

Bachelorarbeit

Abdul Waheed Wasiee

Konzepte zur externen Bereitstellung von Informationen aus einem Archivsystem – eine Untersuchung am Beispiel einer Berufsgenossenschaft

Abdul Waheed Wasiee

Konzepte zur externen Bereitstellung von Informationen aus einem Archivsystem – eine Untersuchung am Beispiel einer Berufsgenossenschaft

Bachelorarbeit eingereicht im Rahmen der Bachelorprüfung
im Studiengang Angewandte Informatik
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. sc. pol. Wolfgang Gerken
Zweitgutachter : Prof. Dr. rer. nat. Wolfgang Fohl

Abgegeben am 30 März 2010

Abdul Waheed Wasiee

Thema der Bachelorarbeit

Konzepte zur externen Bereitstellung von Informationen aus einem Archivsystem – eine Untersuchung am Beispiel einer Berufsgenossenschaft

Stichworte

Archivserver, Metadaten, Massenspeicher, Datenbank, Metadaten Index, Datenträger, Informationen, Archivierung, Recherche

Kurzzusammenfassung

Die modernen Archivsysteme ermöglichen es, dass jede Art von Information zentral oder in verteilten Informationspools gespeichert und verfügbar ist. Die Informationswiedergewinnung erfolgt in der Regel über die Metadaten, die bei der Archivierung festgelegt werden.

Bei der Berufsgenossenschaft für Gesundheitsdienst und Wohlfahrtspflege (BGW) werden die Informationen zwar ebenfalls durch Festlegung von Metadaten zurückgewonnen, jedoch nicht über eine direkte Archivalsuche, sondern über ein eingebundenes versorgendes System.

In dieser Arbeit werden bei der BGW Möglichkeiten untersucht, Informationen durch eine direkte Archivalsuche bereitzustellen. Hierzu werden die Archivierung, Recherche und Information-Retrieval-Vorgänge betrachtet und die Auswahl und Verfügbarkeit der Metadaten aufgenommen. Außerdem werden Möglichkeiten der Volltextsuche untersucht.

Diese Arbeit zeigt drei mögliche Wege auf, wie die Informationswiedergewinnung durch eine direkte Archivalsuche ermöglicht wird. In diesem Fall werden, mithilfe von Volltextindizierung oder Schlüsselnerfassung, Informationen durch eine direkte Archivalsuche gewonnen.

Abdul Waheed Wasiee

Title of the paper

Concept for the external supply of information from an archive system -
An investigation using a professional association as an example.

Keywords

Archive server, Metadata, Mass storage device, Data base, Metadata index, Data carrier, Information, Archiving, Research.

Abstract

Modern archive systems enable all kinds of information to be centralised or available and stored in distributed information pools. The information recovery occurs, as a rule, via the metadata, which are fixed during archiving.

Using the example of the Professional Association for Health and Public Welfare (BGW), the information is similarly retrieved by fixing the metadata, but not, however, via a direct archive search, but rather via an integrated supply system.

This work concerns itself with the possibilities that the BGW have of providing information using a direct archive search. In addition, the archiving, search and information retrieval processes are examined, as well as the choice and availability of the metadata. Moreover, full-text search possibilities are also examined.

This work indicates the three possible ways that information retrieval is enabled using a direct archive search. In this case, information is retrieved with the help of full text indexing or code recompilation, using a direct archive search.

Danksagung

An dieser Stelle möchte ich die Gelegenheit nutzen, einigen Leuten zu danken, die mich während dieser Arbeit unterstützt haben.

Ein großes Dankeschön geht an:

- Prof. Dr. Wolfgang Gerken, der mir mit seinen vielen Anregungen und Erläuterungen zur Seite gestanden hat.
- Prof. Dr. Wolfgang Fohl, der sich bereit erklärt hat, das Zweitgutachten zu übernehmen.
- Herr Ulrich Eilers von der BGW, von dem ich ausgezeichnet betreut wurde.

Weiterhin gilt mein Dank meiner Frau Anja und meiner Tochter Marie, die mich während meiner ganzen Studienzeit unterstützt haben.

Diese Arbeit widme ich meinen Eltern, die mich immer unterstützt und an mich geglaubt haben.

Inhaltsverzeichnis

ABBILDUNGSVERZEICHNIS	VI
TABELLENVERZEICHNIS	VII
1 EINLEITUNG	1
1.1 MOTIVATION	1
1.2 ZIELSETZUNG UND VORGEHEN	1
1.3 DER BETRIEB BGW	2
2 GRUNDLAGEN	4
2.1 KURZBESCHREIBUNG ELEKTRONISCHER ARCHIVSYSTEME	4
2.2 DIE MERKMALE EINES ARCHIVSYSTEMS	5
3 STAND DER TECHNIK BEI DER BGW	6
3.1 SER-ARCHIVSYSTEM ARCHITEKTUR	6
3.1.1 <i>Systemüberblick</i>	6
3.1.1.1 Die ITA-Server-Verwaltungssysteme	7
3.1.2 <i>Dokumentenformate</i>	9
3.1.3 <i>Der Dokumenten-Header</i>	10
3.1.4 <i>Strukturierte Ablage</i>	10
3.1.5 <i>Ablauf Archivierungsaufträge</i>	12
3.1.6 <i>Strukturierte Recherche</i>	13
3.1.6.1 Ablauf Rechercheaufträge	14
3.1.6.2 Ablauf Retrieval-Aufträge	15
NON-CODED INFORMATION (NCI) DOKUMENTE	9
CODED INFORMATION (CI) DOKUMENTE	9
BINÄRDATEIEN	10
ERSTELLER UND EIGENTÜMER DER DOKUMENTE	11
INHALTLICHE KLASSIFIZIERUNG DER DOKUMENTE	11
ABLAGEDATUM DER DOKUMENTE	11
4 ARCHIVIERUNGSVORGANG ARCHITEKTUR BEI DER BGW	16
4.1 SYSTEMÜBERBLICK	17
4.1.1 <i>Einzugsscanner</i>	18
4.1.2 <i>Elektronischer Dokumenten Manager (EDM) Server</i>	18
4.1.3 <i>Intuitiv-Client</i>	18
4.1.4 <i>ITA-Archivserver</i>	18
4.1.5 <i>BG-Standard</i>	19
4.1.6 <i>Computer Output on Laser Disk (COLD) Client</i>	19
4.2 POSTEINGANGSARCHIVIERUNG	19

4.3	POSTAUSGANGSARCHIVIERUNG	23
4.4	DOKUMENTENRECHERCHE	25
5	ANALYSE DER ANFORDERUNGEN	30
5.1	BG-STANDARD-AUSFALL	30
5.2	ARCHIVSUCHE ZUR ERSTELLUNG VON HEUTE NICHT BEKANNTEN AUSWERTUNGEN ÜBER ARCHIVBESTÄNDE	32
6	REALISIERUNGSMÖGLICHKEITEN	33
6.1	VOLLTEXTSUCHE	33
6.2	SER-VOLLTEXTSUCHE	33
6.2.1	<i>Verity-Volltextsuche</i>	35
6.2.1.1	Ablauf der Volltextindizierung mit Verity	35
6.2.2	<i>ASSA-Volltextsuche (assoziative Suche)</i>	36
6.2.2.1	Ablauf der Volltextindizierung mit ASSA	36
6.3	REALISIERUNGSWEGE	37
6.3.1	<i>Die Eigenentwicklung eines Recherche-Tools</i>	37
6.3.2	<i>Der Rendition-Server</i>	38
6.3.3	<i>ITA-COLD-Schlüsselnacherfassung</i>	39
6.4	BESTÄTIGUNG DURCH DIE HAMBURGER GESCHÄFTSSTELLE DER FIRMA SER	39
7	BEWERTUNG DER REALISIERUNGSWEGE	40
7.1	BEWERTUNG	40
7.1.1	<i>Die Eigenentwicklung eines Recherche-Tools</i>	40
7.1.2	<i>Der Rendition-Server</i>	41
7.1.3	<i>ITA-COLD-Schlüsselnacherfassung</i>	42
7.1.4	<i>Bewertungstabelle</i>	43
7.2	FAZIT	43
	ABKÜRZUNGSVERZEICHNIS	45
	LITERATURVERZEICHNIS	47

Abbildungsverzeichnis

Abbildung 3.1: ITA-Archivsystem	7
Abbildung 3.2: Ablauf Archivierungsauftrag	12
Abbildung 3.3: Ablauf Rechercheauftrag	14
Abbildung 3.4: Ablauf Retrieval-Auftrag	15
Abbildung 4.1: BGW Archivserver-Konfiguration	17
Abbildung 4.2: Intuitiv-Client	20
Abbildung 4.3: Posteingangsarchivierung	22
Abbildung 4.4: Postausgangsarchivierung	25
Abbildung 4.5: PEPA-Übersichtstabelle	26
Abbildung 4.6: Recherche über PEPA-Tabelle	27
Abbildung 4.7: Direkte Archivrecherche im Bereich UB	28
Abbildung 4.8: Direkter Recherchevorgang im Bereich UB	29
Abbildung 5.1: Direkte Archivsuche im Bereich Finanz	30
Abbildung 5.2: Direkte Archivrecherche im Bereich UB	31
Abbildung 6.1: Volltextsuche mit dem Evita-Client	34
Abbildung 7.1: Bewertungstabelle	43

Tabellenverzeichnis

Tabelle 4.1: Aufbau der Posteingangsbelegnummer	21
Tabelle 4.2: Aufbau der Postausgangsbelegnummer	23

1 Einleitung

1.1 Motivation

In Unternehmen und Verwaltungen werden heute von Computersystemen riesige Datenmengen erzeugt. Diese sogenannten Massendaten müssen aufgrund gesetzlicher Vorschriften und betrieblicher Anforderungen viele Jahre ordnungsgemäß und revisions sicher aufbewahrt werden.¹

Der Aufbau eines „Information Repository“ (Aufbewahrungsort) wird durch die elektronische Archivierung ermöglicht, indem sich jede Art von unstrukturierten bzw. strukturierten Informationen zentral oder in verteilten Informationspools speichern lässt.

Ein weiterer wesentlicher Vorteil der elektronischen Archivierung ist, einen einheitlichen Zugriff auf alle Informationen des Unternehmens zu haben, um sie jederzeit in den Geschäftsprozessen nutzen zu können.

Die Informationswiedergewinnung ist entweder durch direkten Zugriff auf das Archivsystem oder durch indirekten Zugriff über eingebundene Systeme möglich.

