit verbreiteten und verwendeten Indexstrukturen gibt es in [BBK01, Cas02, CB02, FSZ03, BK+07]. Eine in diesem Kontext häufig verwendete Anfrageart, die auch für das in diesem Teilvorhaben behandelte Thema relevant ist und deshalb von der zu erforschenden Referenzdatenbank unterstützt werden muss, ist der nearest neighbour algorithm, zu dem es zahlreiche Vorarbeiten, unter anderem in [Bal04, BB+97, GO04, KK+09] gibt. Das dabei auftretende Problem des „Fluchs der hohen Dimensionen“ wird in [BG+99, Sch01, WSB98] untersucht.

Darüber hinaus ist es erforderlich, die Anfragebehandlung, d. h. wie wird die Ähnlichkeitsanfrage übersetzt, damit eine entsprechende Suche in der DB möglich ist, an den Kontext, in diesem Fall 2D/3D Sensordaten, anzupassen. Wesentliche Punkte, die in diesem Teilvorhaben erforscht werden, sind allgemeine Konzepte für das Anfragesystem, Datenbankmodelle und

Anfragesprachen, wobei letzteres auch die Nutzerschnittstelle mit einbezieht. Vorarbeiten zu den allgemeinen Konzepten wie Struktur- oder Attributsbedingungen oder räumliche/zeitliche Beziehungen im Kontext einer Ähnlichkeitsanfrage sind u. a. in [KM03, All83, BR99] zu finden. Einen Überblick über Datenbankmodelle, welche im auch im Multimediatkontext zur Anwendung kommen, sind etwa in [Gep02, HS00] zu finden. Auf dem Gebiet der Anfragesprachen gibt es innerhalb des SQL-Standards einen eigenen Standard, der Basisdatentypen für Mediatypen festlegt, wie z. B. in [ME01, KMW03, Sto02] erläutert wird. Ergänzend wird in [Dun03] auf die Datentypmodellierung von Mediatypen eingegangen. Auch zu konkreten Anfragesprachen für Multimediatdatenbanken gibt es einige



---

Vorarbeiten, die als Grundlage für das Teilvorhaben genutzt werden können. Eine generelle Übersicht über existierende Sprachen geben die Arbeiten [Sch04, Pas99]. In [Sch04] wird zudem die konkrete Anfragesprache WS-QBE vorgestellt, die auf QBE [Zlo75] basiert und die eine Wichtung von Anfragetermen erlaubt. Eine Erweiterung von QBE um nützliche Eigenschaften wie z. B. parametrisierbare Unteranfragen wird in [SLG97] vorgestellt. Die auf QBE basierenden Anfragesprachen sind im Kontext des Verbundprojektes von besonderem Interesse, da sie die Nutzerinteraktion mit der Datenbankanfrage vereinfachen.

### ***Beweissichere Datenhaltung (Chain-of-custody)***

Im Zuge der beweissicheren Datenhaltung (sichere Datenspeicherung) im Gesamtszenario Digi-Dak geht es darum, in einer ganzheitlichen Vorgehensweise sowohl die Authentizität und Herkunft der Daten (und deren Zusammenhänge) einwandfrei nachzuweisen als auch die Möglichkeit der Manipulation von Datenstammbäumen durch Sicherung der Integrität auszuschließen bzw. nachvollziehbar zu machen. In [GD07, IW09, Tan04] gibt es Informationen zu bestehenden Ansätzen, zu deren vorhandenen bzw. nicht vorhandenen Merkmalen als auch zu Datenabstammungsmodellen und deren Kategorisierung. Zusätzlich muss auch die IT-Sicherheit zur Vermeidung oder Erkennung von Manipulationen betrachtet werden. In [HR+09] werden Manipulationen zwar betrachtet und durch kryptologische Verfahren gesichert, allerdings gilt der Ansatz nur für Datenstücke, die ausschließlich auf den lokalen Rechnern der Nutzer gespeichert werden. In [ZC+09] werden im Vergleich zu [HR+09] auch komplexe Datentypen betrachtet. Schließlich beschreibt [Cha09] das generelle Vorgehen zur Integration von Datenstammbäumen in Datenbanksystemen. Des Weiteren könne auch Zugriffsstrategien wie existierende Zugriffsmatrizen, Zugriffslisten oder Rollen-basierte Ansätze helfen. Insgesamt bieten die Vorarbeiten eine gute Grundlage, sind jedoch für die im Verbundprojekt vorliegenden Szenarien unzureichend oder müssen spezifisch angepasst werden, da weder die hohen Anforderungen der Gerichtsverwertbarkeit/Revisionssicherheit noch komplexe, mehrdimensionale Daten und komplexe, nachvollziehbare und dynamisch an das Szenario angepasste Zugriffsstrategien und Managementstrategien (wie Schlüssel- oder Rechteverwaltung) betrachtet wurden.

### ***1.5 Kooperation mit anderen Stellen***

Neben der Intensivierung der Kooperation mit den Projektpartnern, hier seien vor allem METOP, AMSL und FH-BRB zu nennen, konnte, vor allem gegen Ende des Vorhabens auch weitere Kontakte zu Forschungseinrichtungen geknüpft werden, die keine Verbundpartner sind. Hervorzuheben ist hier die Kooperation mit der TU Braunschweig und der Uni Passau, die in der Kernpublikationen des AP 5 mündeten [VLDB2014].

---

## II Eingehende Darstellung

### ***II.1 Verwendung der Zuwendung, erzielte Ergebnisse und Gegenüberstellung der Ergebnisse mit den Vorhabenszielen***

#### **Verwendung der Zuwendung**

Die folgende Darstellung schildert der Verwendung der Zuwendung im vergleicht sie mit der Planung des Vorhabens aus Kapitel I.3. Sie entspricht der Darstellung der einzelnen Zwischenberichte.

#### ***Jahr 1 - 01.01.2010 bis 31.12.2010***

Aufgrund der Verzögerung in notwendigen Vorarbeiten verschob sich die Arbeits- und Zeitplanung um ca. 6 Monate nach hinten. Evident wird dies beispielsweise bei der Einreichung der internen Meilensteine 3.1 und 5.1, für die UNI-MD-DB verantwortlich war. Entgegen der ursprünglichen Planung diese zum 30.06.2010 anzufertigen, konnten sie erst zum 30.12.2010 eingereicht werden. Eine wesentliche Ursache hierfür waren die verspätete Auslieferung der Messtechnik und sich daraus ergebende Verzögerungen bei notwendigen Zu- bzw. Vorarbeiten.

Die Vorziehung von anderen Teilaufgaben bzw. Arbeitspaketen war nur in begrenztem Maße möglich, so dass der hieraus erzielte Zeitgewinn den zuvor beschriebenen Rückstand nicht ausgleichen kann. Die Verzögerung der einzelnen APs wird im folgenden genauer begründet. Dabei wird von einer Verzögerung von 6 Monaten ausgegangen, um die jedes AP jeweils verlängert werden muss.

#### ***AP 3 Testszenarien, -daten und Evaluierung:***

Das Arbeitspaket AP3 basiert auf der in AP1 vorgenommenen Sensorauswahl und Sensorparameterbestimmung, um „Testszenarien & -daten“ zu definieren und anhand der ausgewählten Oberflächensensoren und deren Parameter für diese Szenarien Referenzdaten zur weiteren Evaluierung zu erheben. Die Verzögerung ergibt sich somit aufgrund der verspäteten Erstellung der M1.1 und M1.2. Da die Arbeiten im AP3 aufeinander aufbauen, ergibt sich auch eine Verspätung der M3.1 in 2010, sowie M3.2 in 2011.

#### ***AP 5 Datenvorverarbeitung:***

Durch die Verzögerungen in AP3 verspätete sich die Sammlung einer entsprechenden Anzahl an Datensätzen (sowohl Nutz- als auch Meta-Daten und deren Fusion) zur Evaluierung der entsprechenden Vorverarbeitungsmethoden. Zusätzlich sind die Vorverarbeitungsschritte (inkl. deren Zugang via wohldefinierter Schnittstellen) abhängig von den in AP3 definierten Testszenarien. Ein weiterer Grund für die Verzögerung von AP5 ergibt sich aufgrund der verspäteten Erstellung des M1.1, wodurch eine Verzögerung des M5.1 in 2010 eingetreten ist.

#### ***AP 8 Beweissichere Datenhaltung:***

Der Entwurf und die prototypische Implementierung einer sicheren Infrastruktur erfordert es, dass das System selbst (AP3 & AP5) in seiner Funktionsweise vollständig beschrieben wird. Zusätzlich müssen auf der Basis dieser Beschreibung Angriffsszenarien definiert werden, denen technische und organisatorische Maßnahmen entgegengesetzt werden. Somit ist die Verzögerung des AP8 direkt bedingt durch Verschiebung in AP3 und AP5, aber auch durch die verspätete Erstellung von M1.1 zu begründen. Da AP2 ebenfalls verzögert ist, konnten die Ergebnisse nur verspätet konkretisiert werden.

#### ***AP 10 Verfassungs- und datenschutzrechtskonforme Gestaltung:***

Da sich die Bearbeitung von AP6, AP7 und AP8 in 2010 und 2011 verspäten, liegen die Ergebnisse für die Begleitung, Bewertung und Gestaltung in und für AP10 erst verzögert vor, wodurch eine Verzögerung des

---

M10.1 in 2011 und eine Verschiebung des M10.2 von 2011 auf 2012 notwendig ist. Die Zuarbeiten von AP10 zur Demonstratorbewertung und zur Einzelbewertungen der Nutzerschnittstelle für AP11, den daraus abgeleiteten Handlungsanleitungen AP12 und der potentiell daraus entstehenden Verwertungsaspekten für AP13 werden sich entsprechend den Verzögerungen anpassen und entsprechende zeitliche Verschiebungen der Arbeiten werden vorgenommen.

*Kostenplan:*

Durch Schwierigkeiten bei der Suche nach fachlich qualifiziertem Personal konnte erst verspätet mit der Einstellung von Personal und studentischen Hilfskräften begonnen werden. Dadurch konnten noch nicht alle Mittel wie geplant verbraucht werden. In gleichem Maße waren davon auch geplante Veröffentlichungen und zugehörige Dienstreisen betroffen, da durch die oben erwähnten Verzögerungen (z.B. aufgrund verspätet eingetreffener Technik) auch die Ergebnisse nicht rechtzeitig zur Publikation vorlagen. Hier wird versucht, die Ergebnisse später auf fachlich äquivalenten Konferenzen zu veröffentlichen.

***Jahr 2 - 01.01.2011 bis 31.12.2011***

*AP3: Referenzdatenbank*

Die Integration der feature-basierten Repräsentation aus Teilen der Referenzdaten in das Index-Evaluations-Framework [B] zeigt, dass die Verteilung, vor allem bei unseren stark geclusterten Daten, der Daten großen Einfluss auf die Performanz (Laufzeit, und Präzision) der genutzten Indexstrukturen hat [4]. Zusätzlich wurde weiter an der Verbesserung der Maßschneiderung des Zusammenspiels aus Datenbank und Anwendung gearbeitet [5] und die daraus gewonnen Erkenntnisse flossen in die Gestaltung der Schnittstellen ein. Die Notwendigkeit und ein erster Entwurf eines generellen Schemas für die Spurenerfassung, mit dem sich z.B. daktyloskopische Spuren mit weiteren Beweisen kombinieren lassen, wurde in [2] beschrieben.

*AP5: Datenvorverarbeitungsmethoden*

Im Zentrum der Forschungsarbeit zum AP 5, steht die rigorose Evaluierung von Indexverfahren unter Zuhilfenahme der Erweiterung des Frameworks [4, B]. Langfristiges Ziel ist es für ein Anwendungsszenario bestehend aus dem 4-Tupel (DatenSatzAnzahl, Verteilung, Dimensionalität, AnfrageParameter) geeignete Indexierungsverfahren für die effizientere Abarbeitung der Suchanfrage im Vergleich zum trivialen Ansatz zu identifizieren. Unsere Ergebnisse zeigen, dass je nach Anwendungsszenario die Bearbeitungszeit teilweise um den Faktor 1000 beschleunigt werden kann, jedoch eine unpassende Auswahl des Verfahrens auch zu signifikanten Performanzeinbußen führt [4]. Zusätzlich Aspekte, wie in-memory vs. Festplattenzugriff [4], Transaktionsverwaltung [C] und unterschiedliche Distanzmetriken [D] werden ebenfalls im Rahmen von studentischen Projekten und Abschlussarbeiten erforscht.

*AP8: Beweissichere Datenhaltung*

Kürzliche Fortschritte auf dem Gebiet der Provenance (z. B. Provenance for aggregat queries, und die Beziehung zur Kausalität) erlauben es uns ein hierarchisches Framework bzgl. Provenance [2, A] vorzuschlagen, welches für den Datenstammbaum angewendet werden kann. Im einzelnen zeigen wir, dass die verschiedenen Ansätze im Bereich Provenance durch zwei Dimensionen (Fragmentationsgrad und Vertrauen) in ein globales (erweiterbares) Framework fassen lassen. Die einzelnen Ausprägungen der Dimensionen beziehen sich auf die Existenz von Fragmenten (gerichteten azyklischen Graphen) auf unterschiedlichen Granularitätsstufen und das Vertrauen in diese Information. Die einzelnen Stufen sind hierbei durch Abstraktionsschichten mit unterschiedlichem Anwendungsfokus miteinander verbunden und erlauben eine schrittweise Verfeinerung. Zusätzlich schließen Erweiterungen des

---

Frameworks die Grenze zu verwandten Themenfeldern der Kausalität (Nicht-Antworten) [3]. Weitere Herausforderungen wie die Speicherung/Propagierung von Provenance-Informationen über Systemgrenzen für die spezielle Daten wurden in [1] beschrieben.

### ***Jahr 3 - 01.01.2012 bis 31.12.2012***

Unsere Kernaufgaben im dritten Projektjahr bestanden im Transfer der Ergebnisse der Grundlagenforschung und der Abstraktion von Detaillösungen aus den beiden vorangegangenen Jahren. Dies beinhaltet zum Einen die Erforschung des Einsatzes unserer Techniken und Ansätze unter Berücksichtigung der speziellen Anforderungsstruktur des Digi-Dak Projektes und zum Anderen die Generalisierung unserer Ansätze in weiteren Anwendungsdomänen zur Vermeidung von Insellösungen.

So wurden z.B. die Ergebnisse der Maßschneidung des Zusammenspiel des Schemas der Referenzdatenbank mit den zugehörigen Anwendungen [2] aus AP 3 verwendet um für AP 8 die minimal-invasive Integration des Provenance-Belangs [1] in bestehende Infrastrukturen vorzuschlagen [11,I,A]. Analog dazu wurden in AP 5 die klassischen Einflussfaktoren [12] beim Suchen in hoch-dimensionalen Datenräumen evaluiert [9a-9d]. Zusätzlich erfolgte eine Berücksichtigung alternativer Speicherformen [3,10], sowie moderner Co-Processing Ansätze [14,15,F]. Der Transfer von Punktdaten auf Zeichenketten erfolgte in [B,H]. Zusätzlich wurden dabei für AP 8 der Sicherheitsbelang betrachtet. Im Fokus standen Sicherheitsaspekte der Gesamtinfrastruktur [9e,13] bzw. der Generalisierbarkeit aufgrund bestehender Standards [6], sowie der forensische Rekonstruierbarkeit, vor allem Sicht des Datenschutzes [5,8].

### ***Laufzeitverlängerung 01.01.2013 bis 31.05.2013***

In der finalen Phase wurden die zuvor getätigten Arbeiten und die daraus resultierten Forschungsergebnisse. So ist beispielsweise die Kernpublikation des AP 5 [VLDB2014], die bei einer der beiden führenden Datenbankkonferenzen (mit angeschlossenen Journal) platziert werden konnte, in dieser Zeit entstanden. Darüber hinaus wurden in dieser Zeit weitere Publikationen, wie [DBSpek] und [DKE2014] verfasst, bzw. begonnen. Abschließend wurden unsere Zuarbeiten zum M5 in dieser Zeit konkretisiert.

## **Erläuterung der Kernergebnisse der einzelnen APs**

Auf den folgenden Seiten erläutern wir die Kernergebnisse zu den Meilensteinen bei denen denen wir **verantwortlich** waren. Wir unterschieden hierbei zwei Arten von Meilensteinen:

- Grundlegende Meilensteine, die Voraussetzungen für weitergehende Grundlagenforschung in sich tragen; dies sind **M3.1, M3.2 und M8.2**. Die Ergebnisse dieser Meilensteine sind nicht in Gänze bereits veröffentlicht, sonder werden nachfolgend eingehend erläutert.
- Grundlagen-Forschungsmeilensteine enthalten bereits publizierte Forschungsergebnisse. Zu jedem Meilenstein existieren eine oder mehrere Publikation(en), die die Resultate zusammenfasst. Eine vollständige Zuordnung der einzelnen Publikationen erfolgt in II.6. Diese APs sind **M5.1 und M8.3 sowie M4**.

---

## ***AP 3 Testszzenarien, -daten und Evaluierung***

### ***Kernerthemengebiete des AP 3:***

**DB Schnittstellen** - Die Definition der DB Schnittstelle war der erste wichtige Schritt zum Aufbau der Referenzdatenbank. Der wissenschaftliche Neuwert liegt in der Verwendung einer variablen auf den Anwendungsfall zugeschnittenen Schnittstellendefinition in Form von Features und der empirischen Evaluierung der praktischen Einsetzbarkeit dieses neuen Ansatzes.

**Erstellung Referenzdatenbank** - Die Sammlung einer ausreichend großen Datenbasis, sowohl vom Rohdaten als auch extrahierte Merkmale von verschiedenen Anwendungsfällen (z.B. Oberflächen Detektion, Fingerabdrucks-Lokalisation oder Alterungsreihen) bildete eine wertvolle Grundlage für die anschließende Optimierung der Suchverfahren. Die Analyse der Datensätze erlaubt es uns von den eigentlichen Daten zu abstrahieren und synthetische Datensätze zu generieren, die beispielsweise über dieselben bzw. ähnliche Eigenschaften (Dimensionalität, Stochastische Verteilung etc.) verfügen, jedoch beispielsweise mit unterschiedlichen Datensatzanzahlen generiert werden können, um somit die Skalierung unserer Suchverfahren bzw. -optimierungen mit real zu erwartenden Datenmengen zu erforschen.

### ***Publikation des Kernergebnisses aus AP 3:***

Wie bereits bei den Kernthemen erwähnt ist der wesentliche wissenschaftliche Neuwert der neuartige Ansatz für zugeschnittenen Schnittstellendefinition in Form von Features und der empirischen Evaluierung der praktischen Einsetzbarkeit dieses neuen Ansatzes. Publiziert wurde das Ergebnis im Rahmen der CAiSE-Konferenz 2012 [CAISE2012].

Die CAiSE ist seit langem eine der am höchsten gerankten wissenschaftlichen Konferenzen im Bereich Datenbanken & Informationssysteme mit durchschnittlichen Annahmehquoten von unter 15%. Somit kann die Publikation als außerordentlicher Erfolg des Teilprojektes angesehen werden.

### ***Eingehende Erläuterung zum Meilenstein des AP 3***

Nachfolgend erfolgt eine Darstellung der Ergebnisse bezogen auf die einzelnen Meilensteine des AP 3 (M3.1 & M3.2), für die wir als Verbundpartner verantwortlich waren. Im Gegensatz zu den Meilenstein M5.1 oder M8.3, sind die Ergebnisse von M3.1 und M3.2 nicht in eigene Publikationen zusammengefasst, die die Gesamttragweite der Ergebnisse beschreiben. Daher werden wir nachfolgend diese im Detail erläutern. Publikationsergebnisse liegen hier zu grundlegenden neuartigen Techniken wie der Nutzung variabler Datenbankschemata vor [BTW11, CAISE12, Sch12]. Diese Techniken wurden dann verwendet um die eigentliche Schnittstelle zu schaffen.

### **Darstellung der Ergebnisse für M3.1 DB-Schnittstelle zum AP 3**

Der Meilenstein 3.1 DB-Schnittstelle zum AP 3 umfasst die Modellierung der Referenzdatenbank (Referenz-DB) und der zugehörigen Datenbankschnittstelle(n) (DB-Schnittstellen), welche für die Datenerhebung als auch für den Datenzugriff, unter Berücksichtigung der IT-Sicherheit, notwendig sind. Die Referenzdatenbank nimmt eine zentrale Rolle sowohl in der Digi-Dak-Infrastruktur als auch im zu entwickelnden Demonstrator ein.

---

Diese Rolle soll im Folgenden kurz illustriert werden. Weiterhin muss die Modellierung der DB in der Lage sein, sämtliche Konzepte der von den Verbundpartnern definierten Testszenarien abzubilden. Zusätzlich müssen DB und Schnittstellen flexibel auf mögliche Anpassungen der Testszenarien reagieren können, die sich aufgrund von neuen Erkenntnissen im Projektverlauf zwangsläufig ergeben werden.

In Abbildung 1 ist die Digi-Dak-Infrastruktur schematisch dargestellt. Im Zentrum steht die Referenzdatenbank (*Fingerprint Verification DB, FiVe DB*), welche von weiteren Komponenten der Infrastruktur umschlossen wird. Weiterhin deuten die Pfeile zwischen den Komponenten und der Datenbank die Kommunikation bzw. den Datenfluss über die entsprechende DB-Schnittstelle an. Dabei kennzeichnen die Pfeile in DB-Richtung mögliche Schnittstellen zur Datenerhebung (z.B., Sensorik → DB) bzw. zur (Vor)Verarbeitung der Daten (z.B. Datenvorverarbeitung → DB). Umgekehrt kennzeichnen Pfeile in Richtung der weiteren Komponenten mögliche Schnittstellen des Datenzugriffs (z.B. DB → AFIS).

Ein wichtiger Punkt, der für alle DB-Schnittstellen beachtet werden muss, ist die Gewährleistung der Authentizität, Integrität und Vertraulichkeit der Daten. Nur so kann eine lückenlose Beweiskette (*engl: chain-of-custody*) sicher gestellt werden, welche unabdingbar ist für die Verwendung der Ergebnisse beispielsweise vor Gericht. Daher müssen weiterführende Arbeitspakete wie z.B. *AP 8 beweissichere Datenhaltung* in die Modellierung der DB und der zugehörigen Schnittstelle als integraler Bestandteil von Anfang mit einbezogen werden. Dieses Vorgehen stellt sicher, dass die Mechanismen zur Gewährleistung von Authentizität, Integrität und Vertraulichkeit basierend auf der Risikoanalyse dem Schutzbedarf der Daten entsprechen.

Im weiteren Verlauf dieses Berichtes werden wir die Realisierung der DB-Schnittstelle näher beschreiben. Dabei wird zunächst in Abschnitt 3.2 auf die notwendigen Voraussetzungen, wie zum Beispiel die Definition der Testszenarien, eingegangen. Anschließend werden in Abschnitt 3.3 die zusammen mit den beteiligten Verbundpartnern ausgearbeiteten Anforderungen an die DB-Schnittstelle erläutert. Die daraus resultierende Modellierung und Implementierung der DB-

Schnittstelle ist Gegenstand von Abschnitt 3.4. Abschließend geben wir einen Überblick über den aktuellen Stand/Funktionsumfang der DB-Schnittstelle, zu weiteren möglichen Anbindungen und zu weiterführenden Anforderungen, die zukünftig noch in der DB-Schnittstelle abgebildet werden müssen.



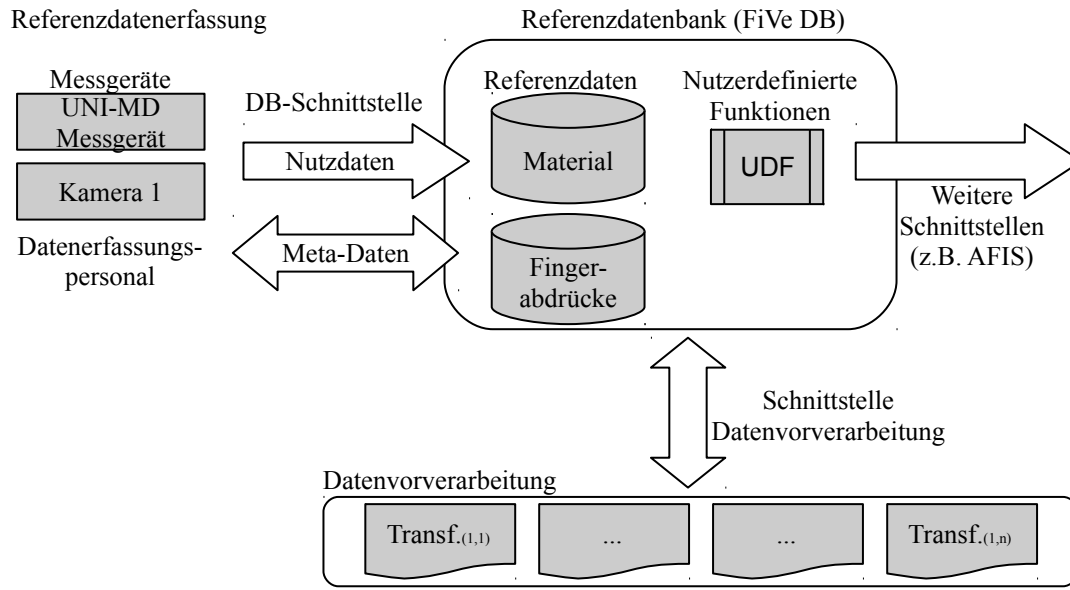


Abbildung 1: Digi-Dak-Infrastruktur

### M3.1 Notwendige Vorarbeiten

Die Definition der DB-Schnittstelle, genauer der Schnittstelle Datenerhebung-Referenzdatenbank, erfordert es, dass die beiden über die Schnittstelle verbundenen Systeme zuvor eindeutig beschrieben wurden. Die Datenerhebung wird einerseits durch die Testszenarien von LKA-ST (AP 3.2.1), UNI-MD-AMSL (AP 3.3.1) und FH-BRB (AP 3.4.1), sowie durch den pauschalisierten Tatort (LKA-ST) beschrieben. Diese informellen Beschreibungen müssen (soweit möglich) vereinheitlicht werden, um daraus den Entwurf des Referenzdatendankschemas zu erstellen, die das modellierte System (Testszenarien inklusive Störeinflüssen) möglichst genau widerspiegelt. Die Modellierung des DB-Schemas dient als Ausgangspunkt zur Definition der Schnittstelle. Die Herausforderung liegt hierbei vor allem darin, die teilweise entgegengesetzten Ziele a) exakte Abbildung des modellierten Systems und b) flexible Anpassbarkeit des DB-Schemas und -schnittstelle an neue Gegebenheiten (Schemaevolution), miteinander zu vereinen.

Durch die nötigen Vorarbeiten muss vor allem definiert werden, welche Objekte in der Modellierung des DB-Schemas aufgenommen werden müssen und wie diese miteinander in Verbindung stehen. Objekte und Konzepte, die hier nicht erfasst sind, nachträglich in ein DB-Schema und in bereits darin enthaltene Daten zu integrieren, ist mit massiven Aufwand verbunden bzw. teilweise nicht möglich. Daher ist dies ein entscheidender Punkt. Für Objekte, die über die Schnittstelle in der DB gespeichert werden sollen, müssen die folgenden Punkte definiert sein:

1. *Semantik des Objekts*: Welches Konzept aus dem modellierten System verbirgt sich hinter dem Objekt? Dazu gehören physisch greifbare Objekte, wie untersuchte Materialien, das (digitale) Abbild eines Fingerabdrucks bzw. Materials und weitere Konzepte wie die Möglichkeit Proben zu behandeln indem sie beispielsweise beruht oder bedampft werden.
2. *Eigenschaften des Objekts*: In der DB wird ein Abbild des modellierten Systems gesichert, welches sich aus den gespeicherten Eigenschaften (Attribute) des Objekts ergeben. Zusätzlich

---

muss zu jeder Eigenschaft eines Objekts der entsprechende Wertebereich festgelegt werden. Die Angaben zur Größe eines Spurenträgers (Länge, Höhe, Breite) enthalten beispielsweise ausschließlich natürliche Zahlen.

3. *Erfassung der Objekteigenschaften:* Die Erfassung der einzelnen Werte einer Objekteigenschaft muss nachvollziehbar und reproduzierbar sein. Das bedeutet, dass den Verbundpartnern klare Richtlinien zur Erfassung einer Objekteigenschaft wie z.B. Schwellwert oder Messparameter zur Verfügung stehen müssen. Hierbei eignet sich insbesondere die Anwendung von normierten Messverfahren. Die Erfassung der Formabweichung (siehe auch [MST5.1]) einer Oberfläche z.B. zum Zweck der Materialklassifikation greift beispielsweise die DIN-Normen [DIN4760] bzw. [DIN 4761] zur Einordnung von Gestaltabweichungen auf. Die Anwendung normierter Methoden ist nicht nur für die eigentliche Datenerfassung sinnvoll, sondern auch für die Vorverarbeitung der Rohdaten der Sensorik [MST5.1]; z.B. bei der Extraktion des Rauheits- und Wellenprofils (siehe hierfür [Sor95]) aus den Topographiedaten mittels phasenkorrekter Filter nach [DIN 4777].
4. *Benamungsschema:* Der Austausch der Referenzdaten (M3.2) erfordert es, ein eindeutiges Benamungsschema für die auszutauschenden Objekte festzulegen. In diesem Zusammenhang muss definiert werden, wie dieses aussieht und wer für die Einhaltung desselben zuständig ist.
5. *Beziehungen der Objekte untereinander:* Es muss festgelegt werden, welche Beziehungen die Objekte untereinander aufweisen können. So ist beispielsweise jeder Fingerabdruck genau einer Person zugeordnet. Jedoch können mehrere Fingerabdrücke der (pseudonymisierten) Person in der DB enthalten sein.

Die in den Vorarbeiten gewonnenen Erkenntnisse fließen in die Definition von Anforderungen der Schnittstelle ein, die nachfolgend erläutert werden.

### **M3.1 Anforderungen an die Schnittstelle**

Die Gesamtanforderungen an die DB-Schnittstelle ergeben sich aus den einzelnen Anforderungen der Verbundpartner für Datenerhebung, -vorverarbeitung und -zugriff. Beachtet werden müssen weiterhin Auflagen der IT-Sicherheit sowie Anforderungen zur Verarbeitung unterschiedlicher Datenformate als auch die Anbindung an externe Tools. Dadurch ergibt sich eine komplexe Anforderungsbeschreibung an die DB-Schnittstelle. Im Folgenden werden die wesentlichen Ansprüche an die Schnittstelle dargelegt, die in Zusammenarbeit mit den Verbundpartnern erarbeitet wurden.

#### *Anforderung A1: Datenerhebung und Abbildung der Testszenarien*

Die Datenerhebung ist ein zentraler Bestandteil der Referenzdatenbank und dementsprechend umfangreich sind auch die Anforderungen an die DB-Schnittstelle, welche zur Datenerhebung bereitgestellt wird.

Bei der Datenerhebung wird zunächst eine Differenzierung vorgenommen zwischen Nutzdaten und Meta-Daten. Erstere beschreiben die eigentlichen Scan-Daten während Letztere Zusatzinformationen zu den vorgenommenen Scans (z.B. Scanparameter, Material) beinhalten und somit Aussagen über scan- oder materialabhängige Parameter ermöglichen, die das Scanergebnis beeinflussen. Sie beschreiben somit u.a. mögliche Störeinflüsse, wie Oberflächenbeschaffenheit des Spurenträgers oder Vergleichbares. Dementsprechend müssen diese Daten ebenfalls über die DB-Schnittstelle verarbeitet (z.B. gespeichert) werden.

---

Grundlage für die Modellierung der Referenzdatenbank und der zugehörigen Schnittstelle sind die definierten Testszenarien. Die Szenarien müssen möglichst genau über die Schnittstelle in der Referenzdatenbank abgebildet werden. Damit einhergehend steht auch die Forderung, dass die erhobenen Meta-Daten zu den Materialien und Proben das modellierte System möglichst exakt beschreiben. Dafür wurde vom Verbundpartner LKA-ST ein pauschalisierter Tatort entworfen der hier als Anforderungskatalog dient. Im Wesentlichen lassen sich aus diesem Katalog folgende Anforderungen ableiten:

1. Materialien – Über die DB-Schnittstelle müssen Materialien angelegt werden können. Im pauschalisierter Tatort des LKA-ST wird darüber hinaus auf Materialien verwiesen, auf denen häufig Fingerabdrücke an Tatorten entdeckt werden. Ziel hierbei ist es, u.a. den Einfluss unterschiedlicher Materialien auf die Scan-Ergebnisse besser untersuchen zu können.
2. Spureträger – Basierend auf den Materialien müssen konkrete Spureträger zusammen mit weiteren Informationen über ihr Erscheinungsbild (z.B. Oberflächenbeschaffenheit und Form) angelegt werden können. Auf diese Weise können z.B. Rückschlüsse auf materialunabhängige Einflüsse gezogen werden, die allein auf Eigenschaften wie die Form des Spureträgers zurückzuführen sind.
3. Proben – Diese Anforderung bezieht sich auf die eigentlichen Objekte, die letztendlich vom Scanner erfasst werden, im weiteren Verlauf auch Probe genannt. Über die DB-Schnittstelle muss es möglich sein Proben, basierend auf dem zugrunde liegenden Spureträger, in der DB anzulegen. Dabei sollen weitere Informationen wie zum Beispiel Lagerort, Typ (Material- oder Fingerabdruck-Scan) oder Institution (wo wurde die Probe angelegt) und Lagerbedingungen (wie Temperatur, Luftfeuchtigkeit etc.) erfasst werden.
4. Scan – Für eine angelegte Probe muss die DB-Schnittstelle das Einfügen der dazugehörigen Scans ermöglichen, wobei jeder Scan eindeutig einem Material, Spureträger und einer Probe zugeordnet sein muss. Die eigentlichen Nutzdaten (FRT-Datei) enthalten zusätzliche Meta-Daten, wie z.B. Scan-Parameter, die aus diesen extrahiert werden.
5. Proben verwalten – Für angelegte Proben muss es möglich sein, einen Lebenszyklus bzw. eine Historie zu verwalten. Das bedeutet, dass für Proben nach ihrer initialen Erstellung weitere Bearbeitungsschritte bzw. Scans erfasst werden müssen. Insbesondere ist hier darauf zu achten, dass bei einem erneuten Zugriff auf eine Probe (siehe auch Datenzugriff) die Beweiskette aufrechterhalten wird. Die Daten sind nicht in den eigentlichen Scans vorhanden, sondern werden als Meta-Daten in der DB hinterlegt, wobei einige der Daten automatisiert ermittelt werden.

Dabei ergibt sich aus den Anforderungen, dass das Anlegen bzw. Bearbeiten einer Probe nicht zwangsläufig einen Scan zur Folge hat.

Neben der Speicherung der Meta-Daten ist die Speicherung der eigentlichen Scan-Daten eine Hauptanforderung an die DB-Schnittstelle zur Datenerhebung. Grundlage für die Anforderungsanalyse sind dabei zwei definierte Datenformate. Zum einen das durch den Verbundpartner FRT vorgegebene proprietäre BinärfORMAT, welches nativ vom Scanner zur Verfügung gestellt wird. Zum anderen das vom Verbundpartner FH-BRB entwickelte Containerformat [KVL2011], welches sowohl die eigentlichen Scan-Daten in unterschiedlichen

---

Darstellungsformen bereitstellt als auch Meta-Daten vom Scanner (z.B. Scan-Parameter, Scan-Maße; siehe auch Abschnitt 3.3). Dabei stellen die Formate selbst schon die erste Anforderung an die DB-Schnittstelle, nämlich die Daten ohne Informationsverlust und nachverfolgbar in der Referenzdatenbank abzuspeichern. Die Anforderung der Nachverfolgbarkeit ergibt sich dabei aus der Anforderung nach einer lückenlosen Beweiskette (siehe auch Anforderung A2).

Im Rahmen des Projektes werden ebenfalls alternative Aufnahmegeräte zum Scannen von Material- oder Fingerabdruckobjekten analysiert, daher ergibt sich hier zwangsläufig die Anforderung an die DB-Schnittstelle, flexibel hinsichtlich des Datenformates zu sein.

#### *Anforderung A2: Sichere Datenbankinfrastruktur*

Die DB-Schnittstelle muss nachweislich bestimmte Aspekte der IT-Sicherheit sicherstellen, namentlich die Authentizität, Integrität und Vertraulichkeit [KVL2011,SSK2011]. Durch die ersten beiden Aspekte soll die Beweiskette gewährleistet werden. Das heißt u.a., dass alle Zugriffe auf oder Änderungen an den Daten dürfen a) nur durch autorisierte Personen (oder Tools) erfolgen und b) müssen lückenlos nachvollziehbar sein. Der letzte Aspekt dient dem Schutz personenbezogener Daten und der Erfüllung geltender Datenschutzbestimmungen. Um diese Bedingungen zu erfüllen, wurden von den Verbundpartnern UNI-MD-AMSL und UNI-MD-DB die Sicherheitsrisiken analysiert und entsprechende Richtlinien aufgestellt, wie demzufolge die Sicherheitsaspekte abzusichern sind. Die Ergebnisse lassen sich wie folgt zusammenfassen:

- *Pseudonymisierte Datenerhebung:* Die Referenzdaten werden pseudonymisiert erhoben. Folglich ist aus den Daten kein direkter Rückschluss auf die zugehörige Person möglich. Dies gilt sowohl für in der Datenbank gespeicherte Daten als auch für Daten, die zum Zwecke der Verifikation der Datenbank temporär zugeführt werden.
- *Einbettung in sichere Infrastruktur:* Die Referenzdatenbank als auch die DB-Schnittstelle werden in eine, entsprechend der Risikoanalyse, sichere Infrastruktur eingebettet. Diese umfasst eine verschlüsselte Übertragung und kann optional auf eine verschlüsselte Speicherung der Daten erweitert werden. Weiterhin wird die Datenbank nur Offline (d.h. auf einem separaten Rechner ohne Internetverbindung) oder im Intranet (vom Internet abgeschottetes lokales Netzwerk) verfügbar sein, um somit das Risiko unerlaubter Zugriffe zu minimieren.
- *Einheitliches sicheres Austauschformat:* Der Austausch von Scan-Daten (sowohl in die DB als auch von der DB zur Vorverarbeitung) erfolgt über ein einheitliches Austauschformat, welches von der FH-BRB entwickelt wurde. Dieses Austauschformat definiert einen Datencontainer, durch den u.a. Authentizität und Integrität der Daten überprüft werden können.