Die Berufsgenossenschaft für Gesundheitsdienst und Wohlfahrtspflege (BGW) archiviert ihre Dokumente in einem Archivsystem der Firma SER Solutions Deutschland GmbH. Die Archivrecherche und Informationswiedergewinnung erfolgen über ein eingebundenes System.

Zusätzlich möchte die BGW durch die direkte Archivsuche flexibel und unabhängig von dem eingebundenen System Informationen gewinnen.

1.2 Zielsetzung und Vorgehen

Das Ziel dieser Arbeit ist es, Möglichkeiten zu untersuchen, Informationen aus dem Archivsystem bereitzustellen, welche außerhalb von Zugriffslogiken versorgender Sys-

¹ Vgl. Handelskammer Hamburg.

teme liegen. Hierfür soll in der Arbeit zunächst erklärt werden, wie im Allgemeinen moderne Archivsysteme arbeiten und welche gemeinsamen Merkmale diese haben.

Zunächst sollen die Aufgaben und Dienste des Archivservers der BGW näher betrachtet werden. Da für die Informationsbereitstellung aus dem Archivsystem die sogenannten Deskriptoren benötigt werden, werden nachfolgend die Vorteile und die Wichtigkeit von Deskriptoren in Archivierungs- und Recherchevorgängen sowie in den Vorgängen „Information Retrieval“ des Archivservers dargestellt. Anschließend werden die Posteingangs- und Postausgangsarchivierungsvorgänge sowie die Recherche und der Vorgang „Information Retrieval“ bei der BGW genauer beschrieben.

Anhand von zurzeit verfügbaren Hilfsmitteln zur Informationswiedergewinnung wird eine direkte Archivsuche analysiert und die Anforderungen der BGW an eine direkte Archivsuche im weiteren Verlauf beschrieben.

Für die effiziente Informationswiedergewinnung über die direkte Archivsuche sollen verschiedene Lösungswege vorgestellt werden. Desweiteren werden die Vorteile der verschiedenen Lösungswege beschrieben.

Diese dienen als Bewertungsgrundlage für eine spätere Entscheidung für einen der verschiedenen Lösungswege.

1.3 Der Betrieb BGW

Die Berufsgenossenschaft für Gesundheitsdienst und Wohlfahrtspflege (BGW) wurde 1929 in Berlin gegründet. Die BGW bietet gesetzliche Unfallversicherung an nichtstaatlichen Einrichtungen an und ist Teil des deutschen Sozialversicherungssystems.

Auch nach ersten erfolgten Fusionen unter den Berufsgenossenschaften aufgrund von gesetzlichen Bestimmungen ist die BGW eine der größten gewerblichen Berufsgenossenschaften in Deutschland. Die BGW hat ca. 565.000 Mitgliedsunternehmen mit insgesamt rund 6.615.000 Versicherten. Bundesweit ist die BGW an zwölf Standorten vertreten und beschäftigt 1.787 Mitarbeiter.

Zu den Versicherten gehören Unternehmen aus dem Gesundheitsdienst und der Wohlfahrtspflege. Genauer gesagt sind dies private Krankenhäuser, Ärzte, Apotheken, gemeinnützige Organisationen, Pflegedienste, Kosmetikstudios und Friseursalons. Die Unternehmen sind verpflichtet, ihre Arbeitnehmer während der Arbeitszeit und auf ihrem Weg zur Arbeit gegen Unfälle und Berufskrankheiten zu versichern. Berufs-

krankheiten sind z. B. Hauterkrankungen durch Allergien in der Friseurbranche. Bei Unfällen während der Arbeitszeit werden die Behandlungskosten von der BGW übernommen. Im Gegensatz zu anderen gesetzlichen Versicherungen werden die Beiträge nur vom Arbeitgeber bezahlt.²

² Vgl. BGW(2009).

2 Grundlagen

2.1 Kurzbeschreibung elektronischer Archivsysteme

Ein elektronisches Archiv besteht aus mehreren Komponenten und bildet die Kernkomponente der elektronischen Archivierung.

Die Dokumente werden, falls sie nicht elektronisch, sondern in Papierform erstellt wurden, über einen Scanner digitalisiert und vor der Archivierung indexiert. Beim Indexieren werden den Dokumenten Metainformationen, wie zum Beispiel Abteilung, Autor, Aktenzeichen, Sachgebiet oder Suchschlagwörter, hinzugefügt.

Danach werden die Dokumente in das Archivsystem archiviert. Ein Archivsystem besteht aus einem Archivserver, an dem ein Massenspeicher angeschlossen ist. Die Verwaltung von Massenspeichern wird in der Regel von einem Plattenwechselautomaten (Jukebox) übernommen.

Als Speicher werden einmal beschreibbare optische Medien wie DVD (Digital Versatile Disc), CD ROM (Compact Disc), oder UDO Discs (Ultra Density Optical) verwendet. Diese Medien nennt man auch WORM (write once read many). Ist die Performance sehr wichtig, werden sogenannte CAS (Content Adressed Storage) Systeme eingesetzt. CAS-Systeme sind Festplattensysteme mit gleichen Eigenschaften wie ein herkömmliches WORM-Medium. Diese Eigenschaften werden durch ihren Aufbau und spezielle Software erreicht.

Die Speichersysteme werden durch die Codierung bei der Speicherung und der speziellen Adressierung vor Überschreiben oder Ändern der Informationen geschützt. Die Metadaten des Archivguts werden dann an einen Datenbankserver weitergegeben und dort auf Festplatte gespeichert. Von diesem Server aus werden dann Suchanfragen beantwortet.³

Diese sehr vereinfachte Architektur und der Aufbau eines elektronischen Archivs kann je nach Archivlösung geändert werden. Die Architektur kann mit zusätzlichen Komponenten wie OCR (Optical Character Recognition) Server, Konvertierungsserver, Zeit-

³ Vgl. Schwalm/Ullrich(2008).

stempelinheiten, HSM (Hardware-Security-Module) oder Modulen zur automatischen Klassifizierung (Support Vector Machines) erweitert werden.

2.2 Die Merkmale eines Archivsystems

Für alle möglichen Anwendungsfälle und Anforderungen gibt es elektronische Archive. Aus diesem Grund sind sie zum Teil unterschiedlich aufgebaut und bieten unterschiedliche Funktionen. Obwohl die Systeme sehr unterschiedlich sein können, lässt sich im Allgemeinen aber sagen, dass elektronische Archive sich durch folgende Merkmale auszeichnen:

- einmal gespeicherte Informationen können nicht mehr verändert werden;
- zum Schutz vor unberechtigtem Zugriff findet eine Zugriffskontrolle/Autorisierung statt;
- die gespeicherten Informationen (Dokument, Zeichnung, Foto, etc.) können mit Metadaten versehen werden;
- über einen datenbankgestützten Index können die Informationen im Archiv recherchiert werden;
- das Archiv bietet eine standardisierte Schnittstelle, Application Programming Interface (API) an, über die andere Anwendungen auf das Archiv zugreifen können;
- das Archiv unterstützt ein verlustfreies Migrationsverfahren.⁴

Datenträger wie Compact Disc oder DVD müssen je nach Qualität der Datenträger alle drei bis zehn Jahre neu überspielt werden.

Die verschiedenen Archivierungssysteme haben natürlich auch einen unterschiedlichen Umfang, was Funktionen und Features betrifft, aber diesen Aufbau und diese Merkmale haben alle professionellen Lösungen in der Regel gemein.

⁴ Vgl. Schwalm/Ullrich(2008).

3 Stand der Technik bei der BGW

3.1 SER-Archivsystem Architektur

Die Firma SER Solutions Deutschland GmbH bietet für elektronische Archivierung neben dem Vertrieb von Hardware, Software aus eigener Produktion an. Im Kern des Archivsystems ermöglicht der Archivserver die Erfassung, Archivierung, Recherche und Weiterbearbeitung von Dokumenten. In diesem Zusammenhang stellt der Archivserver „ITA“ die folgenden Funktionen bei der BGW bereit:

- Dokumente archivieren
- Dokumente recherchieren
- Dokumente und Daten sichern
- Dokumente und Daten rekonstruieren
- Speichergeräte verwalten
- Medien und Laufwerke verwalten

3.1.1 Systemüberblick

Die folgende grafische Darstellung zeigt die wichtigsten Komponenten des SER-Archivsystems mit ITA-Server im Kern.

Nachfolgend werden die ITA-Server Verwaltungssysteme und die Komponenten des Archivsystems erläutert.

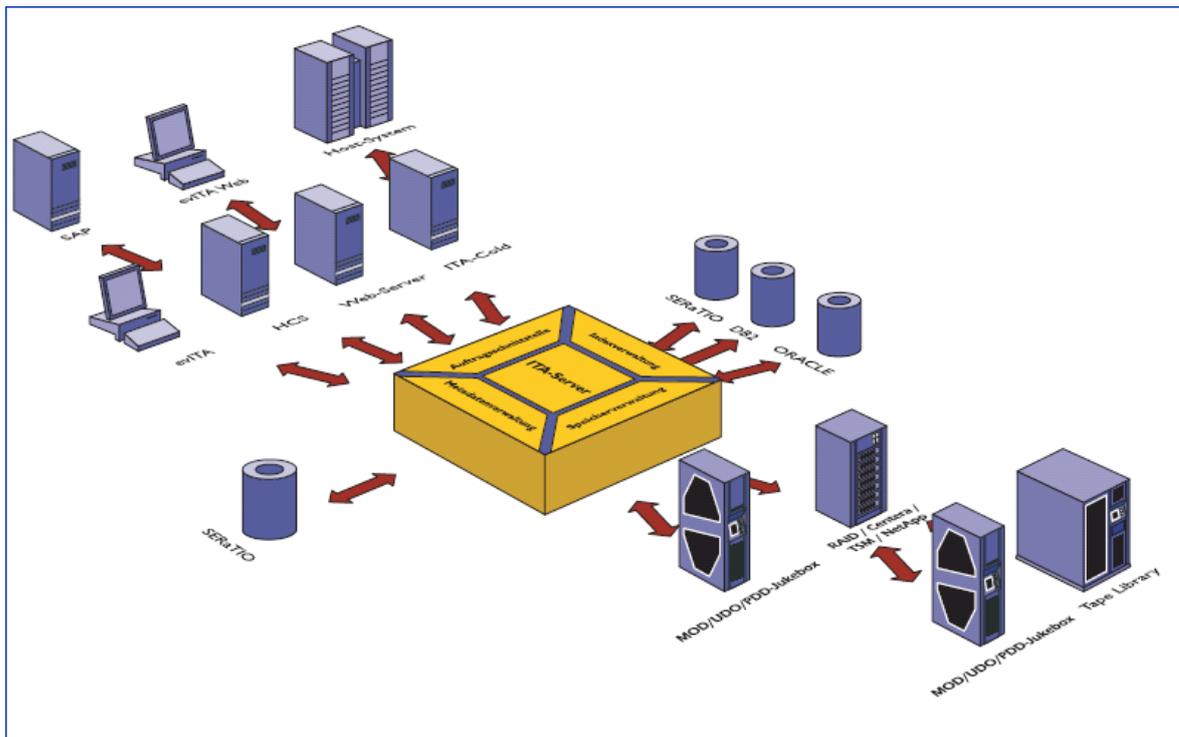


Abbildung 3.1: ITA-Archivsystem
Quelle [SER Solution 2007]

3.1.1.1 Die ITA-Server-Verwaltungssysteme

Medienverwaltung:

Die Medienverwaltung des Servers steuert die Speichergeräte und Medien und erlaubt wahlweise oder in Kombination das Speichern der Dokumente auf magneto-optische Platten oder Magnetplatten als primärem Speichermedium.