Die Anforderungen und Festlegungen bzgl. der IT-Sicherheit müssen zu jedem Zeitpunkt durch die DB-Schnittstelle überprüfbar gewährleistet sein.

#### *Anforderung A3: Datenzugriff*

Neben der initialen Erhebung von Referenzdaten muss die DB-Schnittstelle den Zugriff auf bereits bestehende Referenzdaten kontrollieren. Die Anforderung besteht darin, dass die DB-Schnittstelle resistent gegen unautorisierte Datenzugriffe sein muss. Dabei wird zwischen datenverändernden

---

Zugriffen (z.B. Verbessern der Datenqualität) und datenerhaltenden Zugriffen (z.B. Lesen von Daten) unterschieden. Unabhängig von der Art des Zugriffs besteht jedoch die Anforderung, jeden Zugriff einwandfrei nachzuverfolgen und auf die initialen Daten zurückzuführen, um somit die lückenlose Beweiskette zu gewährleisten. Speziell bei der Anbindung externer Tools (siehe Abbildung 1, unten) muss überprüft werden, ob diese zur Digi-Dak-Infrastruktur gehören. Ist dies nicht der Fall, so ist ein Zugriff von vornherein zu untersagen.

#### *Anforderung A4: DB-spezifische Anforderungen: Flexibilität und Skalierbarkeit*

Neben den bereits behandelten Anforderungen, welche in erster Linie auf die eigentlichen Daten fokussiert waren, existieren zusätzliche Anforderungen, die sich durch die Referenzdatenbank selbst und deren Nutzer ergeben. Auch die Anbindung bereits bestehender Systeme wie zum Beispiel AFIS ist dabei zu berücksichtigen. Das alles führt zur Anforderung, die DB-Schnittstelle flexibel hinsichtlich der Nutzungsanforderungen und damit der dahinterliegenden Referenz-DB zu gestalten. Die Anforderungen lassen sich dabei wie folgt spezifizieren:

- *Varianten (A4.1).* Aus den bisherigen Treffen der Verbundpartner UNI-MD-DB, UNI-MD-AMSL, FH-BRB und LKA-ST hat sich ergeben, dass es unterschiedliche Anforderungen an Funktionalitäts- und Datenumfang gibt. Hierzu gehört beispielsweise, dass gewisse Daten von einigen Projektpartnern z.B. aufgrund von Datenschutzbestimmungen nicht erhoben oder weitergegeben werden dürfen. Daher darf, um eine Sammlung dieser Daten von vornherein zu verhindern, die entsprechende Infrastruktur nicht vorhanden sein. Das bedeutet, dass die Schnittstelle für einzelne Varianten einen unterschiedlichen Funktionsumfang aufweisen muss. In gleichem Maße muss diese Funktionalität konsistent mit der dahinterliegenden Datenbank sein (z.B. mit dem Datenbankschema, siehe auch [SLSK2011]).
- *Versionen (A4.2).* In gleichem Maße muss die Flexibilität über den Zeitraum des Projektes (und darüber hinaus) gegeben sein. Dies ist dem Fakt geschuldet, dass Datenbank und Schnittstelle als Software zu betrachten ist, die im zeitlichen Verlauf gewissen Änderungen (z.B. durch veränderte oder neue Anforderungen) unterliegt (Softwarelebenszyklus und DB-Schemaevolution). Diese sollten ohne großen Aufwand in die bestehende DB-Schnittstelle integriert werden können. Zusätzlich dürfen mögliche Änderungen an der DB oder der DB-Schnittstelle die Integrität der bereits enthaltenen Daten nicht beeinträchtigen.
- *Skalierbarkeit (A4.3).* Die DB-Schnittstelle muss sich leicht auf die jeweilige IT-Infrastruktur des einzelnen Verbundpartners anpassen bzw. integrieren lassen. So muss es möglich sein, eine arbeitsfähige Variante auch mit limitierten Ressourcen zu betreiben, trotz großer Datenbeständen und hohem Datenverkehr. Weiterhin muss die DB-Schnittstelle die Anforderungen an die (Daten)Sicherheit unabhängig davon erfüllen, ob Sie nur auf einem einzelnen, isoliertem Rechner genutzt wird oder in einem (lokalen) Netzwerk.

### **M3.1 Modellierung und Implementierung der DB-Schnittstelle**

Nachfolgend wird die Modellierung und Implementierung der DB-Schnittstelle erläutert. Hierbei werden die Anforderungen A2 bis A4 aufgenommen und beschrieben inwiefern die Schnittstelle diese erfüllt. Die Erfüllung der Anforderung A1 wird im Abschnitt 3.5 veranschaulicht.



---

Die Schnittstelle selbst wird durch eine von UNI-MD-DB entwickelte prototypische Anwendung (DD-Client<sup>1</sup>) realisiert, welche u.a. die Übertragung der Daten der Testszenarien in die eigentliche DB<sup>2</sup> (FiVe DB) ermöglicht. DD-Client ist ein Forschungsdemonstrator, der vom Verbundpartner METOP in den Softwaredemonstrator integriert werden soll und den Ausgangspunkt zur Verbesserung der Usability bildet. DD-Client wird ebenfalls für M5.1 (Schnittstelle Datenvorverarbeitung) verwendet, wodurch Teile der Modellierung und Implementierung wiederverwendet werden können.

### *Umsetzung der Anforderung A2: Sichere Datenbankinfrastruktur*

Unter sicherer Datenbankinfrastruktur wird die Gewährleistung der Authentizität, Integrität und Vertraulichkeit der Daten verstanden, die über diese Schnittstelle ausgetauscht werden. Die ersten beiden Punkte werden vor allem durch das einheitliche sichere Austauschformat erzielt. Die Datenbankinfrastruktur zielt vor allem (jedoch nicht ausschließlich) auf die Sicherung der Vertraulichkeit der Übertragungsverbindung der Schnittstelle mit einem, je nach eingesetzter Technik, spezifischen Restrisiko. Es werden im Folgenden die einzelnen Punkte der Infrastruktur erläutert.

#### *Absicherung der Datenverbindung*

Jede Verbindung des Digi-Dak Database Client (DD-Clients) zur DB wird via SSL (Secure Socket Layer) gesichert. Der Datenbankserver kann so eingestellt, dass er ausschließlich per SSL gesicherte Verbindungen akzeptiert. Zu diesem Zweck wird eine X-509 Zertifikatinfrastruktur erstellt. Der Einsatz von SSL hat folgende Vorteile:

- **Verschlüsselte Verbindung:** Das Abhören von zwischen der DB und DD-Client versendeten Daten wird deutlich erschwert, da deren Übertragung verschlüsselt erfolgt.
- **Client-Server-Authentifizierung:** Der Einsatz von SSL ermöglicht es, dass DD-Client Daten ausschließlich an einen DB-Server mit dem entsprechenden (Server-) Zertifikat versendet. Somit wird verhindert, dass sich ein Rechner fälschlicherweise als DB-Server ausgibt. Umgekehrt kann der DB-Server anhand des (Client-) Zertifikats überprüfen, ob dieser sich mit der DB verbinden darf.

Der Einsatz von SSL wird sowohl vom DB-Server als auch von der Programmiersprache, in der DD-Client entwickelt wurde, nativ unterstützt. Somit kann auf bereits getestete Software zurückgegriffen werden, deren Funktionsfähigkeit vielfach erprobt ist. Die Sicherheit der Verbindung hängt von der Geheimhaltung der privaten Schlüssel der Zertifikatinfrastruktur ab.

Zusätzlich zur Nutzung von SSL wird der IP-Bereich der Rechner beschränkt, die sich mit der DB verbinden können. Diese Funktion wird ebenfalls direkt von der DB angeboten. Die IP-Adresse ist in Rechnernetzwerken eindeutig und (wie in unserem Fall) meist statisch zugewiesen; sie ist allerdings nicht fälschungssicher. Es wird eine Liste von IP-Adressen definiert, die sich mit der DB verbinden können (Whitelist). Alle anderen Verbindungsversuche von Rechnern mit IP-Adressen, die nicht in der Liste enthalten sind, werden ohne weitere Prüfung abgelehnt.

Weiterhin ist eine Verbindung zum DB-Server via DD-Client nur möglich, wenn ein korrektes Paar aus Nutzernamen und -passwort angegeben wird.

---

1 Die Software ist in der Programmiersprache JAVA entwickelt worden.

2 Als Datenbankmanagementsystem wird Postgres 8.4 verwendet.



---

Für einen erfolgreichen *Verbindungsversuch* eines Rechners mit der DB muss

1. der Rechner über eine IP-Adresse verfügen, die auf der Whitelist der DB enthalten ist,
2. der Rechner ein Zertifikat besitzen, welchem die DB vertraut,
3. ein korrektes Paar bestehend aus Nutzernamen und Passwort an die DB übermittelt werden.

Somit existieren drei Sicherungsmechanismen, die *alle* umgangen werden müssten, um unautorisierten Zugriff über ein Rechnernetz zu erlangen. Die Daten, die ein Nutzer sieht bzw. editieren kann, sind abhängig von dessen Rechten (siehe A3: Datenzugriff).

#### *Absicherung der einzelnen Rechner*

Zur sicheren Infrastruktur gehört nicht nur die Absicherung der Verbindungen, sondern auch der einzelnen Rechner inklusive Datenbankserver. Zu diesem Zweck wurden von UNI-MD-AMSL verschiedene technische und organisatorische Maßnahmen erarbeitet, die dies gewährleisten.

#### *Weitere Schutzmaßnahmen*

Die einzelnen Standorte sind aufgrund von organisatorischen Maßnahmen unterschiedlich stark gesichert. Der Zugang zum Verbundpartner LKA-ST ist z.B. deutlich restriktiver als bei den Partnern UNI-MD und FH-BRB, so dass Maßnahmen in Zusammenarbeit mit UNI-MD-AMSL entwickelt wurden, die abhängig von der Grundsicherungslage und vorhandener IT-Ressourcen eingesetzt werden.

- *Lokales Netz:* Abschottung der Digi-Dak-Infrastruktur in einem Intranet ohne Internetverbindung und Begrenzung des Personenkreises der Zugriff auf dieses Netz hat.
- *Feingranulares Logging.* Die Mechanismen sollen nicht nur Manipulationsversuche verhindern, sondern diese auch erkennen und sichtbar machen. Zu diesem Zweck wird u.a. die interne Loggingfunktion der DB genutzt als auch der Mitschnitt der gesamten Kommunikation im Rechnernetz.
- *Grundverschlüsselung der DB:* Die Daten der DB sollen nicht im Klartext auf dem DB-Server abgelegt werden, sondern in verschlüsselter Form. Hierzu existieren zwei Möglichkeiten:
  - (a) Verschlüsselung aller Daten auf allen Festplatten des DB-Servers
  - (b) Verschlüsselung der DB-Daten durch die DB selbst.

#### *Umsetzung der Anforderung A3: Datenzugriff*

Vorausgesetzt ein Verbindungsversuch zur DB war erfolgreich (siehe Umsetzung A2) muss definiert sein, welche Aktionen ein Nutzer auf den Daten ausführen darf. Hierbei gilt das Prinzip des *mandatory access control* [Bis]. Ein neuer DB-Nutzer kann sich anfangs ausschließlich mit der DB verbinden, hat jedoch keinerlei Zugriff auf irgendein Datenobjekt. Der Zugriff muss ihm erst erteilt werden. Im Datenbankenbereich wird generell zwischen drei Zugriffsformen unterschieden, im Einzelnen sind dies:

1. Lesender Zugriff: Hierbei erfolgt keinerlei Veränderung der bestehenden Daten.
2. Einfügen neuer Daten: Schreibender Zugriff bei dem neue Daten in die DB eingefügt, bestehende jedoch nicht verändert werden.

- 
3. Modifikation bestehender Daten: Im Gegensatz zu Punkt zwei werden hierbei bestehende Datensätze der DB verändert, jedoch keine neuen Daten eingefügt.

Die Umsetzung der Zugriffskontrolle erfolgt durch ein Rollen- und Rechtemanagementsystem, welches als Teil des SQL-Standards fest in die DB integriert ist. Somit setzt die Umgehung dieses Mechanismus die Manipulation der DB selbst voraus. Manipulationen an der Schnittstelle (DD-Client) selbst sind hierbei wirkungslos (z.B. SQL Injection). Die Erstellung der einzelnen Rollen und deren zugehörigen Rechten erfolgte in Abstimmung mit UNI-MD-AMSL.

Beispielhaft sei hier die Rolle *Datenerhebung* erläutert, da diese für die Personen konzipiert wurde, die die Erhebung der Referenzdaten ausführen und daher die hier beschriebene Schnittstelle nutzen. Nutzern der DB, denen die Rolle *Datenerhebung*<sup>3</sup> zugeteilt wurde, haben z.B. das Recht neue Spureinträger und Scans in die DB einzutragen. Sie können weiterhin sehen, welche Materialien und Spureinträger sich bereit in der DB befinden. Zusätzlich können sie die Meta-Daten zu jedem in der DB befindlichem Scan lesen. Das eigentliche Bild (auf dem sich u.U. ein Fingerabdruck befindet) ist ihnen jedoch nicht zugänglich, obwohl sie diese Daten einfügen können. Generell darf diese Rolle derzeit keinerlei Daten in der DB modifizieren.

#### *Umsetzung der Anforderung A4: DB-spezifische Anforderungen*

Aufgrund der heterogenen Aufgaben der Verbundpartner ergeben sich teilweise unterschiedliche Anforderungen an die DB-Schnittstelle. Hinzukommt, dass zu erwartende neue Erkenntnisse in Bezug auf die Funktionsweise der Sensorik in unserem speziellen Anforderungsbereich und ein besseres Verständnis des Einflusses von Umweltfaktoren auf die Messergebnisse, eine Anpassung der Schnittstelle erfordern. Beide Punkte werden unter dem Begriff Revisionen zusammengefasst, die Varianten und Versionen umfassen. Weiterhin müssen die unterschiedlichen IT-Landschaften der Verbundpartner berücksichtigt werden.

*Varianten (A4.1).* Varianten umfassen unterschiedliche Implementierung der Schnittstelle zur selben Zeit aufgrund unterschiedlicher Anforderungen, wobei sich Anforderungen durchaus widersprechen können. Die Erstellung von Varianten auf DB-Schemaebene ist dabei als wissenschaftliches Neuland zu betrachten, zu dem von uns in [SLSK2011] Grundlagenforschung betrieben wurde. In dieser Arbeit konnten wir zeigen, dass sich im DB-Bereich Varianten erstellen lassen. Unterschiedlichen Anforderungen zeichnen sich für die hier beschriebene Schnittstelle z.B. auf folgenden Gebieten ab:

- Optionale automatisierte Erfassung von Lagerungsbedingungen (Temperatur, Luftdruck etc.) für die Bestimmung von Alterungseffekten bei Fingerabdrücken und Verunreinigungen.
- Analyse des Einfluss von personenbezogenen Faktoren auf die Qualität der Fingerabdrücke einer Testperson und deren chemische Zusammensetzung (nur LKA-ST) wie beispielsweise der Gebrauch gewisser Medikamente.

*Versionen (A4.2).* Mit Versionen sind die Veränderungen an der Schnittstelle über die Zeit gemeint (Schemaevolution), die sich z.B. aufgrund neuer Erkenntnisse aus dem Projekt ergeben. Die Schnittstelle (DD-Client) ist daher so entworfen worden, dass sie sich mit möglichst wenig Aufwand flexibel an die folgenden Änderungen anpassen kann:

---

<sup>3</sup> Einem Nutzer können mehrere Rollen zugeteilt werden und eine Rolle kann einem Nutzer wieder genommen werden, falls er die damit verbundenen Rechte nicht länger benötigt.

- *Anpassung der Pflichtfelder:* Zu jedem Datenobjekt in der DB ist hinterlegt, ob dessen Attribute (Eigenschaften des Objekts, siehe 3.2) Pflichtfelder sind. DD-Client muss für jedes der Felder anzeigen, ob es sich um eine Pflichteingabe handelt, um Datenbankfehler beim Versuch des Einfügens in der DB zu verhindern. Die Information, ob ein Feld ausgefüllt werden muss, werden durch die Schnittstelle direkt von der DB bezogen und sind nicht statisch in der Schnittstelle gespeichert. Die Anpassung, inwiefern ein Attribut ein Pflichtfeld ist, muss daher nur in der DB geändert werden und wird automatisiert überprüft.
- *Überprüfung der gültigen Eingaben bereits durch die Schnittstelle:* Analog zur Anpassung der Pflichtfelder kann sich der Wertebereich eines Attributes im Laufe des Projektes ändern. Die Überprüfung von Eingaben auf Schnittstellenseite (z.B. ob bei der Größenangabe eines Spurentägers ausschließlich Zahlen angegeben werden) muss in diesem Fall angepasst werden. An dieser Stelle wird derselbe Mechanismus wie bei Pflichtfeldern verwendet und die Informationen zum Wertebereich und Länge bzw. Präzision des Attributes werden direkt aus der DB gezogen und sind nicht redundant in DD-Client gespeichert. Somit reduziert sich der Aufwand zum Anpassen der Schnittstelle.
- *Einfügen zusätzlicher Attribute:* Im Laufe des Projekts kann es, z.B. aufgrund neuer Erkenntnisse, notwendig sein zusätzliche Attribute zu einem in der DB gespeichertem Objekt hinzuzufügen. Dieses Attribut muss somit ebenfalls in die Definition der Schnittstelle aufgenommen werden. Der Vorgang wird unterstützt, indem das zusätzliche Attribut lediglich an ein Element des Eingabeformulars im DD-Client gebunden werden muss. Die Informationen zum Wertebereich und Pflichtfeldcharakter des Attributes werden, wie zuvor beschrieben, direkt aus der DB gezogen und müssen nicht redundant in der Schnittstelle hinterlegt werden.
- *Anpassungen des Rechtemanagements:* Anpassungen der zugehörigen Rechte einer Rolle sowie die Modifikation der Zuordnung von Rollen zu einem DB-Nutzer in der DB müssen sich ebenfalls auf die Schnittstelle auswirken. Zu diesem Zweck werden die Rechte beim Anmelden eines Nutzers direkt aus der DB gelesen. Somit ist sichergestellt, dass beide Systeme synchronisiert werden und die Rollen und Rechtesystem dennoch fest in die DB integriert ist. Alternativ dazu müssten die Rechte manuell in DD-Client gepflegt werden, was einen erheblichen Mehraufwand verursacht. Die Vorbereitung der Anpassbarkeit der Schnittstelle für die soeben genannten vier Punkte (diese Punkte stellen nach derzeitigem Kenntnisstand die wahrscheinlichsten Modifikationen dar) wirkt sich teilweise negativ auf andere (weniger wahrscheinliche) Modifikationen aus. Zu diesen Modifikationen zählt beispielsweise das Umbenennen von Attributen in der DB, da diese Attribute über deren Namen fest an die Schnittstelle gebunden sind und daher eine Anpassung der Schnittstelle erfordern. Wir gehen jedoch davon aus, dass Umbenennungen während der Projektzeit selten vorkommen werden.

*Skalierbarkeit (A4.3).* Die heterogene IT-Landschaft der Verbundpartner erfordert es, dass sich die Gesamtlösung inkl. DB und zugehöriger Schnittstellen flexibel in die bestehende Infrastruktur einbetten kann. Wir adressieren diesen Punkt, indem sich die Implementierung der Schnittstelle DD-Client folgendermaßen entworfen/implementiert wurde:

- *Plattformunabhängigkeit:* DD-Client ist in JAVA implementiert und nutzt durch das SWT Paket nativ den Funktionsumfang des verwendeten Betriebssystems und ist daher unabhängig vom verwendeten Betriebssystem. Getestet wurde die Implementierung bislang auf Windows 7, Windows XP und Ubuntu 10.04.