Der ITA-Server partitioniert Speichermedien in sogenannte Oberflächen. Die Oberflächen haben eine definierte Größe. Diese entspricht der Speicherkapazität auf einer Seite einer magneto-optischen Platte.

Mit ITA-Server können unterschiedliche Technologien für die Speicherung von Dokumenten herangezogen werden. Die folgenden Speichertechnologien werden derzeit unterstützt:

- Magneto Optical Disc (MOD)-Medien
- Ultra Density Optical (UDO)-Medien
- Professional Disc for Data (PDD)-Medien

- Dateisystem (Magnetplatte, RAID, etc.)
- IBM - Tivoli Storage Manager for Data Retention (TSM for DR)
- EMC Centera
- Network Appliance (NetApp) - SnapLock
- Advanced Intelligent Tape (AIT)-Magnetband

Indexverwaltung:

Die strukturierte Suche nach Dokumenten wird ermöglicht, indem die Indexverwaltung des Servers die Dokumentenattribute in Datenbanken ablegt.

Zum Speichern des Index wird das servereigene Datenbanksystem SERaTIO eingesetzt. Alternativ kann Oracle bzw. Datenbank 2 (DB 2) als relationales Datenbanksystem eingesetzt werden.

Metadatenverwaltung:

Die Metadatenverwaltung des ITA-Servers speichert anwendungsspezifische Definitionen wie zum Beispiel die Benutzer des Systems in der Systemdatenbank. Hierzu wird das servereigene Datenbanksystem SERaTIO eingesetzt.

Auftragsschnittstelle:

Über die Auftragsschnittstelle stellt der ITA-Server die Basisdienste von Dokumentenablage und -recherche den jeweils eingesetzten Clients zur Verfügung. Der ITA-Server kommuniziert mit den folgenden Standard-Clients:

Intuitiv-Client ist ein Windows-Client, der zuständig ist für die Erfassung, Indizierung, Ablage, Recherche, Weiterleitung und Bearbeitung von Dokumenten.

Intuitiv-Web-Client ermöglicht den Zugriff auf die archivierten Dokumente über den Web-Browser.

Ähnlich wie der Windows-Client verfügt das Programm über Funktionen wie Ablage, Recherche, Weiterleitung und Bearbeitung von Dokumenten. Es erlaubt neben dem firmeninternen Einsatz auch externen Anwendern wie Kunden oder Lieferanten den Zugriff auf archivierte Dokumente.

Computer Output on Laser Disk (COLD) Client

COLD Client ist ein auf die Massenarchivierung von digital erstellten Dokumenten spezialisierter Client. Neben der COLD-Archivierung ermöglicht das Programm auch die Batch-Archivierung von gescannten Dokumenten und Dateien.

SAP-Schnittstellen

Die Schnittstelle ITA-HTTPContentServer ermöglicht die Integration des SER-Archivsystems mit SAP R/3.

3.1.2 Dokumentenformate

Im SER-Archivsystem können Dokumente beliebigen Formats archiviert und verwaltet werden. Grundsätzlich besteht jedes Archivdokument aus zwei Teilen:

einem Datenteil: dem eigentlichen Dokument oder den Dokumentendaten;
einem Metadatenteil: dem Dokumenten-Header.

Die Dokumentendaten lassen sich abhängig von ihrem Datenformat in folgende Typen einteilen:

Non-Coded Information (NCI) Dokumente

Gescannte Dokumente, die aus Bildpunkten aufgebaut werden (Rasterbilder). NCI-Dokumente werden zum Beispiel per Scanner erfasst und dann mit dem Intuitiv-Client in das SER-Archivsystem archiviert.

Coded Information (CI) Dokumente

CI-Dokumente sind elektronisch erzeugte Dokumente, die auf einer Zeichensatztafel wie zum Beispiel American Standard Code for Information Interchange (ASCII) oder Extended Binary Coded Decimals Interchange Code (EBCDIC) oder auf einer Drucksprache beruhen.

CI-Dokumente werden mit ITA-COLD in das SER-Archivsystem archiviert.

Binärdateien

Binärdateien sind Dokumente, die elektronisch mit Anwendungsprogrammen erzeugt werden. Zu dieser Gruppe gehören zum Beispiel Dateien, die mit einem Textverarbeitungsprogramm wie Microsoft Word oder einem Grafikprogramm erstellt wurden. Jede dieser Anwendungen benutzt ein eigenes Datenformat.

Während NCI-Dokumente typischerweise aus externen Quellen stammen, werden CI-Dokumente und Binärdateien in der Regel vom Unternehmen selbst erstellt.

3.1.3 Der Dokumenten-Header

Der Dokumenten-Header besteht aus einem Satz von Schlüsseln unterschiedlichen Typs, die das Dokument beschreiben. Diese Schlüssel werden bei der Indizierung eines Dokuments mit Werten belegt.

Dokumentendaten und Dokumenten-Header werden bei der Erstarchivierung eines Dokuments zusammen auf den Oberflächen gespeichert. Der Dokumentenheader wird zusätzlich auf der Festplatte des ITA-Servers als Bestandteil des Dokumentenindex gespeichert.

Wird die Indizierung eines Dokuments nachträglich geändert (Schlüsseländerung), dann werden die Dokumentendaten nicht nochmals archiviert. Stattdessen wird nur ein neuer Header mit einem Verweis auf die Dokumentendaten des Originaldokuments gespeichert.

3.1.4 Strukturierte Ablage

Der ITA-Server archiviert Dokumente in eine hierarchisch gegliederte Ablagestruktur. Die Ablagehierarchie berücksichtigt sowohl organisatorische als auch technische Aspekte und wird kundenspezifisch eingerichtet.

Die folgenden Hierarchieebenen werden unterschieden:

Ersteller und Eigentümer der Dokumente

Der ITA-Server verwendet separate Systeme, um die Daten verschiedener Organisationseinheiten voneinander getrennt abzulegen. Diese Systeme dienen zum Beispiel zur Abbildung der Unternehmensbereiche (zum Beispiel verschiedene Abteilungen) eines Konzerns.

Inhaltliche Klassifizierung der Dokumente

Innerhalb eines Systems können Dokumente nach beliebigen inhaltlichen Gesichtspunkten den verschiedenen Datenbanken zugeordnet werden.

Jede Datenbank enthält wiederum eine Gruppe von verwandten Dokumentenklassen. Beispielsweise kann eine Datenbank Rechnungswesen eingerichtet werden, in der Bestellungen, Rechnungen und Lieferscheine voneinander getrennt abgelegt werden.

Ablagedatum der Dokumente

In einem Index werden alle Dokumente einer Datenbank zusammengefasst, die im selben zeitlichen Intervall archiviert wurden.

Aus der Datenbank „Rechnungswesen“ entstehen so beispielsweise die Indizes „Bestellungen im Jahr 2007“, „Bestellungen im Jahr 2008“ und „Bestellungen im Jahr 2009“. Die Länge des Intervalls wird in Abhängigkeit vom Dokumentenaufkommen gewählt: je größer das Volumen, desto kürzer der Zeitraum. Die Unterteilung einer Datenbank in Indizes, die jeweils die Dokumente eines bestimmten Zeitraums enthalten, optimiert die Performance bei Such- und Ablageoperationen. Weiterhin ermöglicht die Unterteilung einer Datenbank eine quasi unbegrenzte Datenhaltung und vereinfacht die Reorganisation der Daten.

Standardmäßig werden in dem ITA-Server die folgenden Zeiträume für Indizes unterstützt:

- Wochen
- Monate
- Jahre

3.1.5 Ablauf Archivierungsaufträge

Archivierungsaufträge laufen im Wesentlichen wie folgt ab:

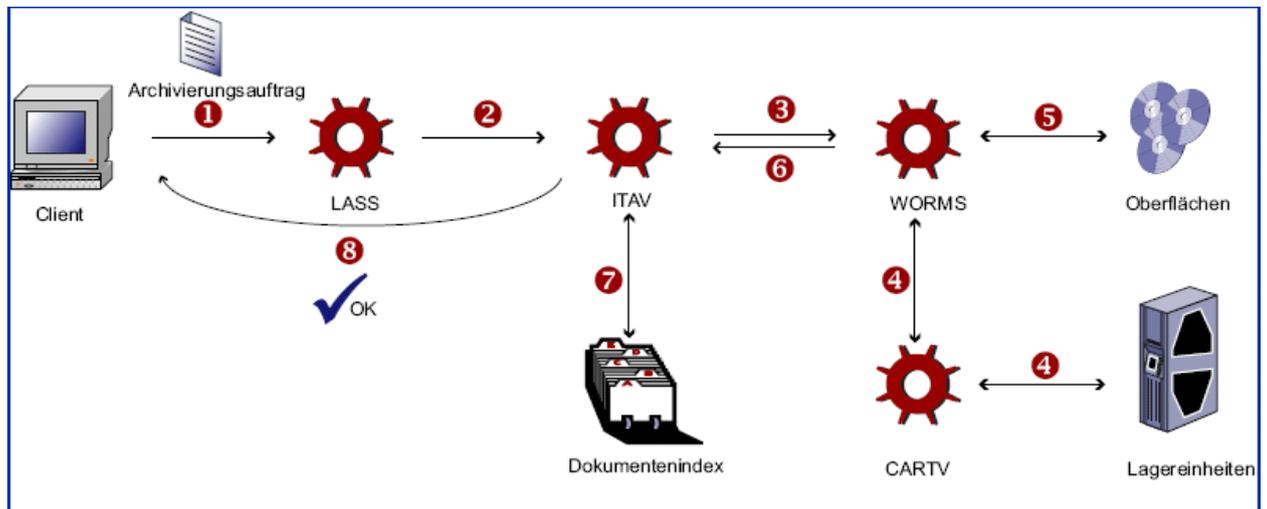


Abbildung 3.2: Ablauf Archivierungsauftrag
Quelle [SER Solution 2007]

1. Der Client schickt das zu archivierende Dokument an den ITA-Server.
2. Der Auftrag wird von dem Prozess LASS angenommen. Der LASS-Prozess ist für die Netzwerkkommunikation bei Archivierungs- und Rechercheaufträgen zuständig.
3. LASS leitet die Auftragsdaten an den ITAV-Prozess weiter. ITAV ist für Zugriffe auf die Dokumentendatenbanken des ITA-Servers verantwortlich.
4. ITAV leitet die Auftragsdaten an den WORMS-Prozess weiter, der für Zugriffe auf die Medien verantwortlich ist.
5. WORMS fordert die aktuelle Schreiboberfläche der entsprechenden Datenbank von dem CARTV-Prozess an. Der Prozess ist für Zugriffe auf die Lagereinheiten verantwortlich.
6. WORMS schreibt das Dokument auf die Oberfläche.
7. ITAV trägt das archivierte Dokument in den Index ein.
8. ITAV übermittelt eine O.-K.-Meldung an den Client.⁵

⁵ Vgl. SER Solutions(2007a).

3.1.6 Strukturierte Recherche

Als Basis für die spätere Suche nach individuellen Dokumenten wird bei der Ablage für jedes Dokument eine Metadatenbeschreibung erstellt. Diesen Vorgang nennt man Indizierung.

Die Metadaten (Schlüssel/Deskriptoren) enthalten spezifische Dokumenteigenschaften wie zum Beispiel Angaben über Autor, Titel oder Thema. Für jede einzelne Eigenschaft wird ein sogenannter Schlüssel definiert.

Je nach Verwendungszweck existieren unterschiedliche Arten von Schlüsseln. Der sogenannte Externe Primärschlüssel dient zur eindeutigen Identifizierung eines Dokuments.

Suchschlüssel dienen zur Formulierung von Dokumentenrecherchen. Dokumentenrecherchen liefern einen Satz von zu den Suchkriterien passenden Dokumenten.

Auswahlschlüssel helfen dabei, aus diesem Satz ein bestimmtes Dokument herauszusuchen. Zur Formulierung von Dokumentenrecherchen sind sie jedoch ungeeignet, da zum Beispiel zu viele Treffer erzielt würden. Suchschlüssel und Auswahlschlüssel beschreiben ein Dokument aus der Sicht des Anwenders. Sie werden daher auch Deskriptoren genannt (das lateinische Wort „describere“ steht für „beschreiben, schildern, darstellen“).

Die Suchschlüssel werden in den Suchbaum eines Dokumentenindex geschrieben, der außerdem noch zu jedem Dokument den Primärschlüssel enthält.

Bei einer Dokumentenrecherche findet eine SQL-Abfrage über die Einträge im Dokumentenindex statt.

Die Gesamtheit aller Metadaten eines Dokuments einschließlich Suchschlüssel, Auswahlschlüssel und Externem Primärschlüssel befindet sich im Dokumenten-Header. Dieser wird zusammen mit den Dokumentendaten auf die Speichermedien des ITA-Servers geschrieben. Als Ergebnis einer Dokumentenrecherche liefert der ITA-Server die Header der passenden Dokumente als sogenannte Trefferliste an den Client zurück. Über den Client kann der Anwender dann bestimmte Dokumente aus der Trefferliste vom Server anfordern und herunterladen.

3.1.6.1 Ablauf Rechercheaufträge

Dokumentenrecherchen laufen in zwei Schritten ab:

- die eigentliche Recherche
- das Retrieval

Rechercheaufträge laufen im Wesentlichen wie folgt ab:

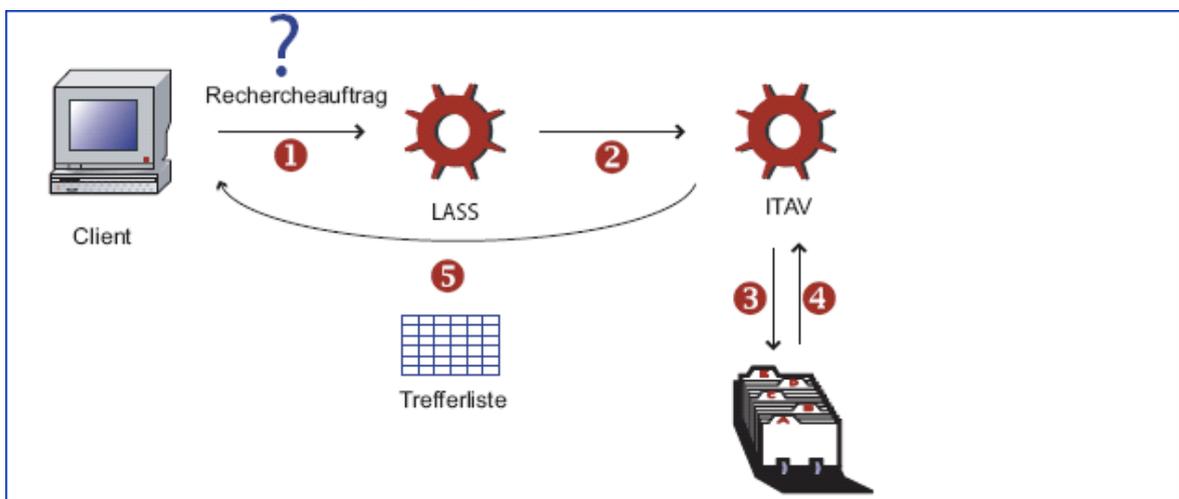


Abbildung 3.3: Ablauf Rechercheauftrag
Quelle: [SER Solution 2007]

1. Der Client schickt eine Anfrage an den ITA-Server. Der Auftrag wird vom LASS-Prozess angenommen.
2. Der LASS-Prozess leitet die Auftragsdaten an den ITAV-Prozess weiter.
3. Der ITAV-Prozess recherchiert im Dokumentenindex nach passenden Dokumenten.
4. Als Ergebnis der Recherche übermittelt der ITAV-Prozess eine Trefferliste an den Client.⁶

⁶ Vgl. SER Solutions(2007a).

3.1.6.2 Ablauf Retrieval-Aufträge

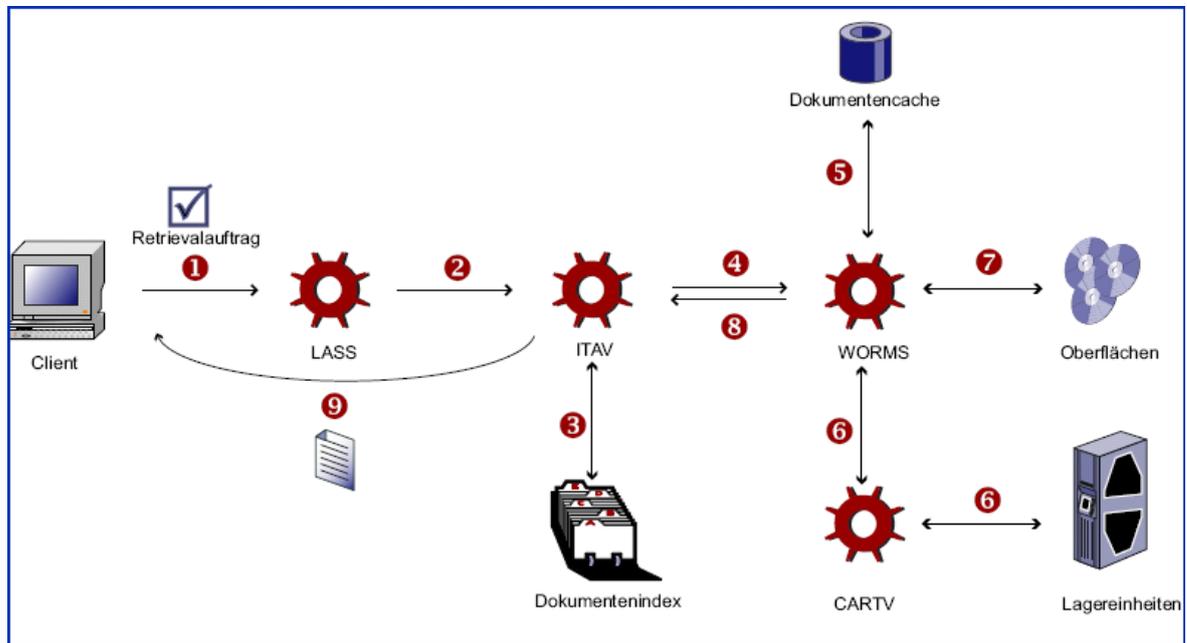


Abbildung 3.4: Ablauf Retrieval-Auftrag

Quelle: [SER Solution 2007]

Aufträge zum Retrieval laufen im Wesentlichen wie folgt ab:

1. Der Client schickt eine Anfrage an den ITA-Server. Der Auftrag wird vom LASS-Prozess angenommen.
2. Der LASS-Prozess leitet die Auftragsdaten an den ITAV-Prozess weiter.
3. Der ITAV-Prozess verwendet den Externen Primärschlüssel aus den Auftragsdaten, um im Dokumentenindex nach dem Internen Primärschlüssel des Dokuments zu suchen.
4. Der ITAV-Prozess übermittelt den Internen Primärschlüssel an den WORMS-Prozess.
5. Der WORMS-Prozess durchsucht im ersten Schritt den Dokumentencache nach dem Dokument.
6. Befindet sich das Dokument nicht im Cache, dann fordert der WORMS-Prozess die entsprechende Oberfläche vom CARTV-Prozess an.
7. Der WORMS-Prozess greift von der jeweiligen Oberfläche das Dokument heraus.
8. Der WORMS-Prozess übermittelt das Dokument an den ITAV-Prozess.
9. Der ITAV-Prozess schickt das Dokument an den Client.⁷

⁷ Vgl. SER Solutions(2007a).