- 
- *Unterstützung verschiedener Datenformate:* Die Untersuchung zur Eignung verschiedener Scanner für den Grobscan erfordert es, neben dem universellen Austauschformat, weitere Bildformate in der DB ablegen zu können. Zu diesem Zweck werden sämtliche Scan-Ergebnisse als Binärdatenstrom (BLOB) behandelt. Für sämtliche unterstützten Formate muss dann lediglich eine Bibliothek zur Verifikation, dass es sich bei dem Binärdatenstrom um das spezifizierte Format handelt, in die Schnittstelle und DB integriert werden.
  - *Betrieb von DD-Client und DB auf einem Rechner:* Aufgrund der Platzformunabhängigkeit von DD-Client und dem verwendeten Datenbankmanagementsystem ist es möglich alle Softwareinstanzen auf ein und demselben Rechner zu betreiben. Dies kann zur Not der Messrechner sein. Weiterhin können, wie in UNI-MD, die einzelnen Softwareinstanzen auf verschiedenen Rechnern mit unterschiedlichen Betriebssystemen laufen.

### **M3.1 Zusammenfassung**

Nachfolgend wird der derzeitige Funktionsumfang der *Implementierung* der Schnittstelle in DD-Client bezüglich *AI* zusammengefasst und zukünftige Arbeitsschritte aufgezählt.

#### *Funktionsumfang DD-Client*

Der derzeitige Funktionsumfang der DB-Schnittstellenimplementierung in DD-Client v.02 Revision 163 ermöglicht das Ausführen folgender Aktionen, die sich aus den Beschreibungen der Testszenerarien ergeben. Die einzelnen Aktionen sind ausschließlich dann möglich, wenn der Nutzer über die entsprechenden Rechte verfügt.

#### Meta-Datenakquise

Die Beschreibung der Umgebungseinflüsse erfolgt durch die Aufnahme von Meta-Daten, die den Zustand des modellierten Systems bei der Aufnahme der eigentlichen Nutzdaten (Scan-Ergebnisse) nach derzeitigem Kenntnisstand beschreiben.

*Materialien.* Derzeit können Materialien in der Referenzdatenbank angelegt und sämtliche bereits in ihr vorhandenen Materialien angezeigt werden. Dies ermöglicht, dass der Einfluss materialabhängiger Parameter untersucht werden kann.

*Spurenräger:* Das Einfügen von Spurenrägern beinhaltet die Erfassung der Größen des Objekts sowie dessen Oberflächeneigenschaften. Zu diesen gehören u.a. die Strukturierung der Oberfläche und deren Absorptionsverhalten bezüglich der Fingerabdruckmasse. Zusätzlich ist jeder Spurenräger einem in der DB vorhandenen Material zugeordnet, um zwischen materialabhängigen und materialunabhängigen Einflüssen unterschieden zu können.

*Proben:* Jede Probe befindet sich auf genau einem in der DB befindlichen Spurenräger. Zusätzlich werden Lagerort und Art der Probe vermerkt. Hierbei wird zwischen Material- und Fingerabdruck-Scan unterschieden. Die Zuordnung der Klimadaten erfolgt über den Lagerort.

*Probenhistorie:* Die Liste von Aktionen zu den einzelnen Proben (anlegen, scannen, bedampfen, reinigen etc.) kann bei entsprechenden Rechten angezeigt werden.

---

## Nutzdatenakquise

Nutzdaten sind die eigentlichen Scan-Ergebnisse. Sie liegen im proprietären Binärformat der FRT GmbH vor und werden im weiteren Verlauf des Projektes durch das universelle Austauschformat der FH-BRB ersetzt. An die eigentlichen Nutzdaten sind weitere Meta-Daten geknüpft. Diese stammen z.T. vom Sensor selbst (dazu gehört beispielsweise die Auflösung). Sie sind in die Datei(en) des Austauschformats integriert. Weiterhin wird über die Schnittstelle manuell das Aufnahmegerät und der Scan-Typ (Grob oder Fein-Scan) erfasst. Die Schnittstelle ist derzeit unabhängig vom verwendeten Austauschformat (siehe hierzu auch 3.5.3).

### *Zusätzliche Funktionen der DB-Schnittstelle*

Der Abschnitt zusätzliche Funktionen der DB-Schnittstelle enthält Operationen auf den Daten der Referenzdatenbank, die die Schnittstelle bereits (für andere Tools) anbietet, jedoch bisher nicht in DD-Client integriert sind. Es ist vorgesehen, diese zeitnah in den Funktionsumfang von DD-Client aufzunehmen.

### Meta-Datenakquise

*Probenhistorisierung:* Die Funktionen zum Erzeugen von Historisierungseinträgen sind bereits in die Schnittstelle integriert. Sie beinhaltet eine Liste von vordefinierten Aktionen (erarbeitet durch UNI-MD-DB, UNI-MD-AMSL und LKA-ST), sowie einen Mechanismus zur Erweiterung der vordefinierten Aktionen.

*Automatisierte Klimadatenerfassung:* Die automatisierte Erfassung der Klimadaten der einzelnen Lagerorte (siehe auch A4.1) zur Vervollständigung des Meta-Datenmodells der Referenzdatenbank ist über die DB-Schnittstelle möglich. Jedem Lagerort wird eine Wetterstation zugeordnet, die deren Klimadaten aufnimmt. Erfasst werden derzeit Temperatur und Luftfeuchtigkeit. Die Nutzung der Schnittstelle ist z.B. via CSV-Dateien möglich, welche von der Messstation erzeugt werden.

---

Analog zur Zusammenfassung der Ergebnisse des M3.1 werden nun die Ergebnisse des M3.2 *Zusammenführung und Einpflegen der Referenzdaten* erläutert.

## **Darstellung der Ergebnisse für M3.2 „Zusammenführung und Einpflegen der Referenzdaten“ zum AP 3**

Der Inhalt des Meilensteins gliedert sich wie folgt. Zuerst wird die Zielsetzung der Referenzdatenerfassung erläutert. Auf Basis dieser Zielsetzung ergeben sich Anforderungen an die einzelnen Datensätze, die in der Definition verschiedener Referenzdaten für unterschiedliche Teile der Referenzdatenbank münden. Darauf aufbauend werden die Zuarbeiten der einzelnen Projektpartner, auf Basis ihres individuellen Forschungsprofils, aufgelistet. Anschließend werden erste Anwendungen der Referenzdaten in Bezug auf AP5 vorgestellt.

### **Zielsetzung der Referenzdatenerfassung**

Die Sammlung der Referenzdaten soll gezielt entsprechend der Verwendung in den einzelnen Anwendungsszenarien erfolgen. Aus diesem Grund werden solche Anwendungsszenarien und dafür zur Verfügung stehende Daten in diesem Abschnitt erläutert.

Bevor Ziele bzw. darauf beruhende Anforderungen definiert werden können, wird nachfolgend veranschaulicht, welche Nutzdatenobjekte in der DD-Infrastruktur für die Sammlung von Referenzdaten relevant sind. Wie in Abbildung M3.2.1 dargestellt, existieren zwei dominante Dimensionen (Art der Daten & Verarbeitungsstatus) anhand derer sich die Daten bezüglich Verarbeitungsstatus und Datenschutzrechtsanforderungen unterscheiden.

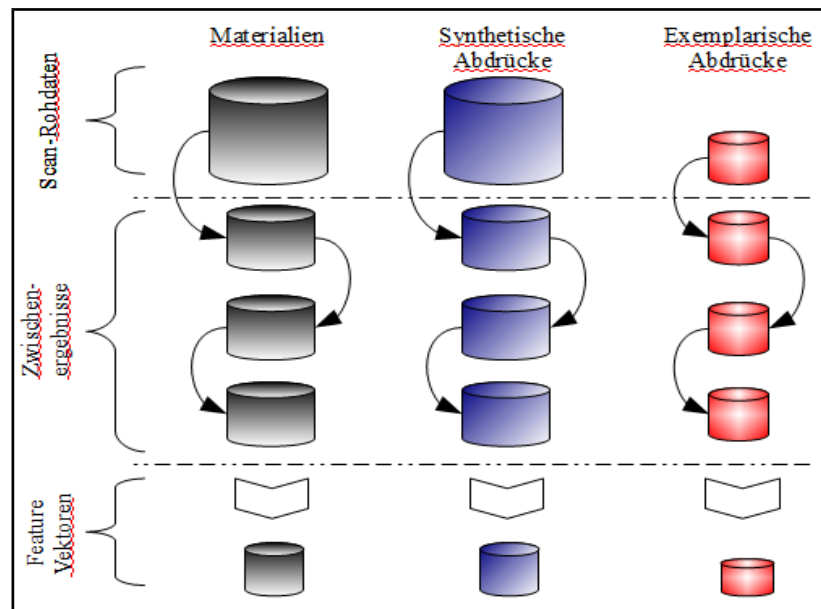


Abbildung M3.2.1: Nutzdaten der DD-Infrastruktur

Art der Daten. Wir unterscheiden an dieser Stelle grundsätzlich drei Arten von Daten: (1) Reine Materialscans, (2) Synthetische Abdrücke sowie (3) Exemplarische Abdrücke.

(1) *Materialien*: Hierbei handelt es sich um Datenobjekte, die sich ausschließlich auf reine Materialscans zurückführen lassen. Sie enthalten keinerlei Fingerabdrücke und fallen daher nicht in die Kategorie der personenbeziehbaren Daten und können somit ohne besondere Berücksichtigung von Datenschutzrechtsbestimmungen ausgetauscht werden.

(2) *Synthetische Abdrücke*: Hierbei handelt es sich um generierte Fingerabdrücke, die mittels spezieller Verfahren gedruckt und unter Nutzung der Scantechnik erfasst wurden. Sie sind bei gewissen Auflösungen optisch kaum von realen Abdrücke zu unterscheiden. Diese Daten enthalten analog zu den reinen Material-Scans keine Merkmale echter Personen und fallen daher nicht unter die Bestimmungen des Datenschutzgesetzes.

(3) *Exemplarische Abdrücke der AG-DB*: Daten dieser Kategorie enthalten personenbeziehbare Daten, so dass deren Erfassung und Lagerung durch die Bestimmungen des Landesdatenschutzgesetzes vorgegeben sind.



Verarbeitungsstatus. Je nach Verarbeitungsstand der Daten in bezug auf das Vorgehensmodell von UNI-MD-AMSL (AP12) unterscheiden sich die Daten bezüglich Größe, Informationsgehalt und Aussagekraft. Sie eignen sich daher für bestimmte Einsatzszenarien, wie in 3.1.2 beispielhaft dargestellt.

(1) *Scan Rohdaten*: Hierbei handelt es sich um die direkten Ergebnisse des Scanprozesses. Diese Daten enthalten den größtmöglichen Informationsgehalt und Datenmenge, allerdings müssen meist Vorverarbeitungstechniken angewandt werden um einzelne Informationen herauszustellen (Aussagekraft).

(2) *Zwischenergebnisse*: Diese Daten entstehen durch die Anwendung von Vorverarbeitungsmethoden auf Rohdaten (oder andere Zwischenergebnisse). Dabei werden die Rohdaten in inhaltsbasierte Darstellungen überführt.

(3) *Feature-Vektoren* sind das finale Ergebnis der Vorverarbeitung. Hierbei sind verschiedene inhaltsbasierte Darstellungen zu einem (normalisierten) Vektor zusammengefasst, der aufgrund der geringen Größe und der Eigenschaften des aufgespannten Raumes sich für Anfrageverarbeitung und -optimierung (AP5) bezüglich eines oder mehrerer Anwendungsszenarien eignet. Hierbei können gezielt Informationen ausgelassen werden, die beispielsweise an dieser Stelle nicht von Interesse sind. Daher sollen, sofern möglich, die Berechnungen auf Feature-Vektoren erfolgen. Formal ist ein Feature-Vektor ein Punkt (p) in einem d-dimensionalen Raum (R) [WSB98, ZRL96].

$$\vec{p} \in [0, a]^d \subset R^d$$

### Anwendungsbeispiele für die verschiedenen Referenzdaten

Die Sammlung der Referenzdaten erfolgt im Hinblick auf die verschiedenen Datenobjekte unter Berücksichtigung ihrer Einsatzszenarien. Aus diesem Grund werden nachfolgend wesentliche Szenarien vorgestellt.

Tabelle M3.2.1: Anwendungsbeispiele der Referenzdaten

	Reine Materialien	Mischformen	Synthetische Fingerabdrücke	UNI-MD-DB: Echte Abdruckssamples
Scan-Rohdaten	Materialbestimmung	---	RoI Bestimmung	Verifikation der Ergebnisse der synthetischen Abdrücke
Featurevektoren	Materialklassifikation	RoI Bestimmung	Klassifikation Ridge/Oberfläche	

ROI = Region of Interest

*Materialbestimmung*. Sowohl für den Fein-Scan, als auch für die individuelle Anpassung der Vorverarbeitungsschritte (z.B. von Schwellwertparametern) ist die Bestimmung des Materials, bzw. die Zuordnung zu einer *Materialklasse* notwendig. Hierzu könnten sich beispielsweise ähnlichkeitsbasierte Ansätze eignen. Die Materialien werden hierbei im wesentlichen, auf die im

---

(LKA-ST) aufgeführten Gruppen begrenzt, da es sich um die häufigsten anzutreffenden Oberflächen eines potentiellen Tatortes handelt.

*RoI-Bestimmung.* Die Ermittlung von Regions of Interest (RoI) ist gleichbedeutend mit der Bestimmung inwiefern bzw. mit welcher Wahrscheinlichkeit sich auf einem rechteckigem Areal eines Grob-Scans ein Fingerabdruck (oder sonstige Anomalie) befindet. Diese RoIs sind dann Eingabe für den anschließenden Fein-Scan in dem zusätzliche Merkmale sichtbar gemacht werden sollen (z.B. Minutien und Schweißsporen), die im Grob-Scan nicht zu erkennen sind.

Für die Bestimmung einer RoI ist zuvor die Identifikation der Materialklasse hilfreich. Weiterhin werden für dieses Anwendungsszenario in der Regel Feature-Vektoren verwendet, die keine personenbezogene Merkmale enthalten. Alternativ dazu stehen Rohdaten mit gedruckten Abdrücken zur Verfügung. Zur *Verifikation* der Ergebnisse der synthetischen Abdrücke, existiert eine kleine Sammlung echter Abdrücke, die ggf. erweitert werden muss.

### **M3.2 Definition der Referenzdaten**

Die Definition der Referenzdaten umfasst sowohl Umfang, Auflösung als auch Qualität und Quantität der Daten. Sie orientiert sich maßgeblich an den zuvor geschilderten Einsatzzwecken. Die nachfolgende Auflistungen beschreiben die gewünschten Materialdaten. Eine Auflistung der bereits erhaltenen und zukünftig noch zu erzeugenden Scans findet sich im folgenden Abschnitt 3.3.

#### *Materialdatenbank*

In der Materialdatenbank werden Scans zu unterschiedlichen Materialien vorgehalten. Somit sollen beispielsweise unbekannte Materialien einer Gruppe mit bekannten Eigenschaften (z.B. bezüglich der verwendeten Scan-Parameter) zugeordnet werden.

#### Art der Scans

Die Scans umfassen sowohl 3D Oberflächenmessungen als auch 2D Profilschans (pro Sample ein horizontaler und vertikaler Scan). Da sich 2D Profilschans aus den 3D Oberflächenmessungen extrahieren lassen und die Dimensionalität, welche sich vor allem für die Dauer und Aussagekraft bei Nachbarschaftsanfragen<sup>4</sup> [BGRS98] positiv auswirkt, besteht die Hoffnung die Materialsbestimmung auf Profilschans durchführen zu können. Sollte dies nicht der Fall sein müssen die entsprechenden 3D Oberflächenmessungen genutzt werden.

#### Umfang der zu erfassenden Materialien

Die zu scannenden Materialien richten sich nach den Materialgruppen des pauschalisierten Tatortes. Zur sinnvollen Betrachtung der intra- und interklassen Varianz sind pro Material mindestens 10 Samples von verschiedenen Messgeräten/Standorten in die Referenzdaten aufzunehmen.

*Repräsentanten der drei Materialklassen.* Die Materialien werden in drei Klassen (kooperative, bedingt kooperative und bislang unkooperative Materialien) eingeteilt, die angeben wie gut sie sich

---

<sup>4</sup> **Beispiel 1.** Bei einem Scan von 1,5cm \* 1,0cm mit 4096 Grauwerten (Intensität oder Topographie) und einem Punktabstand von 100µm ergeben sich im 3D-Profilschan 15.000\*10.000\*4.096 Dimensionen. Hingegen entstehen für die beiden Profilschans nur 15.000\*1\*4069 und 1\*10.000\*4069 Dimensionen ohne weitere Reduktion durch inhaltsbasierte Darstellungen in Feature Vektoren (Trivialmerkmal).

für die derzeit evaluierten Messverfahren eignen. Für die Bestimmung der Stabilität, beispielsweise in Bezug auf aussagekräftige Feature-Vektoren der Referenzdaten, werden hier deutlich größere Mengen an Referenzdaten benötigt. Hierbei ist darauf zu erachten, dass die Daten von allen drei zur Verfügung stehenden Oberflächenmessgeräten stammen.

### Größe und Auflösung der Daten

Zur Sicherung der Vergleichbarkeit der verschiedenen Scanreihen werden nachfolgend die Auflösung und Größe der Referenzdaten festgelegt. Da final Featurevektoren auf den Referenzdaten berechnet werden sollen und zu große Samples zurecht geschnitten werden können, sind die Angaben zur Größe als Richtlinie anzusehen. Wir empfehlen jeweils Ausschnitte von 1,5cm \* 1,0cm aufzunehmen (ungefähre Größe eines Abdruckes & Geringe Scanzeit).

Hingegen ist die Auflösung ein entscheidender Faktor, da bei höheren Auflösungen beispielsweise Minutien, Schweißsporen und einzelne verteilte Ablagerungen der Abdrucksmasse (im Gegensatz zu durchgezogenen Linien) zu erkennen sind. In Sammlung der Referenzdaten werden folgende Auflösungen aus Tabelle 2 berücksichtigt.

Tabelle 3.2.2: Auflösung der Materialreferenzdaten

Art der Daten	Auflösung	Anwendungszweck
2D Profilsan	100µm	Schneller Grobscan zur Materialklassifikation (Scan soll sich nicht oder nur bedingt zur Identifikation eignen)
3D Oberflächenmessung	100µm	(Alternative) Berechnung der 2D Profilsans auf bestehenden Datensammlungen und Fallback, falls Klassifikation auf Profilen nicht möglich.
3D Oberflächenmessung	50µm	Einbeziehung und Unterscheidung von Fingerabdruckmerkmalen von Oberflächenstrukturen (RoI).
3D Oberflächenmessung	25µm	Weitere Einbeziehung und Unterscheidung von detaillierteren Fingerabdruckmerkmalen von Oberflächenstrukturen (RoI).
3D Oberflächenmessung	10µm	<b>Optional Erweiterung</b> falls im hochdetaillierten Messbereich wesentliche Merkmale z.B. für die anschließende Vorverarbeitung von Fein-Scans ermittelt werden.
3D Oberflächenmessung	2µm	<b>Optional Erweiterung</b> falls die kleinstmögliche Auflösung des FRT-Messgerätes wesentliche Merkmale z.B. für die Vorverarbeitung von Fein-Scans ermittelt.