4 Archivierungsvorgang Architektur bei der BGW

Bei der BGW benutzen verschiedene Umgebungen das Archivsystem gemeinsam. Der ITA-Server verwendet separate Systeme, um die Daten dieser Umgebungen voneinander getrennt abzulegen.

Jede Umgebung hat einen eigenen Benutzerstamm und verwendet eigene Attribute zur Charakterisierung der Dokumente.

Innerhalb einer Umgebung können Dokumente nach beliebigen inhaltlichen Gesichtspunkten in Datenbanken zusammengefasst werden.

Bei der Konfiguration des Archivs wurde im ITA-Server unter anderem die Umgebung Produktion (PROD) definiert, diese wurde in vier unterschiedliche Bereiche (Finanz, Prävention, Reha und UB) unterteilt.

Jeder Bereich hat einen eigenen Satz von Datenbanken. Jede Datenbank verfügt wiederum über einen Standardpool und einen Reservepool, auf deren Medien die entsprechenden Dokumente archiviert werden.

Die folgende grafische Darstellung zeigt die Struktur von Systemen und ihrer Umgebung.

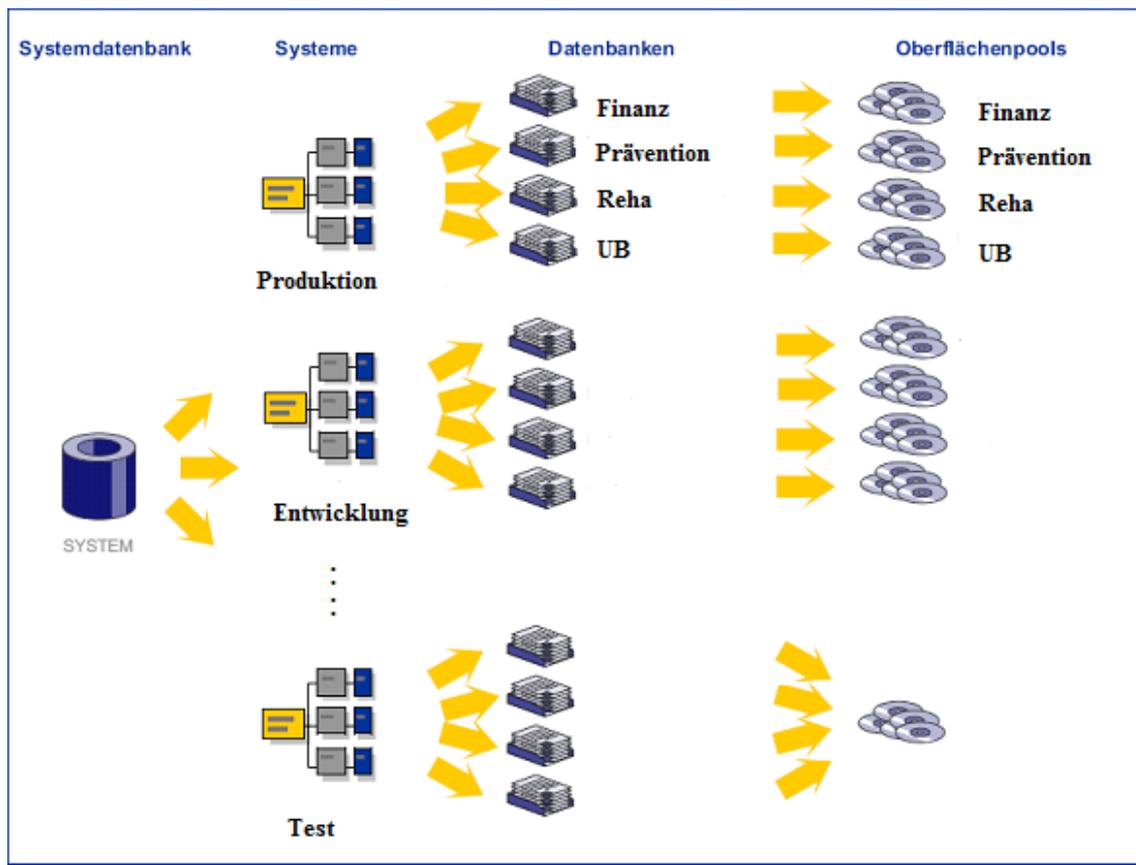


Abbildung 4.1: BGW Archivserver-Konfiguration
Vgl.: [SER Solution 2007]

4.1 Systemüberblick

Jeden Tag werden bei der BGW während der Realisierung von Geschäftsprozessen zahlreiche Dokumente erzeugt. Diese Dokumente werden als sogenannte Postausgangsdokumente bezeichnet.

Zu den Posteingangsdokumenten zählen sämtliche eingehenden Rechnungen, die in Papierform bei der BGW eintreffen.

Um die Posteingänge und -ausgänge zu verarbeiten, sind folgende Komponenten bei der BGW im Einsatz:

4.1.1 Einzugsscanner

Der Scanner mit integriertem FTP-Modul ist an das Netzwerk der BGW angeschlossen und ermöglicht die Übertragung von Daten auf den EDM-Server.

4.1.2 Elektronischer Dokumenten Manager (EDM) Server

Der EDM-Server wird zur Ablage und Vorverarbeitung von zu archivierenden Dokumenten aus Posteingangs- und Postausgangsverfahren verwendet. Der Server wird als sogenannter OCR-Server eingesetzt. Die zu archivierenden Dokumentenformen bestehen aus verschiedenen Anwendungen wie z. B. Form Text, Word BGS usw.

Die Dokumente werden durch eine OCR-Lesung im TIFF-Format konvertiert und geprüft.

Der Vorteil des EDM-Servers besteht darin, dass eine asynchrone Archivierung ermöglicht wird. Außerdem wird erreicht, dass bei Ausfall des Archivservers die zu archivierenden Dokumente zwischengespeichert werden. Die Archivierung findet später statt.

Der Nachteil des EDM-Servers besteht darin, dass bei Ausfall das ganze System zusammenstürzt.

4.1.3 Intuitiv-Client

Intuitiv ist ein SER-Archiv-Client mit Benutzeroberfläche für Posteingangsarchivierung, Rechercheaufträge und Retrieval-Aufträge.

4.1.4 ITA-Archivserver

Der ITA-Server stellt Anwendungen zur Verfügung, um die Posteingangs- und Postausgangsdokumente zu archivieren.

4.1.5 BG-Standard

BG-Standard wurde 2003 bei der BGW eingeführt und ist das zentrale Informationssystem. BG-Standard ist eine Standard-Softwarelösung für Unfallversicherungsträger. In der Anwendung BG-Standard werden u. a. Geschäftsprozesse der Berufsgenossenschaft für Gesundheitsdienst und Wohlfahrtspflege verwaltet.

4.1.6 Computer Output on Laser Disk (COLD) Client

ITA-COLD ist ein auf die Massenarchivierung von digital erstellten Postausgangsdokumenten spezialisierter Client.

4.2 Posteingangsarchivierung

Die Posteingangsdokumente sind Rechnungen, die in nichtdigitalisierter Form bei der BGW mit Scanner erfasst werden und dann mit Intuitiv in das SER-Archivsystem archiviert werden. Folgende Schritte sind für die Posteingangsarchivierung nötig:

1. Auf der jeweils ersten Seite einer Rechnung wird ein Barcode aufgeklebt. Der Zielpfad wird durch Eingabe am Scanner ausgewählt und auf den EDM-Server übertragen.

Der EDM-Server ist so aufgeteilt, dass jede Umgebung, z. B. Produktion und ihre Bereiche Finanz, Reha, UB und Prävention, ihr eigenes Zielverzeichnis haben.

Dann können beliebig viele Dokumente über Scanner per FTP-Transfer an den EDM-Server übertragen werden.

Pro Scan-Durchgang wird ein Multipage-TIFF- und eine Scanner-IDX-Datei an den EDM-Server übertragen.

2. In dem EDM-Server wird die Multipage-TIFF-Datei von dem Konvertermanager überprüft. Die TIFF-Datei muss lesbar und darf nicht defekt sein. Bei einer fehlerhaften Datei wird sie in das dafür vorgesehene „Error“-Verzeichnis verschoben. In der Multipage-TIFF-Datei muss pro Dokument ein Barcode vorhan-

den sein. Beim Auslesen des Barcodes wird pro Dokument eine neue IDX-Datei mit einem Barcode erzeugt. Im Anschluss wird pro Dokument eine Multipage-TIFF-Datei erzeugt, danach werden die Dokumente in dem „Work“-Verzeichnis auf dem EDM-Server abgelegt.

- Um die Posteingangsdokumente aus dem EDM-Server zu archivieren, stellt SER die Dienste des Intuitiv-Client bereit. Der Benutzer startet die Benutzeroberfläche des Intuitiv-Client für Posteingangsarchivierung, und es folgt eine manuelle Verarbeitung der Dokumente.