---

## Qualitative Anforderungen

Bei der Bestimmung der Qualität eines Samples müssen sowohl objektive als auch subjektive (nicht oder nur schwer messbare) Einflüsse berücksichtigt werden müssen.

*Objektive Merkmale.* Die bisherigen Erkenntnisse zeigen, dass Umgebungseinflüsse möglichst konstant gehalten werden müssen, deshalb sollen die Klimadaten zu den einzelnen Messungen mit aufgenommen werden. Die Angabe von maximal gültigen Änderungen erscheint, aufgrund fehlenden Wissens über den tatsächlichen Einfluss, derzeit nicht sinnvoll.

*Subjektive Merkmale.* Zu den subjektiven Kriterien gehören beispielsweise die Berücksichtigung sämtlicher, nicht in den Klimadaten erfassten Ereignisse, die Einfluss auf das Messergebnis haben können. Dazu gehören beispielhaft Schwingungen (z.B. Straßenbahn) und Änderungen des Umgebungslichtes. Zusätzlich muss beachtet werden wie viele ungültige Messwerte enthalten sind. Ein weiterer Punkt sind Artefakte, die durch Verschleiß und unbeabsichtigte Beschädigung der Oberfläche (z.B. durch Mehrfachverwendung der Probe) das Messergebnis beeinträchtigen können.

Daher kann die qualitative Eignung eines Samples nur durch die aufnehmenden Person bestimmt werden. Hierbei muss erwähnt werden, dass eine gewisse Intraklassenvarianz, wie im realen Einsatz zu erwarten, durchaus erwünscht ist.

### *Feature Vektoren für die Daten der Materialdatenbank*

Neben der Aufnahme der Scan-Rohdaten ist die Sammlung (geeigneter) Feature-Vektoren bzw. inhaltsbasierter Darstellungen, die anschließend zu Feature-Vektoren zusammengefasst werden, von Bedeutung.

Für die Zuordnung eines (unbekannten) Materials zu einer der drei Materialklassen wurden bzw. werden noch von der FH BRB Repräsentanten für diese Klassen ausgewählt und die entsprechenden Materialien in großer Zahl gescanned. Diese Scans sollen als Referenz z.B. für die Zuordnung eines unbekanntes Materials zu einer dieser Klassen über die Bestimmung der nächsten Nachbarn dienen. Zur Reduktion der Dimensionalität und Erhöhung der Aussagekraft der Scan-Ergebnisse werden beschreibende Oberflächenmerkmale verwendet, die beispielsweise die Angaben über die Homogenität der Oberfläche (sowohl Intensitäts- als auch Topographiedaten).

*Anonymisierung von Mischformen.* Durch die Verwendung von Oberflächenmerkmalen enthalten die Daten, selbst wenn auf den Rohdaten ein Fingerabdruck enthalten ist, keine personenbeziehbaren Daten, sodass Mischformen erschaffen werden können, die wir zur Materialdatenbank rechnen. Diese Mischformen dienen vor allem der Unterscheidung zwischen reinen Materialdaten und Daten mit Anomalien, die potentiell einen Fingerabdruck darstellen (RoI).

*Berechnung der Feature-Vektoren auf weiteren Daten.* Sobald sich die Feature Vektoren (bzw. eine Auswahl derer) als geeignet für gewisse Anwendungszwecke, wie die Materialklassifikation oder RoI-Bestimmung eignen (AP 6 & 7), können die Feature-Vektoren auf weiteren Referenzdaten berechnet werden und sind damit Ausgangsmaterial für die Anfrageoptimierung in AP5.

## **M3.2 Synthetische Fingerabdrücke**

Bei der Erstellung einer geeigneten Datenbasis an Fingerabdrücken besteht das Problem der datenschutzrechtskonformen Speicherung und Verwendung der Daten (siehe hierzu Landesdatenschutzgesetz). Dazu gehört beispielsweise die Einwilligung der betroffenen Personen zu dem konkreten Verwendungszweck. Eine Lösung dieses Problems ist die Verwendung synthetischer (gedruckter) Fingerspuren.

---

### *Detail- und Komplettaufnahmen*

Derzeit existieren bei UNI-MD AMSL zwei Arten von synthetischen Spuren: zum einen Detailaufnahmen einzelner Papillarlinien vor allem zur Unterscheidung echter Fingerspuren von künstlichen, sowie zur Klassifikation von Oberfläche und Papillarlinien. Die zweite Spurensammlung umfasst die Aufnahme kompletter Abdrücke, vor allem zur RoI-Bestimmung. Beide Sammlungen werden den Referenzdaten hinzugefügt. Zukünftige Scans sollen folgende Eigenschaften besitzen:

Tabelle M3.3.3: Größen und Auflösung der Synthetischen Fingerspuren

Auflösung	Anwendungszweck
1,5cm x 1cm bei 50µm Punktabstand	<i>RoI-Bestimmung &amp; Klassifikation Ridge/Oberfläche:</i> Gutes Mittelmaß zwischen Scanzeit, Detailgrad & unerwünschten Druckartefakten.
1,5cm x 1cm bei 25µm Punktabstand	<b>Optional Erweiterung</b> zur 50µm Sammlung, wenn die Druckqualität verbessert wird, so dass weniger unerwünschte Druckartefakte auftreten.
1,5cm x 1cm bei 10µm Punktabstand	<b>Optional Erweiterung</b> wie 25µm
2mm x 2mm bei 2µm Punktabstand	<i>Klassifikation</i> echter bzw. synthetischer Abdruck z.B. aufgrund der Druckartefakte.

### *Umfang und Heterogenität der Datenbasis*

Insgesamt werden ca. 100 Samples erfasst werden, um eine aussagekräftige Datenbasis zu erschaffen. Derzeit ist das Drucken qualitativ hochwertiger Abdrücke nur auf Folie möglich. Sofern die technischen Problemen gelöst sind, soll die Sammlung auf die Materialien des pauschalisierten Tatortes erweitert werden.

### *Exemplarische Samples echter Abdrücke der AG-DB*

Zur Überprüfung der Ergebnisse der künstlichen Fingerspuren existiert eine kleine Sammlung echter Abdrücke der Projektmitarbeiter der AG-DB in verschiedenen Auflösungen & unterschiedlichen Materialien.

### *Alterungs- und Überlagerungssamples*

Derzeit ist es nicht geplant eine Basis aus echten oder überlagerten Fingerspuren für Alterung und Überlagerung anzulegen. Bei gedruckten Abdrücken ist derzeit nicht belegt, dass sich diese in Alterung oder Überlagerung analog zu echten Spuren verhalten. Weiterhin bestehen bei der Sammlung echter Abdrücke wieder rechtlich Probleme mit der Übergabe der Spuren von den Projektpartnern, so dass ein Aufbau einer Referenzdatenbasis allein aus den begrenzten Samples der AG-DB nicht sinnvoll erscheint.

---

### M3.2 Verwendung der Referenzdaten

Die Feature-Vektoren der Referenzdatensammlung dienen als zusätzlicher Input für die Anfrageoptimierungsforschung (AP 5). Bisher wurden hier verschiedene künstliche Verteilungen für die Daten im hochdimensionalen Raum erzeugt. Nach der Transformation der Daten in den reellwertigen Einheitswürfel  $[0-1]^n$  bzw. den diskreten Byte-Würfel  $[0-255]^n$ , bilden sie eine Erweiterung der Testsets mit unbekannter Verteilung.

AP 5 Testsznarien, -daten und Evaluierung

#### **Kernthemengebiete des AP 5:**

**(partielle) Ähnlichkeitssuche** - Das erste Ergebnis hierfür ist Definition des Variantenraumes (dies beinhaltet beispielsweise Anfragetypen und Distanzmetriken), der die zu betrachtenden Einflussfaktoren, basierend auf der aktuellen Literatur, umfassend beschreibt. Dies bildet die Basis für eine weitere zielgerichtete Erforschung von Einflussfaktoren und Lösungsansätzen zur Optimierung der Suchanfragen.

**Abstraktes Fingerspurenmodell & AFIS Anbindung (Interpol)** - Die Definition eines abstrakten Fingerspurenmodells in Form eines DB-Schemas für unsere Referenzdatenbank FiVe-DB (Fingerprint Verification Database) bildet eine notwendige Grundlage um im nächsten Schritt die eigentlichen Datenbasis erfassen und entsprechend auswerten zu können. Weiterhin wurde gezeigt, wie unsere Spurenmodell in das AFIS Austauschformat (Interpol-Spezifikation) übertragen werden kann und welche Unterschiede zwischen den Datenmodellen existieren.

**Optimierung der Suchverfahren** durch: Mehrdimensionale Indexverfahren, Co-processing und Speicherstrukturen

- Das erste Ergebnis ist eine umfangreiche empirische Evaluierung, welchen Einfluss der Einsatz verschiedenster mehrdimensionaler Indexstrukturen für verschiedene Ausprägungen des Variantenraumes (Abstraktion vom eigentlichen Anwendungsfalles aufgrund von Merkmalen, wie stochastischen Verteilung etc., zum Zwecke der Generalisierbarkeit der Ergebnisse) auf die Antwortzeit hat. Unsere Ergebnisse zeigen, dass die Wahl einer geeigneten Indexstruktur nicht trivial ist. So konnten sowohl Beschleunigungen im Bereich von Faktor 10 hoch 5, als auch Performanzeinbußen bis Faktor zwei im Vergleich zur üblichen Messgrundlage, dem sequentiellen Scannen aller Einträge, erreicht werden. Weiterhin wurde hierbei der Effekt von approximierenden Indexstrukturen in Bezug auf die Präzision der gelieferten Ergebnisse untersucht, wobei festgestellt wurde, dass, bezogen auf die Anwendungsfälle aus dem Digi-Dak Projekt, sich diese Indexstrukturen sehr gut für Datensätze eignen, bei denen der sogenannte „Fluch der hohen Dimensionen“ einen effizienten Einsatz klassischer Techniken nicht mehr erlaubt.
- Aufgrund der Schwierigkeit bei der Wahl einer geeigneten Indexstruktur bündelten wir unsere Erfahrungen, Erkenntnisse, Ergebnisse und Softwaregrundlagen in dem Evaluierungsframework QuEval. Es ermöglicht, auf Basis einer abstrakten Beschreibung des Anwendungsfalles (was einer Ausprägung unseres Variantenraumes entspricht), die optimale Indexstrukturimplementierung effizient und mit stochastisch valide zu ermitteln. Darüber



---

hinaus offeriert QuEval erstmals die Möglichkeit neuartige Indexstrukturen gegen die derzeit besten Verfahren umfassend zu evaluieren.

- Wir konnten darüber hinaus zeigen, dass die Verwendung von Trends aus der modernen Hardware, wie Multi-Core Architekturen GPU-Coprocessing signifikante Verbesserungen in Bezug auf die Antwortzeit hat. Dies zeigt sich vor allem in Verbindung mit modernen Speicherstrukturen, wie der spaltenweisen (im Gegensatz) zu der klassischen zeilenbasierten Speicherung und Verarbeitung.
- Unsere Studierenden konnten die von uns erzielten Grundlagenforschungsergebnisse erfolgreich in mehreren internationalen Wettbewerben einsetzen. Hervorzuheben ist hierbei vor allem der SIGMOD Programming Contest.

### ***Publikation des Kernergebnisses aus AP 5:***

Publiziert wurde das Ergebnis im Rahmen der CAiSE-Konferenz 2012 [VLDB14]. Die VLDB gehört zusammen mit der ACM SIGMOD zu den beiden am höchsten gerankten Datenbankkonferenzen mit einem H5 Index von 66 (Google Scholar). Somit kann die Publikation als außerordentlicher Erfolg des Teilprojektes angesehen werden.

### ***Eingehende Erläuterung zum Meilensteinen des AP 5***

Nachfolgend erfolgt eine Darstellung der Ergebnisse bezogen auf die einzelnen Meilensteine des AP 5 (M5.1), für die wir als Verbundpartner verantwortlich waren.

### ***Darstellung der Ergebnisse für M5.1 Schnittstellendefinition Datenvorverarbeitung zum AP 5***

Die grundlegende Problemstellung und Komplexität der damit verbundenen Herausforderungen sind in [GVD12] beschrieben. Generell ist hierbei zu bemerken, dass bisherige Evaluierungen nur auf einzelne Teilaspekte eingehen und somit untereinander schwer vergleichbar sind [Tob13]. Daher schließen wir mit unserem Evaluationsframework QuEval [A] genau diese Lücke und konnten das Framework und die dazugehörigen Ergebnisse auch auf einer der beiden hochrangigsten Datenbank-Konferenzen veröffentlichen [VLDB14]<sup>5</sup>. QuEval und seine Erweiterungen [B] bilden die Basis für unsere Zuarbeiten zum M4. Die einzelnen Evaluierungen sind dabei in [ASB12] veröffentlicht.

---

5 Die Proceedings werden im PVLDB als Journal veröffentlicht.

---

## ***AP 8 Testszzenarien, -daten und Evaluierung***

### ***Kernthemengebiete des AP 5:***

#### ***Eingehende Erläuterung zum Meilenstein des AP 8***

Nachfolgend erfolgt eine Darstellung der Ergebnisse bezogen auf die einzelnen Meilensteine des AP 8 (M5.8), für die wir als Verbundpartner verantwortlich waren.

#### *Darstellung der Ergebnisse für M8.2 zum AP 8*

Der Meilenstein (MS) 8.2 Schnittstellenzusammenfassung enthält die Zusammenfassung der Schnittstellen aus den Vorarbeiten MST 3.1 & MST 5.1. Weiterhin sind Zusätze, die sich durch Modifikation der Anforderungen z.B. aus neuen Forschungsergebnissen resultieren, enthalten. Die Zusammenfassung ist eine der wesentlichen Vorarbeiten, die für Integration der DB über die entsprechenden Schnittstellen in den Demonstrator (PM) notwendig ist. Zusätzlich werden die Schnittstellen sowohl für die Integration von Daten aus der Datenerfassung (AP 1), als auch für Weitergabe der enthaltenen Daten an die Datenvorverarbeitung & Mustererkennung (AP5 & AP6) benötigt. Weiterhin müssen die Anforderungen an den beweissichere Datenhaltung (AP8), sowie die verfassungs- & datenschutzrechtskonforme Gestaltung (AP10) berücksichtigt werden. Nachfolgend werden die Schnittstellen und deren Bezug zu den einzelnen AP aufgeführt und erläutert.

#### **M8.2: Gliederung**

In den nachfolgenden Abschnitten werden die Einzelbausteine der Schnittstellenzusammenfassung erläutert. Begonnen wird mit notwendigen Vorarbeiten, bevor die grundlegenden Entwurfsprinzipien vorgestellt werden, aus denen sich die eigentliche Architektur ergibt. Die Umsetzung der der Schnittstelle zu demonstrations- (zuarbeit PM) und forschungszwecken erfolgt im *DD-Client*. Abschließend werden mögliche Erweiterungen der Schnittstellen diskutiert.

#### **M8.2: Notwendige Vorarbeiten**

Der folgenden Abschnitt erläutert direkt notwendige Vorarbeiten der vorliegenden Schnittstellenzusammenfassung.

#### *AP 3.1 Datenerfassungsschnittstelle (UNI-MD-DB)*

Im MST 3.1 wurde die Datenerfassungsschnittstelle definiert. An der Schnittstelle wurden seit Erstellung des MST3.1 keine grundlegenden Änderungen vorgenommen. Sie wird daher wie in MST 3.1 beschrieben in MST 8.2 übernommen.

Die Datenerfassungsschnittstelle (Abb. M8.2.a) definiert welche Daten bei der Erfassung von Spuren aufgenommen werde, wie diese miteinander verknüpft sind. Zusätzlich ist festgehalten welche Operationen auf den Daten ausgeführt werden dürfen (AP 8), was letztendlich in einem feingranularen Rechtemodell mündet. Beispielsweise dürfen die eigentlichen Scan-Daten (die Bilddaten) nur in die Datenbasis aufgenommen und anschließend nicht mehr verändert werden. Ebenfalls ist der lesende Zugriff über diese Schnittstelle zu unterbinden (AP 10). Im Gegensatz dazu muss eine Möglichkeit geschaffen werden auf Meta-Daten zuzugreifen und eine wohldefinierte Menge derer über entsprechende Funktionen zu modifizieren.

---

### *AP 5.1 Schnittstelle Datenvorverarbeitung (UNI-MD-DB)*

Im MST 5.1 wurde die Schnittstelle zur Datenvorverarbeitung definiert. Analog zu M3.1 wurden auch an dieser Schnittstelle seit Erstellung des MST5.1 keine grundlegenden Änderungen vorgenommen. Sie wird daher wie in MST 5.1 beschrieben in MST 8.2 übernommen.

Die Datenerfassungsschnittstelle (Abb. M3.1.a) definiert Art, Umfang, Struktur und Granularität der Daten, welche an die Forschungsprototypen der Datenvorverarbeitung weitergegeben werden. An dieser Stelle muss vor allem auch festgelegt werden, wer unter welchen Umständen auf welche Daten zugreifen kann. Dies wird, wie in M3.1, in einem feingranularen Rechtemodell festgehalten. Zusätzlich zum Austausch der eigentlichen Nutzdaten (Bild-Daten), erfolgt derzeit über diese Schnittstelle die Verwaltung der Proben, sowie die Weitergabe der Klimadaten, vor allem für die Alterungsdetektion (AP 4).

### *Generelles Austauschformat (FH BRB)*

Vor dem Hintergrund eine möglichst einheitliche Schnittstellenlandschaft unter Berücksichtigung der Anforderungen der beweissicheren Datenhaltung zu schaffen, wurde von der FH BRB ein universelles Austauschformat (Container) speziell für den Fingerabdruckskontext entwickelt [KVL11].

Dieses XML-basierte Format, dient als Container in dem die einzelnen Datenobjekte transportiert werden können. Durch die Verwendung von Signaturen kann (mit einem bestehenden Restrisiko) die Authentizität und Integrität der enthaltenen Daten überprüft werden.

### *Weitere Vorarbeiten*

Weitere Vorarbeiten, wie beispielsweise die Vorgehensmodelle (UNI-MD-AMSL) auf denen die Trennung in die einzelnen Datenobjekte beruht, werden an dieser Stelle nicht explizit erwähnt, sondern auf die MST3.1 und MST 5.1 verwiesen.

## **M8.2: Grundsätzliche Entwurfsprinzipien**

Beim Design der Schnittstelle müssen die folgenden drei grundlegende Entwurfsprinzipien, die sich quer durch die gesamte Architektur ziehen. Die Prinzipien und deren Auswirkungen bzw. Notwendigkeit werden nachstehend erläutert.

### *Prinzip der Datensparsamkeit und angepasstem Funktionsumfanges*

Das Prinzip der Datensparsamkeit und der zweckbezogenen Datenerhebung sind grundlegende Prinzipien des Datenschutzes (siehe hierzu auch MST 8.1). Daher darf die Schnittstelle nur in der Lage sein Daten zu erheben, die für den gegebenen Anwendungszweck benötigt werden. Eine Basis für die Bestimmung der Notwendigkeit bieten die Vorgehensmodelle von UNI-MD-AMSL und die Aufgabenstellungen der einzelnen Verbundpartner.

Daraus ergibt sich letztendlich die Forderung nach Varianten der Schnittstelle, wie auch schon in MST 3.1 angedeutet. Zusätzlich erlaubt es dieses Vorgehen die Schnittstellen aus MST 3.1 und MST 5.1 zu einer Schnittstelle zusammenzufassen.

---

### *Prinzipien der IT Security bezüglich Authentizität, Integrität und Vertraulichkeit der Daten*

Die Sicherstellung der Authentizität, Integrität und Vertraulichkeit der verwendeten Daten sowie und die Bestimmung des Restrisikos, stellen eine fundamentale Prinzipien dar. Sie ergeben sich sowohl aufgrund des Schutzwertes der verwendeten personenbezogenen Daten (AP10), als auch durch die Sicherung des Beweismertes der eigentlichen Fingerspur.

Die Maßnahmen zur Sicherung wurden bereits in MST 3.1 erläutert und werden, wie in Abschnitt 3.2.3 beschrieben, durch das universelle Austauschformat unterstützt.

### *Erweiterbarkeit der Schnittstelle in Bezug auf neue und bzw. sich ändernde Anforderungen*

Ein weiteres grundlegendes Prinzip ist die Forderung nach der Erweiterbarkeit der Schnittstelle. Aufgrund der zu erwartenden Änderungen der Anforderungen durch neue Forschungsergebnisse muss die Schnittstelle in der Lage sich sich mit möglichst wenig Aufwand, flexibel an neue Gegebenheiten anzupassen.

Zusammen mit der 3.3.1 ergibt sich hieraus, dass zum einen *zusätzlich* benötigte Funktionen hinzugefügt werden können, nicht (mehr) benötigte entfernt und bestehende ggf. modifiziert werden.

Insgesamt erzeugen diese drei Prinzipien hohe Anforderungen an die Architektur und die spätere Umsetzung der Schnittstelle im DD-Client. So konnten wir, in Zusammenarbeit mit der METOP, bezüglich der Modifizierbarkeit der Schnittstelle auf DB-Ebene sowie der Abstimmung entsprechender Client-Anwendung (in diesem Fall die Umsetzung der Schnittstelle DD-Client) auf das dahinter liegende DB-Schema eine Forschungslücke identifizieren und eine Lösungsvorschlag publizieren [SLSK11].

## **M8.2: Architektur**

Die Architektur der Schnittstellenzusammenfassung ergibt sich aus den Anforderungen der Grundprinzipien des vorhergehenden Abschnittes sowie den bestehenden Schnittstellen der Vorarbeiten, die in dieser zusammengefasst werden.

Nachfolgend werden zuerst Programmfamilien als Lösung zur Integration der Grundprinzipien und zur Zusammenfassung bestehenden Schnittstellen eingeführt. Anschließend wird die Modellierung der Schnittstelle und die zugehörigen kombinierbaren Funktionen (Feature) erläutert.

### *Programmfamilien zur Integration der Grundprinzipien*

Programmfamilien stellen ein Sammlung von Techniken zur Erzeugung von maßgeschneiderten Softwareinstanzen aus einer Basis an gemeinsam genutzten Merkmalen dar [CN00,BSR04]. Ein Feature stellt in diesem Fall eine Funktionszusatz dar (z.B. zum Speichern Sensor Daten), der für eine beteiligte Partei (bei uns Verbundpartner) von Interesse ist. Durch Grundlagenforschung zur Übertragung der auf Anwendungsseite bekannten Techniken auf den DB-Bereich, und somit auf die Schnittstellenzusammenfassung, konnten wir die Machbarkeit dieses Ansatzes nachweisen [SLSK11]. Weiterhin konnte wir in dieser Veröffentlichung zeigen, dass durch dieses Vorgehen sich Varianten mit vergleichsweise geringem Aufwand erstellen lassen. Weiterhin ergeben sich Vorteile in Bezug auf die Modifikation bestehender Varianten gegenüber traditionell verwendeten Ansätzen. Somit eignet sich dieses Vorgehen besser zur Integration der Grundprinzipien in diese zusammengefasste Schnittstelle.

---

## M8.2: Modellierung der Schnittstelle

Der erste Schritt bei der Entwicklung einer Programmfamilie besteht darin die einzelnen Variationspunkte zu identifizieren. Das bedeutet, dass die einzelnen Feature, die miteinander kombiniert werden können bestimmt werden müsse. Hierzu muss festgelegt werden, auf welche Art und Weise die Bestimmung der einzelnen Feature erfolgt.

Zur Sicherung der Abdeckung möglichst vieler individueller Anforderungen der Projektpartner und der Berücksichtigung einer mit geringem Aufwand verbundenen Integrierbarkeit einzelner Konzepte in den Demonstrator, richtet sich unsere Bestimmung der Feature vor allem nach:

1. Vorgehensmodellen (UNI-MD-AMSL)
2. Container-Spezifikation (FH BRB)
3. Informationen zum Spurensicherungsprozess aus Meetings (LKA-ST)
4. Aufbau & Konzept der ersten Demonstrators (METOP)
5. Zusätzlichen Arbeiten (vor allem UNI-MD-AMSL)

### *Kombinationsmöglichkeiten der Feature - Feature Modell*

Nach der Bestimmung der einzelnen Feature, erfolgt die Definition der Beziehungen der Feature untereinander. Hierbei wird festgelegt welche Feature miteinander kombiniert werden können. Die grafische Darstellung der Schnittstellenzusammenfassung in Abb, 2 erfolgt im Feature Modell [CN00].

Das Feature Modell der Schnittstellenzusammenfassung (Abb. 2) enthält eine Menge an obligatorischen Feature, die in jeder Variante der Schnittstelle (auch genannt FiVe DB) enthalten sind, sofern das Elternmerkmal gewählt wurde. So ist das Merkmal Data Storage in jeder Variante enthalten. Weiterhin existieren eine Anzahl an optionalen Merkmalen, die je nach Bedarf in den Varianten enthalten sein kann. Zusätzlich können ein oder mehrere Feature, die über eine *oder*-Verbindung verknüpft sind, in einer Variante enthalten sein. Daher ist es möglich alle drei Interface Typen (Forensic File Formats, Common SQL und Watermarks) in einer Variante, z.B. zu Evaluierungszwecken, zu integrieren.

Anhand des Feature Modells lassen sich die validen Varianten (Instanzen der Schnittstelle mit DBMS System mit unterschiedlichen Features) ableiten. Auf Basis dieser Information können dann die einzelnen Varianten erstellt werden.

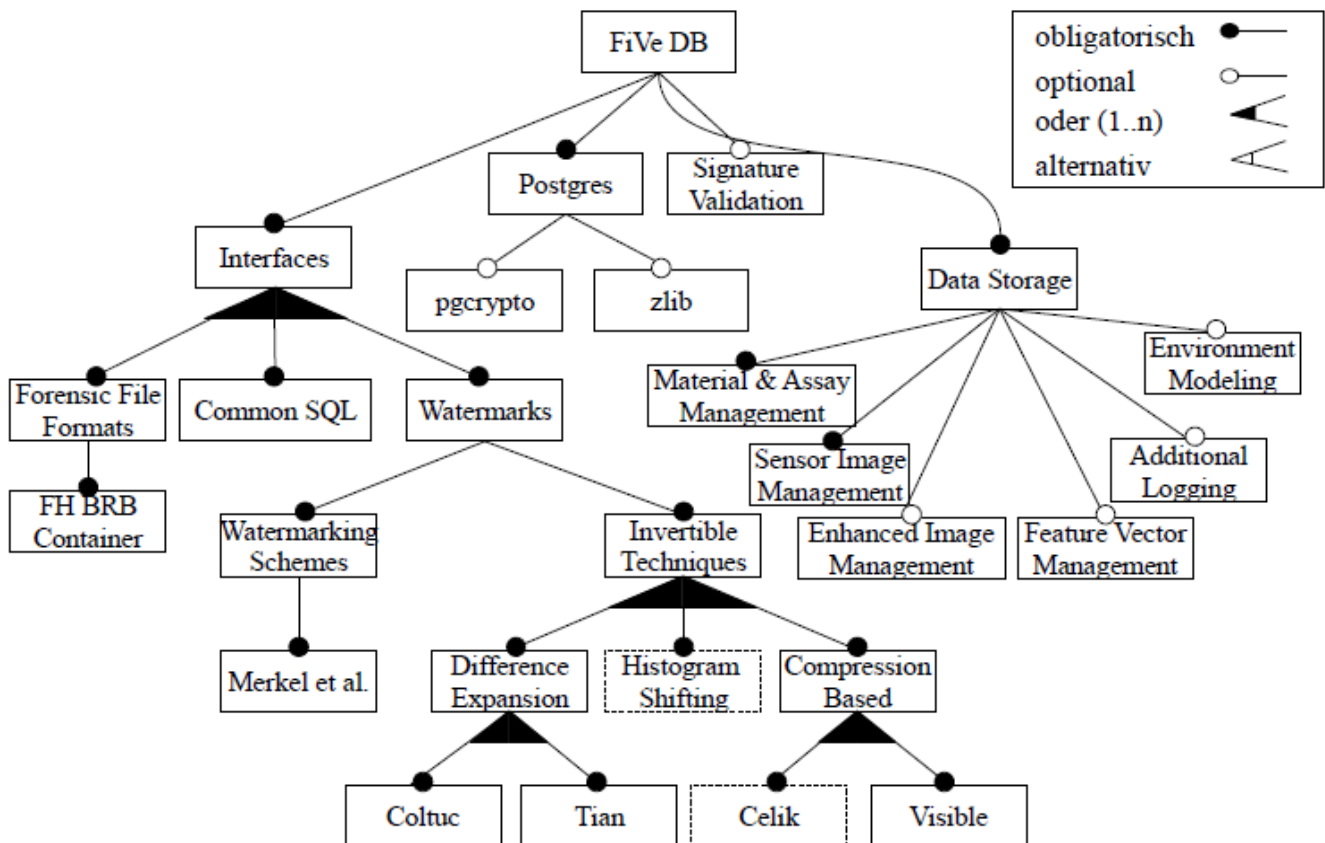


Abbildung 2: Feature Modell der Schnittstellenzusammenfassung (adaptiert aus [BNCOD11])

#### Abhängigkeiten der einzelnen Feature

Zusätzlich zu den Einschränkungen im FM existieren weitere Limitierungen, die die Kombinationsmöglichkeiten einschränken (siehe Abb. 3). So benötigt beispielsweise der Einsatz des universellen Austauschformats (FH BRB Container) sowohl eine Möglichkeit zum entpacken des komprimierten Datenstromes (zlib), als auch eine Funktion zum Validieren der Signaturen, die eingesetzt werden um die Integrität und Authentizität der Daten zu gewährleisten.



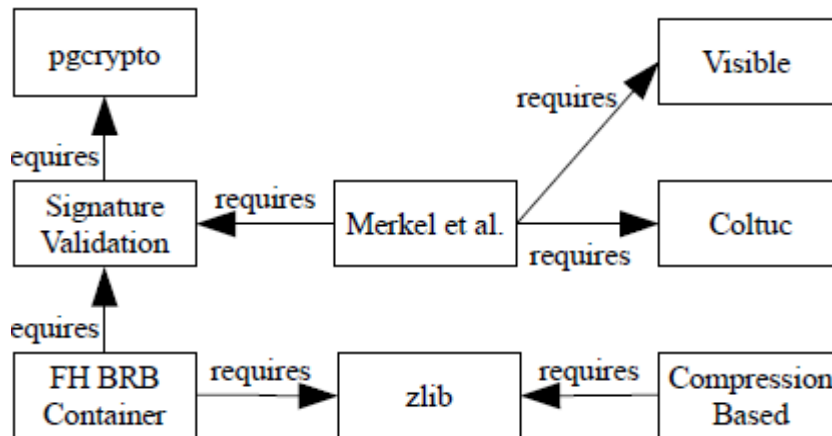


Abbildung 3: Abhängigkeiten der Feature

## M8.2: Beschreibung der einzelnen Feature

Es erfolgt nun eine Beschreibung zu den einzelnen Feature der Schnittstelle. Die Beschreibung orientiert sich im allgemeinen an der Baumstruktur des Feature Modells.

**Konzeptknoten FiVe DB:** Hierbei handelt es sich lediglich um den Konzeptknoten der Programmfamilie FiVe DB (Finger Print Verification Database). Alle weiteren Merkmale dieser Familie sind Kindknoten dieses Elementes.

**Data Storage:** Hierbei handelt es sich um die Basisinfrastruktur für die einzelnen *Datenspeicherungsmerkmale* basierend auf dem Vorgehensmodell von UNI-MD-AMSL.

**Material & Assay Management:** Zu diesem Merkmale gehören grundlegende Funktionen zur Aufnahmen und Verwaltung von Material- und Probanddaten. Das Merkmal ist in jeder Variante enthalten und die Aufteilung in die einzelnen Datenobjekte entspricht den MST 3.1 und MST 5.1.

**Sensor Image Management:** Die Verwaltung der original Sensorbilder ist mit der beweisssicheren Speicherung der Sensoraufnahmen und deren dedizierte Weitergabe an die Vorverarbeitung (siehe auch MST 5.1) zuständig. Dazu gehört beispielsweise ein feingranulares RechteManagement.

**Enhanced Image Management:** Das optionale Merkmal zur Verwaltung der Vorverarbeitungsergebnisse die Speicherung und den Austausch von Vorverarbeitungsergebnissen zwischen der DB und der DD-Infrastruktur.

**Feature Vector Management:** Die Speicherung der Merkmalsvektoren (Bildbeschreibungen) wird durch dieses optionale Merkmal organisiert.

**Additional Logging:** Legt zusätzliche Trigger, Log-Tabellen an und erhöht das Log Level um die einzelnen Arbeitsschritte in der biometrischen Kette detailliert nachvollziehen zu können. Dies beinhaltet beispielsweise auch mögliche Zugriffsverletzungen bzw. deren Versuch.

**Environment Modeling:** Das optionale Merkmal zur Erfassung der Umgebungsbedingungen erlaubt, vor allem für die Alterungsdetektion (AP4), die Aufnahmen der Umgebungsdaten.

**DBMS spezifische Merkmale:** Die Merkmale Postgres, pgcrypto, zlib sind DBMS spezifisch (für die Übertragung in andere DBMS siehe folgenden Abschnitt). In unserer Architektur verwenden wir derzeit ein Postgres DBMS in Version 8.4. Die Merkmale *pgcrypto* und *zlib* sind dabei frei

---

verfügbare optionale Merkmale des DBMS und wurde nicht von uns erstellt. Sie werden von einigen Interface Merkmalen z.B. für kryptografische Funktionen benötigt. Das Merkmal *Signature Validation* erweitert die *pgcrypto* Bibliothek zusätzlich um die Möglichkeit zum validieren von Signaturen.

**Interfaces:** Die Gruppe der Interface Merkmale stellt die eigentliche Schnittstelle dar. Diese Merkmale erlauben den Austausch von Daten zwischen Forschungsprototypen der DD-Infrastruktur und Datenhaltungsmerkmalen unter Sicherstellung IT-Security Anforderungen. Hierbei verwenden die einzelnen Merkmale unterschiedliche Techniken und verfügen u.U. über unterschiedliche Restrisiken.

**Forensic File Formats:** Zu der Gruppe der forensischen Datenaustauschformate gehört derzeit ausschließlich das Container-Format der FH BRB. Es sind jedoch zusätzliche Formate wie AFF denkbar [KVL11], um die Vorteile des Container-Formats für unseren Anwendungskontext evaluieren zu können.

**Common SQL:** Dieses Merkmal dient uns *ausschließlich* als Vergleichsparameter um beispielsweise Performanceunterschiede im Vergleich zu den forensischen Datenaustauschformaten ermitteln zu können. Es sei hier explizit drauf hingewiesen, dass durch dieses Interface keine zusätzlichen Maßnahmen zur Sicherung der Integrität und Authentizität getroffen.

**Watermarks:** Die Merkmale der Wasserzeichengruppe unterteilen sich in *Einbettungstechniken* mit deren Hilfe invertierbare Wasserzeichen in die Datenobjekte eingebettet werden können und *Wasserzeichenschemata* die aus einer Wahl an Einbettungstechniken ein Wasserzeichen mit individuellen Eigenschaften erzeugen. Zu den derzeit verfügbaren Techniken gehören die Ansätze von Coltuc [CC07], Tian [Tia03], einfache verlustlose Kompression. Der Ansatz von Celik [CSTS05] ist derzeit noch in Arbeit. Derzeit wird ein komplexes Wasserzeichenschem (jede Technik ist ein simples Schema) verwendet, nämlich das Schema von Merkel et al. [MKDV11], dessen Anwendung wir im DB-kontext als Teil einer Veröffentlichung demonstrieren konnten [SSM+11].

### *Übertragung in andere DBMS*

Die Übertragung in andere DBMS außer Postgres ist möglich, jedoch mit Umbauarbeiten, sowohl auf konzeptioneller Ebene, als auch auf Seiten der Implementierung verbunden. Es müssen beispielsweise äquivalente Merkmale zu den Merkmalen *pgcrypto* und *zlib* gefunden werden. Auf Anwendungsebene müssen die Interfacemerkmale, die zu wesentlichen Teilen aus Prozeduren bestehen (siehe auch 3.5) in die interne Sprache des neuen DBMS übersetzt werden.

### *Schnittstellen ohne Beteiligung der DB*

Sollen Daten in der DD-Infrastruktur *ohne Beteiligung* der Datenbank ausgetauscht werden, schlagen wir analog die Nutzung des Containers und der entsprechenden Objekte der DB vor, wie sie in der DB und in der ersten Version des Demonstrators verwendet werden. Dies verringert den Aufwand, falls spätere Ergebnisse (mit zugehörigem Datenstammbaum) anschließend persistent in der DB gespeichert werden.

## **M8.2: Umsetzung der Schnittstelle**

In diesem Abschnitt wird die Umsetzung der Schnittstelle beschrieben. Als erstes wird zu diesem Zweck erörtert wie einzelne Varianten aus der Modellierung erstellt werden. Danach wird Implementierung der Schnittstellenzusammenfassung im DD-Client Forschungsprototypen als

---

proof-of-concept betrachtet. Abschließend erläutern wir, wie die Schnittstellenzusammenfassung in weitere Forschungsprototypen der Infrastruktur integriert wird.

### *Erstellung von Varianten*

Bei der Verwendung von Programmfamilien muss eine Strategie gewählt werden, wie aus der Modellierung mit bestehenden Softwareartefakten (Implementierungsbausteinen) basierend auf der Feature-Auswahl eine konkrete Variante der Schnittstelle erstellt wird.

In unserem Anwendungsfall kann die FiVe DB am besten als Plugin-basiertes Framework angesehen werden. So weit möglich sind die einzelnen Merkmale in unterschiedliche DB-Schemata gekapselt und damit physisch getrennt. Dies hat den Vorteil, dass Daten beispielsweise auf physisch unterschiedlichen Table-Spaces geschrieben werden können und die Rechtevergabe dies berücksichtigt.