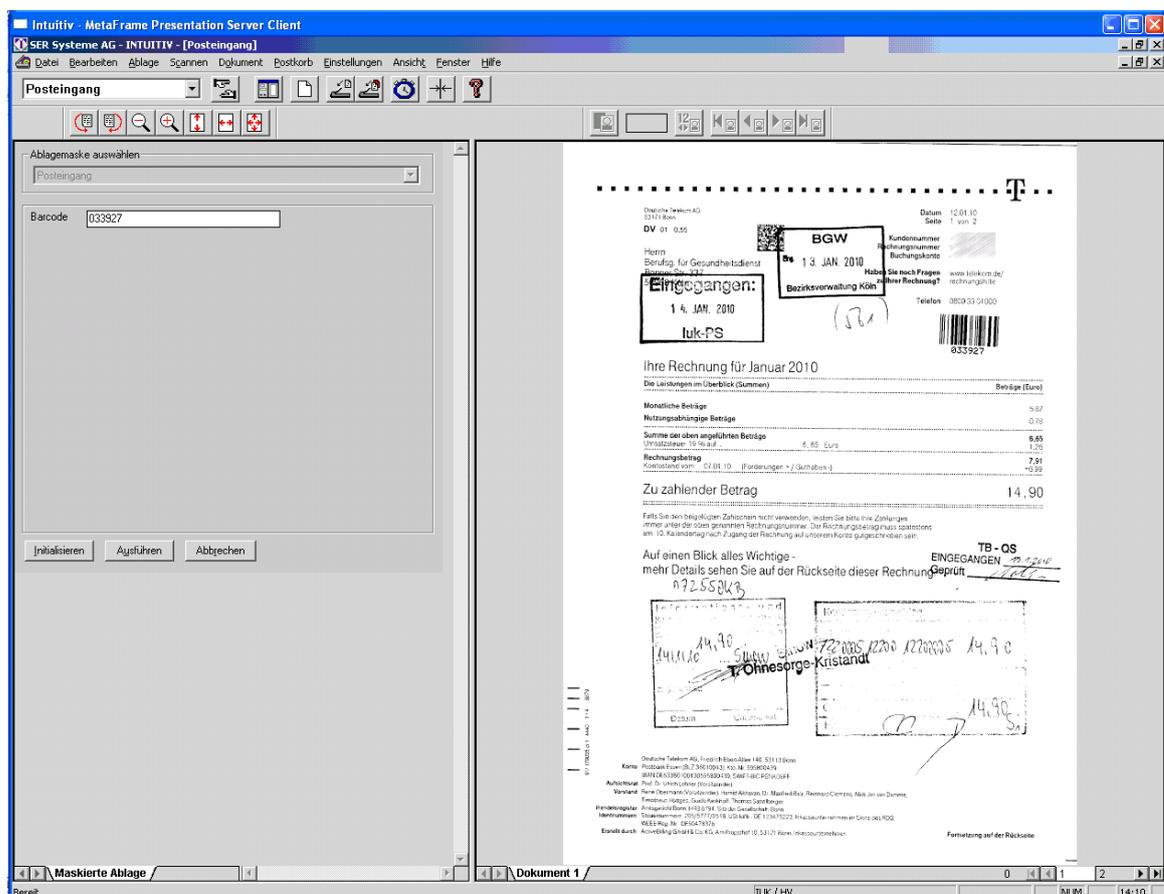


Abbildung 4.2: Intuitiv-Client

Der Benutzer überprüft, ob das Dokument durch die Konvertierung lesbar ist und ob der Barcode richtig erkannt wurde. Fehlerhafte Dokumente werden von dem Benutzer ins „Error“-Verzeichnis auf dem EDM-Server verschoben. Bei fehlerfreien Dokumenten wird die Belegnummer (Posteingangsnummer) automatisch generiert. Die Belegnummer setzt sich aus der Nummer der jeweiligen Bezirksverwaltung (BV), dem Tagesdatum und dem Barcode zusammen.

Folgende Tabelle zeigt den Aufbau der Belegnummer:

Zeichen 1 und 2	BV-Nummer 01–12
Zeichen 3 bis 10	Tagesdatum tt mm jjjj
Zeichen 11 bis 15	Barcode Nr. 00000–99999

Tabelle 4.1: Aufbau der Posteingangsbelegnummer

Quelle: [BGW 2009a]

Die Belegnummer wird als sogenannter Externer Primärschlüssel benutzt, um das Dokument in dem Archivsystem eindeutig zu identifizieren.

4. Der Benutzer legt durch den Intuitiv-Client die Umgebung und den Bereich im Archivserver fest. Korrekte Dokumente werden sofort im Archiv abgelegt. Gleichzeitig wird die Belegnummer in der Transferdatei auf dem EDM-Server gespeichert.
5. Die automatische Übertragung der Belegnummern an BG-Standard wird gestartet, wenn die hinterlegte Dokumentengrenze (Standard 50) erreicht wurde oder der Benutzer die Client-Anwendung beendet hat.

Der Transfer der Belegnummern an BG-Standard erfolgt nun.

Die BG-Standard-Sitzung wird im Hintergrund unsichtbar geöffnet. Die gespeicherten Transferdaten werden an BG-Standard gesendet und an das zugehörige Scan-Postfach übermittelt. Anschließend wird die Transferdatei gelöscht und die BG-Standard-Sitzung beendet.

6. Der Benutzer startet die BG-Standard Sitzung und öffnet das Scan-Postfach. Alle für den Bereich gescannten Dokumente (Belegnummern) werden angezeigt. Der Benutzer leitet die einzelnen Dokumente (Belegnummern) an die zugehörigen Sachbearbeiter der Abteilung weiter. Der Sachbearbeiter öffnet sein Postfach und erfasst die hinterlegten Dokumente (Belegnummern).

Folgende grafische Darstellung verdeutlicht den Vorgang der Posteingangsarchivierung:

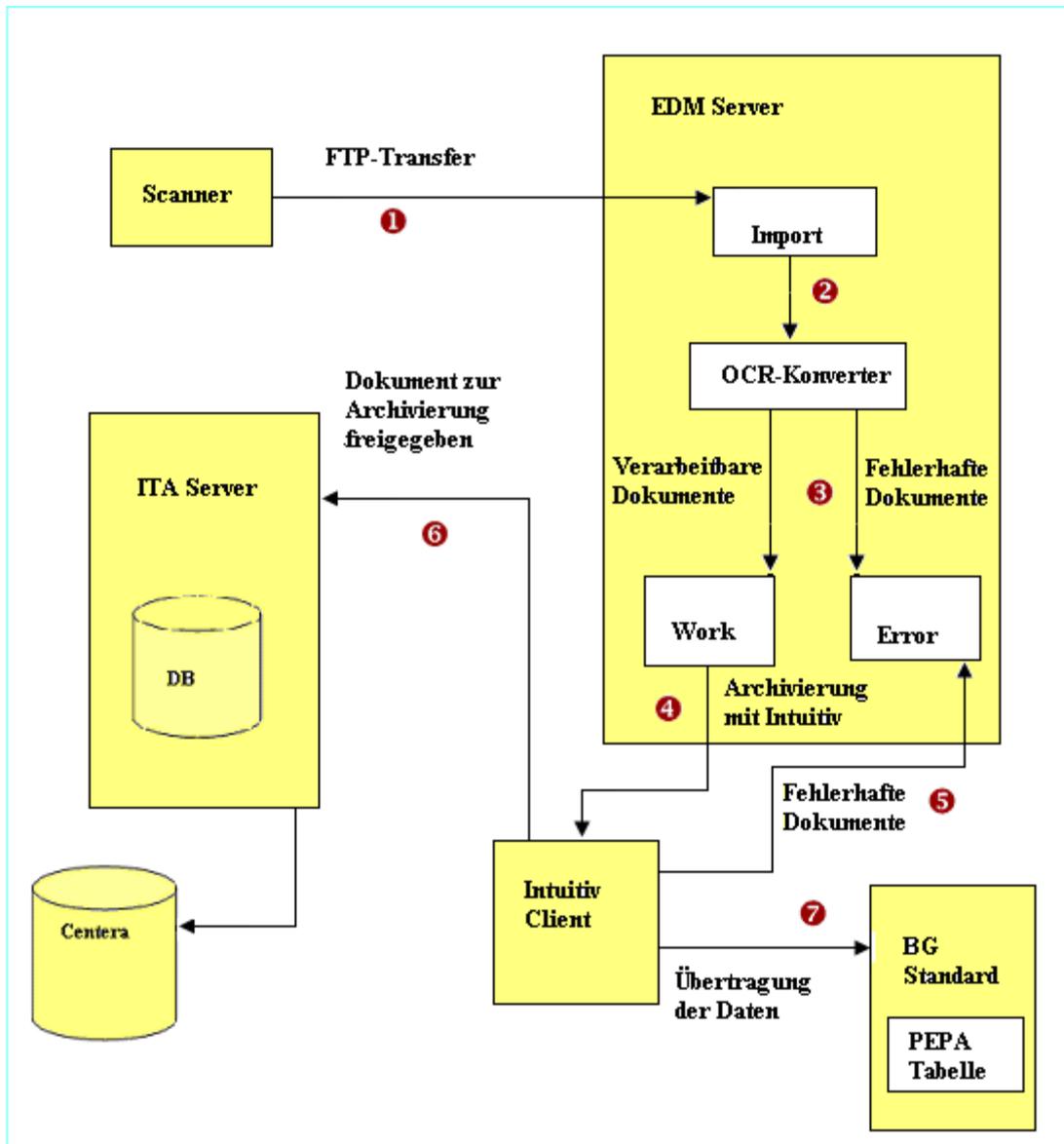


Abbildung 4.3: Posteingangsarchivierung
Vgl.: [BGW 2009a]

1. Durch den Scanvorgang werden die Dokumente in eine Multipage-TIFF-Datei konvertiert und auf den EDM-Server übertragen.
2. Die Datei wird durch den Konvertermanager überprüft.
3. Bei einer fehlerhaften Datei wird diese in das dafür vorgesehene „Error“-Verzeichnis verschoben.
4. Starten des Intuitiv-Clients und Sichtung der Dokumente und manuelle Verarbeitung durch den Benutzer.
5. Fehlerhafte Dokumente werden in das „Error“-Verzeichnis abgelegt.
6. Korrekte Dokumente werden sofort im Archiv abgelegt.
7. Automatische Übertragung der Daten an BG-Standard.

4.3 Postausgangsarchivierung

Im Gegensatz zu den Posteingangsdokumenten werden die Postausgangsdokumente bei der BGW selbst erstellt. Diese Dokumente haben verschiedene Formate und werden auch wie Posteingangsdokumente im TIFF-Format archiviert.

1. Die Postausgangsdokumente werden aus einem Vorgang heraus erzeugt. In BG-Standard wählt der Benutzer für die Erstellung eines Dokuments das gewünschte Aktenzeichen und danach den gewünschten Dokumententyp aus.

Daraufhin startet BG-Standard die Subanwendungen Word oder Form Text Filler.

Außerdem werden auch Postausgangsdokumente durch BG-Standard in bestimmten Vorgängen vollautomatisiert erzeugt, welche dann auch gedruckt und archiviert werden.

2. Der Benutzer druckt das erstellte Dokument endgültig aus (kein Probedruck). Der allgemeine Druckvorgang in BG-Standard erzeugt eine Belegnummer für den Postausgang und legt eine IDX-Datei an, die die Belegnummer (Postausgangnummer) und andere Indexwerte enthält. Die Belegnummer setzt sich aus dem Buchstaben A für Postausgang, BV, Tagesdatum und Barcode zusammen.