### *Modifikation einer bestehenden Variante*

Das hinzufügen eines solchen Merkmals erfolgt durch ausführen eines entsprechenden create scripts, das entfernen durch ein analoges delete script. Somit können Merkmale hinzugefügt (vorausgesetzt die Abhängigkeiten werden berücksichtigt) und auch wieder entfernt werden. Beim entfernen ist zu beachten, dass sich aufgrund der Merkmalskombinationen einige Merkmale u.U. nicht ohne vorherige Entfernung anderer löschen lassen. So kann das Merkmal *enhanced image management* ausschließlich dann entfernt werden, wenn nicht auch das Merkmal *feature vector management* vorhanden ist. Grund hierfür sind bestehende Referenzen (Fremdschlüssel) zwischen den Merkmalen. In so einem Fall verbietet das DBMS die Löschung selbständig, aufgrund der Reihenfolge der Löschkommandos. Die Grundlagenforschung zu diesem Thema, auch unter der Berücksichtigung von Datenschutzrechtsanwendung, stellt ein Alleinstellungsmerkmal dieses Projektes dar, der weiteren Raum für Forschung bietet.

## **M8.2: Implementierungsdetails am Beispiel des Forschungsprototypen DD-Client**

Bei der Implementierung der eigentlichen Softwareartefakte unterscheiden wir auf DB-Seite zwei verschiedene Merkmalstypen:

- dynamische Merkmale, die im wesentlichen aus Prozeduren innerhalb des DBMS bestehen (z.B. Signature Validation, Merkel Watermark)
- statische, die hauptsächlich aus Tabellen und wenig dynamischen Elementen bestehen (vor allem die Elemente der Datenhaltungskomponenten)

Beide Merkmalstypen können über eine entsprechende Verbindung (z.B. via JDBC oder ODBC) via SQL angesprochen werden. Im DD-Client wird die Verbindung via JDBC aufgebaut und die Schnittstelle anschließend

### *DB Services zur Kapselung*

Das Verbergen von Funktionalitäten der Schnittstelle hinter Funktionsaufrufen verfolgt zweierlei Ziele:

- 
- Vereinfachen der Integration der Schnittstelle in die verschiedenen Prototypen, da nicht die gesamten SQL-Kommandos generiert werden müssen, wofür detailliertes Schemawissen von Nöten ist.
  - Verbergen von Implementierungsdetails & Typsicherheit im Funktionsaufruf erschweren die Manipulation der Anfragen (z.B. SQL-Injektion).

Die Funktionen der Schnittstelle wurden von uns als proof-of-concept in DD-Client integriert. Nachfolgend werden exemplarische Beispiele mit entsprechender Visualisierung präsentiert.

## **M8.2: Integration der Schnittstelle in weitere Forschungsprototypen**

In diesem Abschnitt beschreiben wird zwei Integrationsmöglichkeiten der Schnittstellen-zusammenfassung in weitere Forschungsprototypen.

### *Ansprechen der Schnittstelle über DD-Client*

Es ist vorgehen, die Schnittstelle über DD-Client via Kommandozeile mit entsprechenden Parametern benutzen zu können (analog zur Container-Validierung). Dazu enthält der Prototyp ein Downloadverzeichnis indem heruntergeladenen Dateien (z.B. Container) gelagert werden können, auf die dann weiter zugegriffen werden. Die konkrete Erweiterung würde bei entsprechendem Bedarf der Projektpartner implementiert werden.

### *Integration der Schnittstelle*

Die zweite Möglichkeit ist mit Aufwand verbunden, erlaubt es jedoch zum einen individualisiert auf Ergebnisse der Schnittstellenzusammenfassung zu reagieren und Daten zu verarbeiten, die nicht in Dateiform vorliegen. zu diesem Zweck sind folgende Schritte notwendig:

#### Schritt 0: Vorbereitung

- Netzwerkverbindung zur Datenbank (normalerweise Port 5432) mit gewünschten Funktionen
- Anlegen von Nutzer/PW Paaren in der entsprechenden DB & Erteilen der *notwendigen* Rechte

#### Schritt 1: Aufbauen der Verbindung zur DB

- Einbindung des PostgreSQL Treibers für Version 8.4 über JDBC<sup>6</sup> (Java) oder ODBC (sonst), dieser ist für nahezu alle gängigen Sprachen verfügbar
- Instanzierung eines Objektes des Treibers mit URL zum DB Server, Port, Nutzername und Passwort (Beispiel in der Implementierung der DB Klasse im DD-Client zu finden)

#### Schritt 3: Generierung des SQL-Kommandos und Funktionsaufrufe

- Die SQL Kommandogenerierung bzw. Funktionsaufruf aus DD-Client kopieren
- Anbinden der Containerbibliothek über Kommandozeile
- Absenden der Kommandos an die DB über den Instanz des Treiberobjektes

---

6 <http://jdbc.postgresql.org/download.html>

---

(optional) Schritt 4 Visualisierung des Anfrageergebnisses

- Aufbau von entsprechenden Tabellen (Konzepte könne teilweise aus DD-Client übernommen werden)
4. Visualisierung von Bilddaten (nicht in DD-Client vorhanden)

### **M8.2: Zusammenfassung**

Im diesem Meilensteinbericht wurde der Aufbau und die Funktionsweise der Zusammenfassung der bisherigen Einzelschnittstellen erläutert. Besonderes Augenmerk wurde hierbei auf die datenschutzrechtskonforme Gestaltung und die beweissichere Datenhaltung gelegt.

#### *Darstellung der Ergebnisse für M8.3 zum AP 8*

Im Gegensatz zum M8.2. wurden die Ergebnisse M8.3 in eigenen Publikationen veröffentlicht. Kernergebnis ist der Vorschlag eines neuen Ansatzes zur querschnittenden Integration des Provenance Belangs in bestehende Systeminfrastrukturen [TAPP12]. Zu diesem Zweck haben wir ein Hierarchisches Provenance Framework erforscht, welches eine Großteil der derzeit bekannten Ansätze zur Extraktion von Provenance Daten zusammenfasst [FIN12]. Die Ergebnisse werden derzeit aufbereitet und sollen zeitnah als Journalpublikation ([DKE14]) veröffentlicht werden

### **Gegenüberstellung der Ergebnisse vorgegebenen Zielen**

Die Nachprüfbarkeit der Erfüllung der vorgegebenen Ziele zeigt sich am besten in der Erfüllung der einzelnen Meilensteine bzw. in Form der Zuarbeiten zu die AP-spezifischen und teilprojekt-übergreifenden Meilensteine für die wir verantwortlich waren. Diese haben wir nachfolgend in tabellarischer Form aufgelistet. Die Erfüllung kann dabei in zwei verschiedenen Formen erfolgen. Eingehende Beschreibung der Ergebnisse im Abschlussbericht für die grundlegenden Meilensteine M3.1, M3.2 und M8.2, während für M5.1, M8.3 und M4. Eine Besonderheit bildet das AP 3. Hier sind beide Meilensteine grundlegender Natur. Allerdings wurden fehlende grundlegende Technologien identifiziert, für die ein ganzheitlicher Ansatz in [CAISE12] als Abschluss des APs publiziert wurde.

<b>AP</b>	<b>Meilenstein</b>	<b>Grad der Verantwortlichkeit</b>	<b>Kernergebnis bzw. Publikation(en)</b>
AP 3	<b>M3.1</b>	verantwortlich	wie beschrieben
	<b>M3.2</b>	verantwortlich	wie beschrieben
AP 3 Abschlusspublikation			[CAISE2012]
AP 5	<b>M5.1</b>	verantwortlich	[GVD12] [A] [B] [C]
AP 8	<b>M8.2</b>	verantwortlich	wie beschrieben
	<b>M8.3</b>	verantwortlich	[FIN12], geplant [DKE14]

Teilprojekt-übergreifende Meilensteine		
M4	verantwortlich	wesentliche Ergebnisse als Zusatz zur ersten Fassung publiziert in [CAISE2012] aus AP3 in [ASB12] und [VLDB14] aus AP 5 in [TechReport] aus AP 8 in [BTW] aus AP 10

## II.2 Wesentliche Positionen des zahlenmäßigen Nachweises

Die anfallenden Ausgaben erfolgten vor allem im Bereich Personal und Reisekosten. Die wesentlichen einzelnen Posten werden in der folgenden Tabelle mit der zugehörigen Kennung laut AZA dargestellt.

Position	Ausgaben
0811 Beschäftigte E-12 bis E 15	320.108,62 €
Sonstige Beschäftigungsentgelte (HiWi)	22.701,56 €
Dienstreise	8.884,31 €
Investitionen von mehr als 410€ im Einzelfall	5.916,32 €

## II.3 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die Notwendigkeit der geleisteten Arbeiten ergibt sich direkt aus den Planung der Arbeiten des Teilprojektes laut Einzelantrag und Verbundskizze. Dies sind die Kernbeiträge des Teilprojektes zur angestrebten Gesamtlösung eines „rechtskonformen und automatischen Systems“. Dieses System adressiert direkt die förderpolitischen Ziele der Bundesregierung, welche im zugrundeliegenden Call formuliert sind. Im Einzelnen sind dies die (a) generalpräventive und (b) positive Wirkung für das subjektive Sicherheitsempfinden in der Gesellschaft. Somit ergibt sich die Notwendigkeit der geleisteten Arbeit transitiv aus den förderpolitischen Zielen der Bundesregierung.

Die Angemessenheit der geleisteten Arbeit ist zum einen durch die vollständige Erfüllung der gesetzten Ziele, wie Abschnitt II.1 eingehend erläutert, und zum anderen durch die Einhaltung des Kostenplans begründet.

## II.4 Fortschreibung des Verwertungsplans

### Erfindungen und Schutzrechtsanmeldung

Wie in der Anlage „Anzeige von Schutzrechtsanmeldungen und Erfindungen“ dargestellt, wurden im Rahmen diese F+E Vorhabens von uns keine Erfindungen oder Schutzrechte angemeldet.



---

### ***Wirtschaftliche Erfolgsaussichten nach Projektende***

- Untersuchungen zur Optimierung des Demonstrators einschließlich Erweiterungen im Bereich der Datenvorverarbeitung, um das Aufwand-Nutzen-Verhältnis für den Endanwender zu optimieren
- Forschungsk Kooperationen mit der Wirtschaft insbesondere mit PROVET, METOP und FH-BRB im Hinblick auf angewandte Forschung im Bereich der Verfassungs- und Datenschutzrechtlich konformen Gestaltung von DV-Systemen
- Übertragung der Projektergebnisse im Bereich der beweissicheren Datenhaltung auf andere Gebiete mit ähnlichen Sicherheits- und Beweisbarkeitsanforderungen, wie z. B. die Bankbranche

### ***Wissenschaftlich und/oder technische Erfolgsaussichten nach Projektende***

- Weitererforschung abstrakter Datenmodelle, die auch für andere Multimediadaten gültig sind
- Erforschung einer Datenbankinfrastruktur, bestehend aus einer Referenzdatenbank mit definierter Datenbankschnittstelle, die als Grundlage für sämtliche Datenhaltungs-, -verarbeitungs- und -austauschprozesse beim Aufbau des Digi-Dak Demonstrators verwendet wird und auch von anderen Systemen (z. B. AFIS) genutzt bzw. in andere Systeme integriert werden kann
- Zusammenführung und Erweiterung des Know-hows der Vorgehensmodellen
  - für die digitale Fingerspurenerfassung und -auswertung zur Optimierung der Datenvorverarbeitung
  - für die beweissichere Datenhaltung in Hinblick auf die Nachvollziehbarkeit von Verarbeitungsschritten und juristische Verwertbarkeit
- Erweiterung der Erkenntnisse zu Datenschutzkonformität in der IT-Forensik
- Anwendung der Projektergebnisse aus dem Bereich der Datenvorverarbeitung auch für andere Forschung, z. B. Benutzerauthentifizierung
- Ausbau des Forschungsspektrums der Universität
  - im Bereich der Vorverarbeitung von forensischen und biometrischen Daten
  - im Bereich „Sichere Datenhaltung“, insbesondere Protokollierung von Verarbeitungsschritte sowie Datensicherheit und Datenschutz

### ***Wissenschaftlich und/oder technische Anschlussfähigkeiten für eine mögliche nächste Phase***

- Grundlagenforschung: weiterführende Kooperationen von Wissenschaft und Wirtschaft aufbauend auf dem Digi-Dak Demonstrator mit den Verbundpartnern als auch anderen Forschungs- und Anwendergruppen
- Ausbau von Forschungsthemen: revisionssichere und verfassungskonforme Datenhaltung/-verarbeitung, Integration DBMS & Retrieval-Systeme
- Integration in Lehre zu Themen Beweissichere Datenhaltung und -verarbeitung,

---

Nachvollziehbarkeit von Verarbeitungsschritten, Datensicherheit und Datenschutz im Bachelor und Master

- Wissenstransfer in andere Projekte wie CoMo (Competence in Mobility) – Teilprojekt „sichere Datenhaltung“ und Teilprojekt „Virtual Engineering“, ViERforES
- Wissenschaftliche, technische Begleitung der Anwendungs- und Präventivscenarien des LKA-ST, BKA, DNRI und TELEROB sowie interessierter Anwender
- Technische Begleitung weiterer Anwendungen im Bereich der biometrischen Datenhaltung, z. B. biometrischer Reisepass

## ***II.5 Bekanntgewordene Ergebnisse dritter Stellen mit Bezug zu den Forschungsthemen des Projekts***

Wie bereits in den Zwischenberichten beschrieben, sind im Verlaufe des Vorhabens keine Ergebnisse bekannt geworden, die direkt die Kernzeile des Teilprojektes betreffen. Allerdings sind bei unserer strukturierte Recherche nach Nr. 2.1. BNBest-BMBF 98 in gängigen Portalen wie der ACM Digital Library, IEEE Explore und SpringerLink und bei bekannten Konferenzen wie ACM SIGMOD und VLDB zusätzliche neu-publizierte und relevante Grundlagen- oder Forschungsergebnisse aus verwandten Anwendungsbereichen bekannt worden. Diese Erkenntnisse, geordnet nach Projektjahr sind im einzelnen:

### *Jahr 1 - 2010*

Im ersten Jahr wurden keine Publikationen gefunden, die mit den im Teilprojekt „Datenvorverarbeitung und Datenhaltung“ anvisierten Forschungszielen unmittelbar in Konflikt stehen bzw. eine Änderung der Forschungsziele erforderlich machen. Vielmehr zeigen die Veröffentlichungen in verwandten Themengebieten die Notwendigkeit und das Potential weiterer Forschung in diesen Bereichen. Exemplarisch seien hier Veröffentlichungen zum Thema Datenherkunft (Data Provenance) vorgestellt, welche direkt mit AP8 Beweissichere Datenhaltung zusammenhängen.

Durch die zunehmende Vernetzung (Internet, cloud computing) und die Nutzung, Weiterverwendung und Aggregation von Daten unterschiedlicher Herkunft und Qualität ergeben sich vor allem Probleme in Bezug auf die Verlässlichkeit der Daten, ähnlich wie dies auch für die Verwendung unserer Fingerabdrucksdaten der Fall ist. Für wissenschaftlichen Datenbanken wird beispielsweise in [R2] vorgeschlagen einen Verlässlichkeitsaspekt in die Daten zu integrieren und eine entsprechende Architektur zur Auswertung dieser Zusatzinformationen vorgestellt. Dieses Konzept wird in [R1] auf die von uns verwendete Technologie (relationale Datenbanken) ausgeweitet. Obwohl dieses Arbeiten ein ähnliches Problem behandeln, gehen die vom Projekt benötigten Anforderungen weit über den Leistungsumfang dieser Lösungen hinaus. So müssen wir die Integrität, Vertraulichkeit und Authentizität jedes einzelnen Datenobjekts für den gesamten Speicherungsprozess inkl. etwaiger Vorverarbeitungsschritte sicherstellen (Datenstammbaum bzw. engl. chain-of-custody). Weiterhin ist uns die Herkunft und Fehlerraten der einzelnen Daten bekannt, sodass wir nicht mit den gängigen Problem der Unsicherheit in den Daten konfrontiert sind, wie dies in [R3] der Fall ist.

Ähnlich verhält es sich bei der datenschutzrechtskonformen Gestaltung des Gesamtsystems (AP10) als auch der Referenzdatenbank (AP3). Hierfür wurde unlängst die Idee bzw. das Konzept der Hippokratischen Datenbanken vorgestellt [R4]. Jedoch wird in der Literatur nicht beschrieben, welche technischen und organisatorischen Maßnahmen zur Implementierung dieses Konzeptes (bzw. eines vergleichbaren Vorgehens) getroffen werden müssen. Zusätzlich müssen in unserem konkreten Fall die speziellen Gegebenheiten des deutschen Rechtes berücksichtigt werden, die sich beispielsweise vom US-Rechtssystem unterscheiden können.

---

### *Jahr 2 - 2011*

Im zweiten Jahr bestätigen sich die Erkenntnisse des letzten Jahres, die die Notwendigkeit und das Potential vor allem im (Data) Provenance Bereich (AP 8) zeigen.

Dieses Erkenntnisse werden zusätzlich durch eigene Publikationen und Einladungen zu Vorträgen in diesem Fachgebiet unterstützt. Im Bereich Provenance sind vor allem grundlegende Publikationen in hochrangigen Konferenzen und Journals, auf deren Ergebnissen wir aufbauen können, zu nennen. Hervorzuheben ist hier vor allem die Erweiterung des Semiringmodells [1] zur Erfassung und Propagierung von Provenance-Informationen basierend auf der relationalen Algebra auf Aggregationen [2] (inkl. Gruppierung). Ein weiterer wichtiger Beitrag ist die Verbindung der Konzepte hinter dem Semiringmodell (und verwandten formalen Ansätzen wie Lineage [3] und Why-Provenance [4]) zur Kausalität [5,6] basierend auf Halpern Modellen [7] bezüglich Ursache und Wirkung.

### *Jahr 3 & Verlängerung - 2012 bis 31.5.2013*

Während in den letzten beiden Jahren vor allem im Bereich Provenance grundlegende Publikationen entstanden, welche uns (a) als Ausgangspunkt dienten und (b) die Aktualität des Themengebietes verdeutlichen, sind im Jahr 2012 konkurrierende Ansätze zu [A] auf dem Workshop TaPP publiziert worden (siehe hierzu auch entsprechenden Konferenzbericht). Hervorzuheben sind vor allem [C1] und [C2]. Beide Ansätze beschäftigen sich mit der Integration des Provenance Concerns in bestehende Systemlandschaften. Ansatz [C1] ist direkt auf Betriebssystemebene verankert und kann daher nur schwer umgangen werden. Allerdings ist es diesem Ansatz, aufgrund der direkten Integration in das BS, nicht möglich detaillierte Provenance Informationen zu sammeln. Dies wird jedoch im DD-Kontext benötigt. Der Ansatz wird daher als sinnvolle Ergänzung betrachtet.

Anders verhält es sich im Ansatz [C2]. Dieser ist ein direkter Konkurrent. Hier werden modifizierte C-Compiler verwendet um Provenance-Daten mit ähnlichem Detailgrad ab zugreifen. Vorteil des Ansatzes ist die vollkommen transparente und einfache Anwendung der modifizierten Compiler. Im Gegensatz zu unserem Ansatz hat dieser jedoch zwei entscheidende Nachteile: (1) All-or-nothing property: Der Ansatz annotiert entweder ein gesamtes Programm mit dem Provenance Concern oder gar nichts. Dadurch werden große Mengen an nicht relevanten Provenance-Daten erzeugt (die in sich Datenschutzprobleme beherbergen können) aus denen die relevanten Daten anschließend mühevoll extrahiert werden müssen. Potentielle Erweiterungen erreichen aber aufgrund konzeptioneller Unterschiede nie den Maßschneiderungsgrad wie in unserem Ansatz. (2) Zweitens beschränkt sich der Ansatz, im Gegensatz zu unserem [1,11], auf die Verwendung von primitiven Variablen-Typen, schließt also alle objekt-orientierten Ansätze aus, ignoriert call-by-reference Konstrukte und Strukturen.