Folgende Tabelle zeigt den Aufbau der Belegnummer:

Zeichen 1	Kennzeichen A = Postausgang
Zeichen 2 und 3	BV-Nummer 01–12
Zeichen 4 bis 11	Tagesdatum tt mm jjjj
Zeichen 12 bis 16	Barcode Nr. 00000–99999

Tabelle 4.2: Aufbau der Postausgangsbelegnummer

Quelle: [BGW 2009a]

Auch hier wird die Belegnummer als Externer Primärschlüssel benutzt, um die Dokumente im Archivsystem eindeutig zu identifizieren.

3. Der Archivierungsauftrag wird erst aktiviert, wenn der Benutzer in BG-Standard ein Schreiben endgültig ausgedruckt hat.

Die Postausgangsdokumente werden mit den dazugehörigen IDX-Dateien an den EDM-Server gesendet und in die jeweiligen Umgebungen und die dazugehörigen Bereiche gespeichert.

Im EDM-Server werden die Dokumente durch die Konvertermanager verarbeitet. Folgende Anwendungen werden durchgeführt:

Der Konvertermanager überprüft, ob das Quelldokument lesbar und nicht gesperrt ist. Außerdem wird überprüft, ob alle Voraussetzungen erfüllt sind (z. B. ob für eine XML-Datei ein XML-Schema vorhanden ist).

Nach positiver Auswertung wird das Dokument per Image-Drucker in Multi-page-TIFF konvertiert und anschließend auf Korrektheit überprüft. Für jedes TIFF-Dokument wird eine neue IDX-Datei mit Indexwerten erzeugt. Nach erfolgreicher Konvertierung wird eine RDY-Datei als Indikator erzeugt, und die Ursprungsdatei wird gelöscht.

4. Der ITA-Server holt die zu archivierenden Dokumente vom EDM-Server und archiviert sie mit ITA-COLD. Dieser Archivierungsvorgang wird von ITA-COLD automatisch geprüft und durchgeführt. Dabei wird geprüft, ob TIFF- und IDX-Datei lesbar sind und ob die Dateien übereinstimmen. Außerdem wird geprüft, ob alle Pflichtfelder in der IDX-Datei gefüllt sind.

Nach diesen Überprüfungen wird eine Auswertung gestartet, um das Archivdokument in die dazugehörige Umgebung (z. B. PROD) und den dazugehörigen Bereich (z. B. Finanz, Reha, UB oder Prävention) zu archivieren.

Das Archivdokument wird im dazugehörigen Teilbestand archiviert. Bei einer fehlerfreien Archivierung wird die Ursprungsdatei gelöscht.

Folgende grafische Darstellung verdeutlicht die Postausgangsarchivierung:

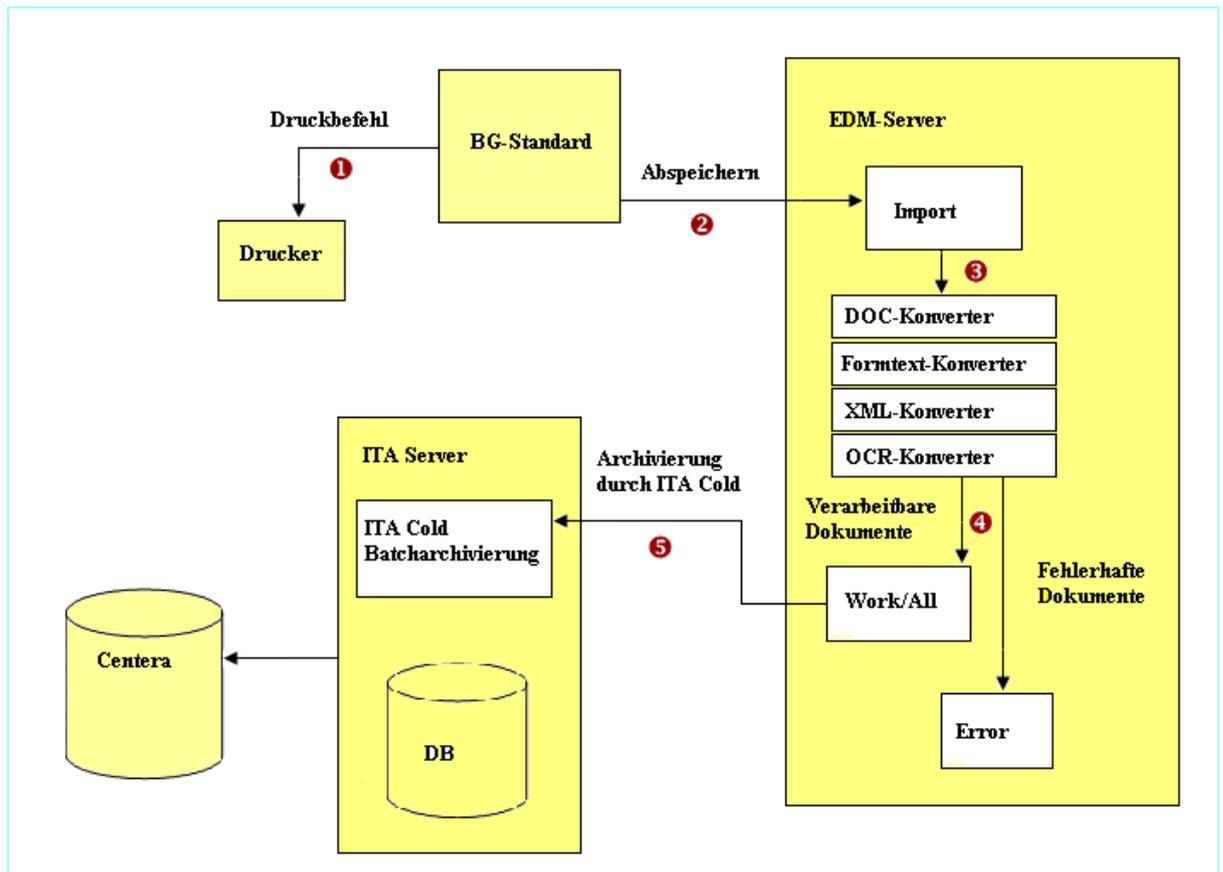


Abbildung 4.4: Postausgangsarchivierung
Vgl.: [BGW 2009a]

1. Der Benutzer druckt das erstellte Dokument endgültig aus (kein Probedruck).
2. Die Postausgangsdokumente werden mit den dazugehörigen IDX-Dateien an den EDM-Server gesendet.
3. Im EDM-Server werden die Dokumente durch den jeweiligen Konvertermanager verarbeitet.
4. Fehlerhafte Dokumente werden in das „Error“-Verzeichnis verschoben.
5. Archivierung der Dokumente durch ITA-COLD.

4.4 Dokumentenrecherche

Für die Archivsuche nach individuellen Dokumenten stellt das SER-Archivsystem die Deskriptoren zur Beschreibung von Dokumenten aus der Sicht des Anwenders zur Ver-

fügung. Außerdem wird im SER-Archivsystem jedes Dokument für eine eindeutige Identifizierung zum Information Retrieval mit einem sogenannten Externen Primärschlüssel beschrieben.

Bei der BGW werden die Dokumente mit den Deskriptoren bereichsabhängig beschrieben. In den Bereichen Finanz, Prävention und Reha werden die Dokumente mit den Deskriptoren Belegnummer und Archivierungsdatum belegt. In diesen drei Bereichen ist keine direkte Archivsuche möglich, da innerhalb von BG-Standard mit anderen Schlüsseln gearbeitet wird. Hier wird eine Umschlüsselung über Tabellen durchgeführt, um an die Belegnummer zu gelangen, mit der der Archivaufruf initiiert wird. Innerhalb von BG-Standard wird z. B. über die Posteingangs-/Postausgangs- (PEPA) Übersicht gearbeitet. Alle darin angezeigten Werte werden aus internen Tabellen gezogen, ohne einen einzigen Archivzugriff.

06-00-WA000200 Kloed, Rainer geb 13.08.1945 VersFall 13.11.1963						
Datum von		16.04.2008	Datum bis		16.10.2007	Notiz(en) vorhanden <input type="checkbox"/>
KZ	Belegnummer	Datum	Formular	Status	Empfangen von / abgewandt an	EDM-Status
A	A05-18.10.2007-00031		A2050 - Allgem. Erinnerung mit Rückantwort	angelegt	Rainer Klöed Pseudo Kackluck, 60407 Frankfurt	archiviert
A	A05-07.02.2007-00032		04320 - Rechnungsürzung	angelegt	Walseral-Apothek Franz Felder, Walserstr. 51, 87567 Riedern	angelegt
E	E08-07.02.2007-14872		07000 - Rechnung	angelegt	Walseral-Apothek Franz Felder, Walserstr. 51, 87567 Riedern	
E	E08-06.02.2007-14061		07000 - Rechnung	angelegt	Walseral-Apothek Franz Felder, Walserstr. 51, 87567 Riedern	
E	E08-05.02.2007-18000		07000 - Rechnung	angelegt	Walseral-Apothek Franz Felder, Walserstr. 51, 87567 Riedern	
E	E08-05.02.2007-18001		07000 - Rechnung	angelegt	Walseral-Apothek Franz Felder, Walserstr. 51, 87567 Riedern	
E	E08-05.02.2007-18003		07000 - Rechnung	angelegt	Walseral-Apothek Franz Felder, Walserstr. 51, 87567 Riedern	
E	E08-05.02.2007-14068		07000 - Rechnung	angelegt	Walseral-Apothek Franz Felder, Walserstr. 51, 87567 Riedern	
E	E08-13.12.2006-14064		H4126 - V1 Übersetzung Diagnose mit Rechnung	angelegt	Walseral-Apothek Franz Felder, Walserstr. 51, 87567 Riedern	

Abbildung 4.5: PEPA-Übersichtstabelle
Quelle: [BGW 2009a]

Es findet in diesen drei Bereichen eigentlich keine Recherche, sondern ein Information Retrieval (Informationswiedergewinnung) statt.