Interessant ist weiterhin die erste Formalisierung der Provenanzdatensammlung für Programmcode in [C3]. Dieser könnte als Nachweis der Korrektheit unseres eignen Ansatzes dienen.

## ***II.6 Erfolgte und geplante Veröffentlichungen***

Nachfolgend listen wir die in unserem Teilprojekt entstandenen Veröffentlichungen geordnet nach der jeweiligen Veröffentlichungsart auf. Dies umfasst (a) Workshop und Konferenzpublikationen, (b) Journalbeiträge, (c) Abschlussarbeiten und Sonstige Veröffentlichungen.

### ***Peer-reviewte Workshop- und Konferenzbeiträge***

Im folgenden Listen wir von uns veröffentlichten Beiträge in peer-reviewten Workshops und Konferenzen. Um dem Leser einen ersten Eindruck für die Wertigkeit der einzelnen Publikationen ist

für alle der Typ der Publikation vermerkt. Darüber hinaus geben wir, sofern es er ein Auszeichnungsmerkmal der jeweiligen Konferenz ist, die Akzeptanzrate mit an.

\*) Der Typ der Konferenz nicht unbedingt gleichbedeutend mit dem Typ der Dienstreise, beispielsweise wenn International Konferenzen in Deutschland stattfinden.

Abgeschlossen		Typ*
[GPCE10] AP 3 & AP 5	Code Clones in Feature-Oriented Software Product Lines. GPCE, 2010.	Internationale Konferenz 31% (18/59)
[BTW11] AP 3	Generierung maßgeschneiderter Relationenschemata in Softwareproduktlinien mittels Superimposition. BTW, 2011.	Nationale Konferenz
[BioID11] AP 8	Database-Centric Chain-of-Custody in Biometric Forensic Systems. BioID, 2011.	Europäischer Workshop
[BNCOD11] AP 8	Reliable Provenance in Multimedia Data using Fragile Watermarks. BNCOD, 2011.	Nationale Konferenz
[SPIE12] AP 3	A First Proposition for a General Modelling of Trace Evidences. SPIE Photonics, 2012.	Europäischer Workshop
[CAISE12] AP3 Kern- publikation	Building Information System Variants with Tailored Database Schemas Using Features. CAiSE, 2012.	Internationale Konferenz 14% (42/297)
[BALTIC12] AP 5	Workload-based Heuristics for Evaluation of Physical Database Architectures. Baltic DB&IS, 2012.	Europäische Konferenz
[TAPP12] AP 8	Toward provenance capturing as cross-cutting concern. TaPP, 2012.	Internationaler Workshop
[GVD12] AP 5	Challenges in finding an appropriate multi-dimensional index structure with respect to specific use cases. GvD 2012.	Nationaler Workshop
[BTW_W13a] AP 8	The Monetary Value of Information. BTW-Workshops 2013.	Nationaler Workshop
[ADBIS_W12] AP 5	Towards Optimization of Hybrid CPU/GPU Query Plans in Database Systems. ADBIS Workshops 2012.	Europäischer Workshop
[ADBIS12] AP 5	Automatic Selection of Processing Units for Coprocessing in Databases. In ADBIS 2012.	Europäische Konferenz 26% (32/122)
[BTW_W13b] AP 5	Dynamic Relational Data Management for Technical Applications. BTW-Workshops 2013.	Nationaler Workshop
[BTW_W13c] AP 5	Extending an Index-Benchmarking Framework with Non-Invasive Visualization Capability. In BTW-Workshops 2013.	Nationaler Workshop
[BTW13] AP 10	Privacy-Aware Multidimensional Indexing. In BTW. 2013.	Nationale Konferenz 13%
[BTW_W13d] AP 10	Secure Deletion: Towards Tailor-Made Privacy in Database Systems. BTW-Workshops. 2013.	Nationaler Workshop
[BTW_W13e] AP 8	Database and Data Management Requirements for Equalization of Contactless	Nationaler

[BTW_W13f] AP 10	Acquired Traces for Forensic Purposes. In BTW-Workshops. 2013.	Workshop
[BTW_W13g] AP 5	Forensics on GPU Coprocessing in Databases. In BTW-Workshops. 2013.	Nationaler Workshop
[BTW_W13h] AP 5	Optimierung der Exact-Match Anfrage eines Lokal Sensitiven Hashverfahrens. In BTW-Workshops. 2013.	Nationaler Workshop
[BTW_W13h] AP 5	Parallel Execution of kNN-Queries on in-memory K-D Trees. In BTW-Workshops. 2013.	Nationaler Workshop

### **Journalbeiträge**

In der Folge listen wir die die von uns publizierten bzw. in Arbeit befindlichen Journal-Beiträge aus peer-reviewten Journals auf. Zur Einordnung der Qualität des entsprechenden Beitrags sind Kennzahlen wie der Impact Factor oder h5-index, je nach Verfügbarkeit, mit aufgelistet.

[C&C12] AP 5	A Framework for Cost based Optimization of Hybrid CPU/GPU Query Plans. DB Control and Cybernetics, 41(4) 2012.	0,32 Impact Factor
[InfSys13] AP 5	Efficient co-processor utilization in database query processing. Information Systems 38(8) 2013.	1,77 Impact Factor
[VLDB14] AP5 Kern- publikation (M4)	QuEval: Beyond high-dimensional indexing à la carte. PVLDB. 2014.	akzeptiert 01/08/13 h5-index: 66
<b>geplant</b>		
[DBSpek13] AP 8	Arbeitstitel: Database and Data Management Requirements for Equalization of Contactless Acquired Traces for Forensic Purposes - Provenance and Performance. Datenbank-Spektrum.	geplant für Ende 2013
[DKE14] AP8 Kern- publikation	Arbeitstitel: A Hierarchical Framework for Provenance Based on Fragmentation and Uncertainty. DKE Journal.	geplant für 2014

### **Abschlussarbeiten (Bachelor und Master)**

Abschlussarbeiten und die Zuordnung zu den jeweiligen Arbeitspaketen sind nachfolgend aufgeführt.

[Hec13]	Sarah Heckel. Lineares Programm zur Prototypenbestimmung für das permutationsbasierte Indexverfahren. Bachelorarbeit, April 2013.	AP 5
[Sch12]	Christopher Schulz. Ansätze zur Erzeugung variabler Datenbankschemata in Softwareproduktlinien. Masterarbeit, Dezember 2012.	AP 3
[Bra12]	Waldemar Braun. Anfrageinterface für zeilenorientierte und spaltenorientierte Datenbanksysteme. Masterarbeit, Oktober 2012.	AP 5
[Lin12]	Alexander Grebhahn. Analyse des Rekonstruktionspotentials von	AP 10

	multidimensionalen Indexstrukturen zum sicheren Löschen. Masterarbeit, Oktober 2012.	
[Bro12]	David Broneske. Visuelle Analyse der Raumaufteilung und Bucketauslastung von permutationsbasierten Indexverfahren. Bachelorarbeit, April 2012.	AP 5
[Gre12]	Ina Lindauer. Forensisch sicheres Löschen in relationalen Datenbankmanagementsystemen. Masterarbeit, Germany, März 2012.	AP10
[Her11]	Tim Hering. Untersuchung der Leistungsunterschiede zwischen Variationen des Vector Approximation File bei k-Nearest-Neighbor-Anfragen im hochdimensionalen Raum. Bachelorarbeit, Dezember 2011	AP 5
in Bearbeitung		
[Her13]	Tim Hering. Parallelizing multidimensional indexes for main memory databases. Masterarbeit.	August 2013
[Tob13]	Martin Tobies. Konzeption, Modellierung und prototypische Umsetzung eines Benchmarks für mehrdimensionale Indexstrukturen. Masterarbeit.	August 2013
[Weh13]	Adrian Wehrmann. Exemplarische Erweiterung mehrdimensionaler Indexstrukturen um Epsilon-Bereichsanfragen. Bachelorarbeit.	Oktober 2013

### **Sonstiges**

Folgende sonstige Veröffentlichungen wurden im Rahmen des Projektes vorgenommen.

[ASB12] AP 5 - M4	Techniken zur forensischen Datenhaltung 1 - Ausgewählte studentische Beiträge (AsB). Universität Magdeburg, 2012.	<b>Herausgeber- schaft</b>
[ASB12a] AP 5	Vergleich von Indexverfahren für hochdimensionale Datenräume. ASB 1(1) 2012.	
[ASB12b] AP 5	Implementierung und Evaluierung hochdimensionaler Indexstrukturen. ASB 1(1) 2012.	
[ASB12c] AP 8	Secure Database Infrastructures. ASB 1(1) 2012.	
[ASB12d] AP 5	Implementierung und Evaluierung von Distanzfunktionen im Digi-Dak-Framework. ASB 1(1) 2012.	
[ASB12e] AP 5	Einfluss von multivariat schief normalverteilten Daten auf multidimensionale Indexstrukturen. ASB 1(1) 2012.	
[FIN12] AP 8 - M8.3	A Hierarchical Framework for Provenance Based on Fragmentation and Uncertainty 2012.	FIN Techreport review durch Fakultät
geplant		
[ASB13]	Techniken zur forensischen Datenhaltung 2 - Ausgewählte studentische Beiträge (AsB). Universität Magdeburg, 2013.	<b>Herausgeber- schaft</b>



---

## Literaturverzeichnis und sonstige Referenzen

### *Literatur des aktuellen Forschungsstandes*

- [Bla98] Blatter, C.: Wavelets – Eine Einführung. Vieweg, 1998.
- [SDS95a] Stollnitz, E.J., DeRose, T.D., Salesin, D.H.: Wavelets for Computer Graphics: A Primer, Part 1. In IEEE Computer Graphics and Applications, May 1995.
- [SDS95b] Stollnitz, E.J., DeRose, T.D., Salesin, D.H.: Wavelets for Computer Graphics: A Primer, Part 2. In IEEE Computer Graphics and Applications, July 1995.
- [Dum93] Dumais, S.T.: Latent semantic indexing (lsi) and trec-2. In Text Retrieval Conference (TREC), 1993, p. 105-116.
- [GKL01] Guan, L., Kung, S.-Y., Larsen, J., editors: Multimedia Image and Video Processing. CRC Press, 2001.
- [Jac03] Jackson, J.E., editor: A User's Guide to Principal Components. Wiley Interscience, 2003.
- [EP+03] Evgeniou, T., Pontil, M., Papageorgiou, C., Poggio, T.: Image Representations and Feature Selection for Multimedia Database Search. In IEEE Transactions on Knowledge and Data Engineering (TKDE), Volume 15(4), 2003, p. 911-920.
- [DK+09] Dias, N.S.M., Kamrunnahar, M., Mendes, P.M., Schiff, S. J., Correia, J.H.: Variable Subset Selection for Brain-Computer Interface - PCA-based Dimensionality Reduction and Feature Selection. In Proc. of the Int. Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS), 2009, p. 35-40.
- [BBK01] Böhm, C., Berchthold, S., Keim, D.: Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases. In ACM Computing Surveys, Volume 33(3), 2001, p. 322-373.
- [Cas02] Castelli, V.: Image Databases – Search and Retrieval of Digital Imagery. Chapter in Multidimensional Indexing Structures for Content-Based Retrieval, 2002, p. 373-433.
- [CB02] Castelli, V., Bergman, L.D., editors: Image Databases – Search and Retrieval of Digital Imagery. John Wiley & Sons, 2002.
- [FSZ03] Feng, D., Siu, W.C., Zhang, H.J., editors: Multimedia Information Retrieval and Management: Technological Fundamentals and Applications. Springer-Verlag, 2003
- [BK+07] Böhm C., Kunath P., Pryakhin A., Schubert M.: Effective and Efficient Indexing for Large Video Databases. In Proceedings GI-Fachtagung für Datenbanksysteme in Business, Technologie und Web (BTW), 2007.
- [BK+07] Böhm C., Kunath P., Pryakhin A., Schubert M.: Effective and Efficient Indexing for Large Video Databases. In Proceedings GI-Fachtagung für Datenbanksysteme in Business, Technologie und Web (BTW), 2007.
- [BB+97] Berchthold, S., Böhm, C., Keim, D., Kriegel, H.-P.: A Cost Model for Nearest Neighbor in High-Dimensional Data Space. In Proc. of the ACM Symposium on Principles of Database Systems (PODS), 1997, p. 78-86.
- [GO04] Goodman, J. E., O'Rourke, J.: Handbook of Discrete and Computational Geometry, 2nd Edition, CRC Press, 2004.

- [KK+09] Kriegel, H.-P., Kröger, P., Renz, M., Züfle, A., Katzdobler A.: Incremental Reverse Nearest Neighbor Ranking. In Proceedings of the Int. Conference on Data Engineering (ICDE), 2009.
- [BG+99] Beyer, K., Goldstein, J., Ramakrishnan, R., Shaft, U.: When is Nearest Neighbor meaningful?. In Proc. of the Int. Conference on Database Theory (ICDT), 1999, p. 217-235.
- [Sch01] Schmitt, I.: Nearest Neighbor Search in High Dimensional Space by Using Convex Hulls. Preprint 6, University of Magdeburg, 2001.
- [WSB98] Weber, R., Schenk, H.-J., Blott, S.: A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces. In Proc. of the Int. Conference on Very Large Data Bases (VLDB), 1998, p. 194-205.
- [KM03] Klettke, M., Meyer, H.: XML & Datenbanken – Konzepte, Sprachen und Systeme. Dpunkt.verlag, 2003.
- [All83] Allen, J.F.: Maintaining knowledge about temporal intervals. In Communications of ACM, Volume 26(11), 1983, p. 832-843.
- [BR99] R. Baeza-Yates, B. Ribeiro-Neto: Modern Information Retrieval. ACM Press, Addison-Wesley, New York 1999.
- [Gep02] Geppert, A.: Objektrelationale und objektorientierte Datenbankkonzepte und -systeme. Dpunkt.verlag, 2002.
- [HS00] Heuer, A., Saake, G.: Datenbanken – Konzepte und Sprachen, 2.Auflage. MITP-Verlag, 2000.
- [ME01] Melton, J., Eisenberg, A.: SQL Multimedia and Application Packages (SQL/MM). In ACM SIGMOD Record, Volume 30(4), 2001, p. 97-102.
- [Sto02] Stolze, K.: Still Image Extension in Database Systems – A Product Overview. Datenbank-Spektrum, 2002, p. 40-47.
- [Dun03] Dunckley, L: Multimedia Databases: An Object-Relational Approach. Addison Wesley, 2003.
- [Sch04] Schulz, N.: Formulierung von Nutzerpräferenzen in Multimedia-Retrieval-Systemen. PhD thesis, University of Magdeburg, 2004.
- [Pas99] Paskamp, M.: Vergleichende Analyse von Anfragesprachen in Multimedia-Datenbanken. Master Thesis, University of Magdeburg, 1999.
- [Zlo75] Zloof, M.M.: Query By Example. In Proc. of the AFIPS National Computer Conference, Volume 44, 1975, p. 431-438.
- [SLG97] Scharnofske, A., Lipeck, U.W., Gertz, M.: SubQuery-By-Example: Eine orthogonale Erweiterung von QBE. In Proceedings GI-Fachtagung für Datenbanksysteme in Business, Technologie und Web (BTW), 1997.
- [Cha08] Chapman, A.P.: Incorporating Provenance in Database Systems, PhD thesis, University of Michigan, 2008.
- [GD07] Glavic, B., Dittrich, K.: Data Provenance: A Categorization of Existing Approaches, In Proc. 12. GI-Fachtagung für Datenbanksysteme in Business, Technologie und Web (BTW), 2007.
- [HR+09] Hasan, R., Sion, R., Winslett, M.: The Case of the Fake Picasso: Preventing History Forgery with Secure Provenance. In Proc. of the 7th USENIX Conference on File and Storage Technologies (FAST), 2009.
- [IW09] Ikeda, R., Widom, J.: Data Lineage: A Survey. Technical Report, Stanford University, 2009.
- [Tan04] Tan, W.-C.: Research Problems in Data Provenance. In IEEE Data Engineering Bulletin, 2004.
- [ZC+09] Zhang, J., Chapman, A., LeFevre, K.: Do You Know Where Your Data's Been? - Tamper-Evident Database Provenance. Technical Report Nr. CSE-TR-548-08, University of Michigan, 2009.

---

## Sonstige Referenzen

- [BSR04] Scaling Step-Wise Refinement. IEEE Transactions on Software Engineering 30(6), 2004.
- [CC07] Very Fast Watermarking by Reversible Contrast Mapping Signal Processing Letters (14), 2007.
- [CN00] Software Product Lines: Practices and Patterns. Addison-Wesley, 2001.
- [CSTS05] Lossless generalized-LSB data embedding, IEEE Transactions on Image Processing (14), 2005.
- [FTC06] Reversible watermarking: Current status and key issues, International Journal of Network Security(2), 2006.
- [KVL11] Automated Forensic Fingerprint Analysis: A Novel Generic Model and Container Format. Workshop on Biometrics and ID Management, 2011.
- [MKDV11] Reversible Watermarking with Digital Signature Chaining for Privacy Protection of Optical Contactless Captured Biometric Fingerprints - A Capacity Study For Forensic Approaches Int'l Conf. on Digital Signal Processing, 2011.
- [Tia03] Reversible data embedding using a difference expansion, Circuits and Systems for Video Technology (13), 2003.

## *Literatur - Bekanntgewordene Ergebnisse*

- [R1] Data aspects in relational databases. CIKM. 2010
- [R2] Weaving temporal and reliability aspects into a schema tapestry. Data Knowl. Eng. 63(3). 2007
- [R3] Databases with uncertainty and lineage. VLDB Journal 17(2). 2008
- [R4] Hippocratic Databases. VLDB. 2002
- [1] Provenance Semirings. In PODS, 2007.
- [2] Provenance for Aggregate Queries. In PODS, 2011
- [3] ULDBs: Databases with Uncertainty and Lineage. In VLDB, 2006.
- [4] Why and Where: A Characterization of Data Provenance. In ICDT, 2001.
- [5] Causality in Databases. In IEEE Data Eng. Bulletin, 2010.
- [6] The complexity of causality and responsibility for query answers and non-answers. In VLDB, 2010.
- [7] Causes and Explanations: A structural-model approach. part-i: Causes. Brit. Journ. Phil. Sci, 2005.
- [C1] A General-Purpose Provenance Library. TaPP 2012.
- [C2] Towards Automated Collection of Application-Level Data Provenance., SRI International. TaPP 2012.
- [C3] Hierarchical Models of Provenance. TaPP 2012.

## Vorträge/Seminare

- [I] Eingeladener Vortrag beim CREST Workshop: Thema: Forms of Provenance. 2011.
- [II] Einladung zum Dagstuhl Seminar on Principals of Provenance. Thema: Toward Provenance as Cross-cutting Concern. 2012.

---

**Forschungsdemonstratoren:**

- [A] Provenance Framework: Integration von Provenance als querschnittender Belang in existierende Software-Lösungen (DBMS).

**Lehre:**

- [B] Ausbau des Evaluations-Frameworks (QueVal): Neue Indexstrukturen, Berücksichtigung weiterer Datenverteilung, Speichertypen & Distanzmetriken.
- [C] Teilnahme am SIGMOD Programming Contest: A Multidimensional Indexing System
- [D] Projekt-Vorlesung: Digi-Dak Database Project.