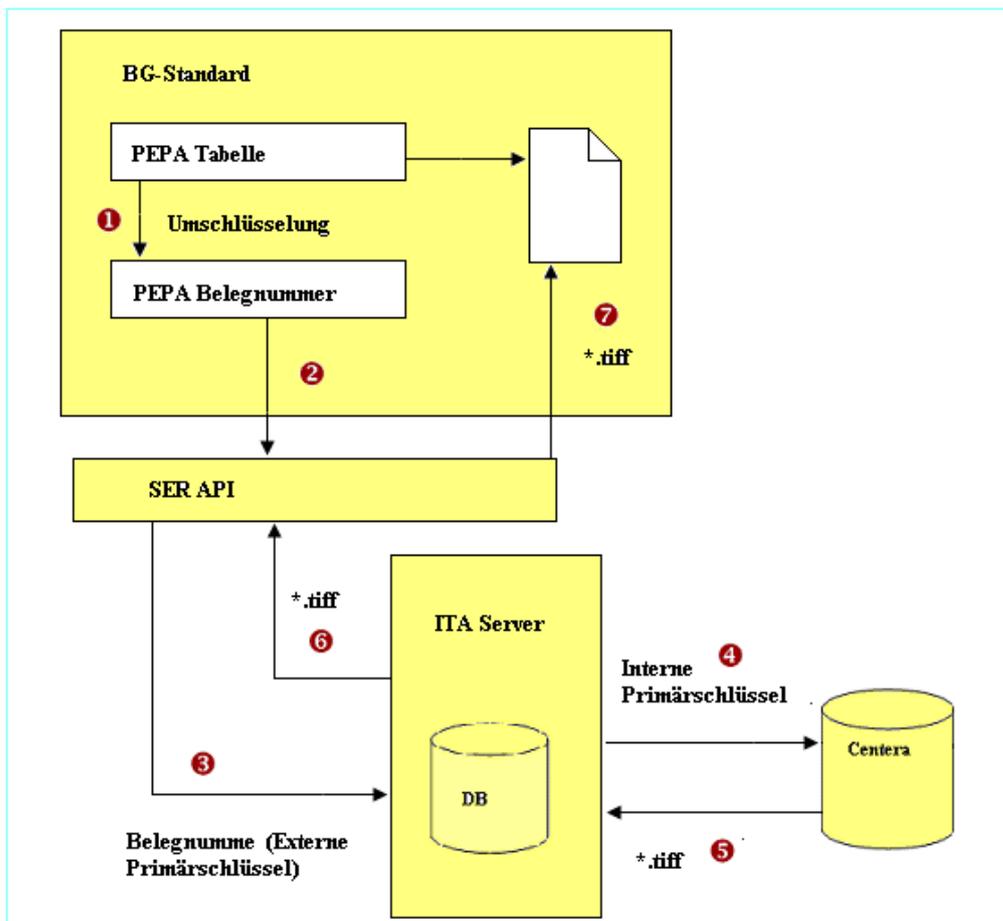


Abbildung 4.6: Recherche über PEPA-Tabelle
Vgl.: [BGW 2009a]

1. Der Benutzer wählt über die PEPA-Tabelle das gewünschte Dokument aus. Es findet eine Umslüsselung über Tabellen statt, um an die Belegnummer zu gelangen.
2. Die Belegnummer (Externer Primärschlüssel) wird an API des ITA-Servers gesendet.
3. Anhand des Externen Primärschlüssels wird der Interne Primärschlüssel ermittelt.
4. Dokumentensuche anhand der Internen Primärschlüssel in Centera.
5. Als Ergebnis wird das TIFF-Dokument an den ITA-Server gesendet.
6. Das TIFF-Dokument wird an BG-Standard übermittelt.
7. Der Benutzer kann das Dokument für die Weiterbearbeitung benutzen.

Im Bereich Unternehmerbetreuung (UB) werden die Archivdokumente anhand folgender Deskriptoren beschrieben:

- Archivierungsdatum
- Dokumentendatum

- Geschäftsvorfall
- Dokumentenart
- Absender
- Aktenzeichen
- Belegnummer

Nur in diesem Bereich ist eine direkte Archivsuche möglich, da innerhalb von BG-Standard mit denselben Indexschlüsseln gearbeitet wird. Hier wird, im Gegensatz zur PEPA-Übersicht, nach der Eingabe eines Aktenzeichens direkt im Archiv gesucht. Die Trefferliste wird vom Archiv an BG-Standard übergeben. Aus dieser Trefferliste kann dann das gesuchte Dokument ausgewählt werden, dadurch wird die Belegnummer an das Archiv gesendet, und das Dokument wird angezeigt.

Kriterien					
Fallnummer		M462452H00		Anzahl der Belege:	
Suchergebnisse					
Eingangsdatum	Erstellungsdatum	Absender	Dokumentart	Geschäftsvorfall	Belegnummer
	11.09.2007	UB	MUB181 N	BE	UW07091116104753
	11.09.2007	UB	MUB058	BE	UW07091116104770
	08.08.2007	UB	MUB181 N	BE	UW07080813591353
	08.08.2007	UB	MUB058	BE	UW07080813591370
	07.08.2007	UB	MUB159A	MA	UW07080711131070
	15.05.2007	UB	MUB181 N	BE	UW07051517073753
	15.05.2007	UB	MUB058	BE	UW07051517073770
	29.03.2007	UB	MUB181 N	BE	UW07032911213553

Suche starten/aktualisieren Indizieren Anzeigen Drucken Ende -> Postf Hilfe

Abbildung 4.7: Direkte Archivrecherche im Bereich UB
Quelle: [BGW 2009a]

Folgende grafische Darstellung verdeutlicht die direkte Archivrecherche im Bereich UB:

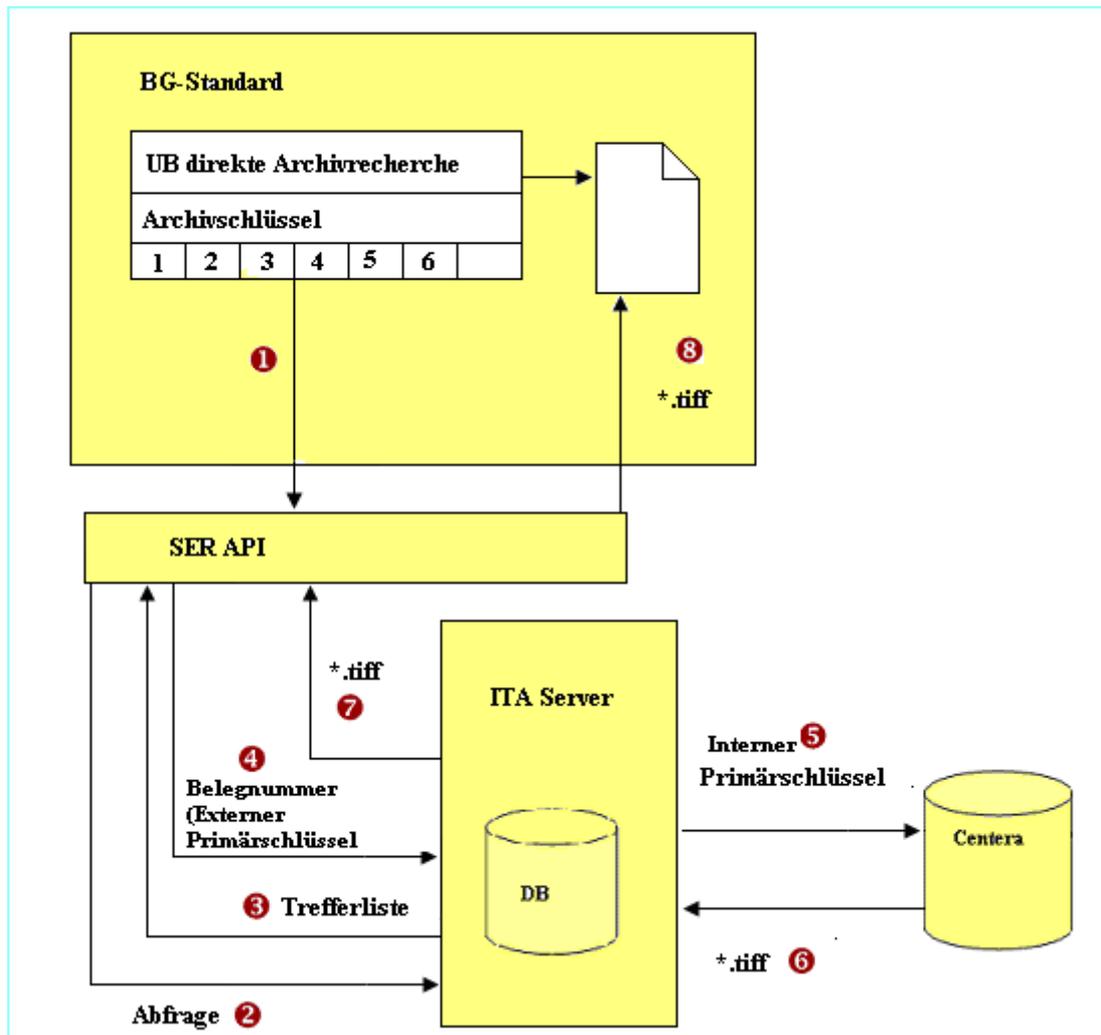


Abbildung 4.8: Direkter Recherchevorgang im Bereich UB
Vgl.: [BGW 2009a]

1. Der Benutzer recherchiert anhand derselben Indexschlüssel direkt im Archiv.
2. Die Abfrage wird an ITA-Server gesendet.
3. Als Ergebnis der Recherche wird eine Trefferliste an BG-Standard übermittelt.
4. Aus dieser Trefferliste kann dann das gesuchte Dokument ausgewählt werden, dadurch wird der Externe Primärschlüssel an das Archiv gesendet.
5. Anhand des Externen Primärschlüssels wird der Interne Primärschlüssel ermittelt.
6. Dokumentensuche anhand des Internen Primärschlüssels in Centera.
7. Das TIFF-Dokument wird an BG-Standard übermittelt.
8. Der Benutzer kann das Dokument für Weiterbearbeitung benutzen.

5 Analyse der Anforderungen

5.1 BG-Standard-Ausfall

Die Recherche und Informationswiedergewinnung von Dokumenten erfolgt zurzeit nur über BG-Standard. Im Falle eines Ausfalls oder einer Ablösung von BG-Standard ist eine effiziente Archivsuche nicht in allen Bereichen gegeben. Hierfür sind die Auswahl und Verfügbarkeit von Deskriptoren verantwortlich.

In den Bereichen Finanz, Prävention und Reha führt eine direkte Archivsuche mit Intuitiv-Client zu einer Vielzahl an Treffern in der Trefferliste. Das Problem für diese ungenaue Archivsuche liegt darin, dass nur das Archivierungsdatum als einziger Deskriptor zur Verfügung steht.

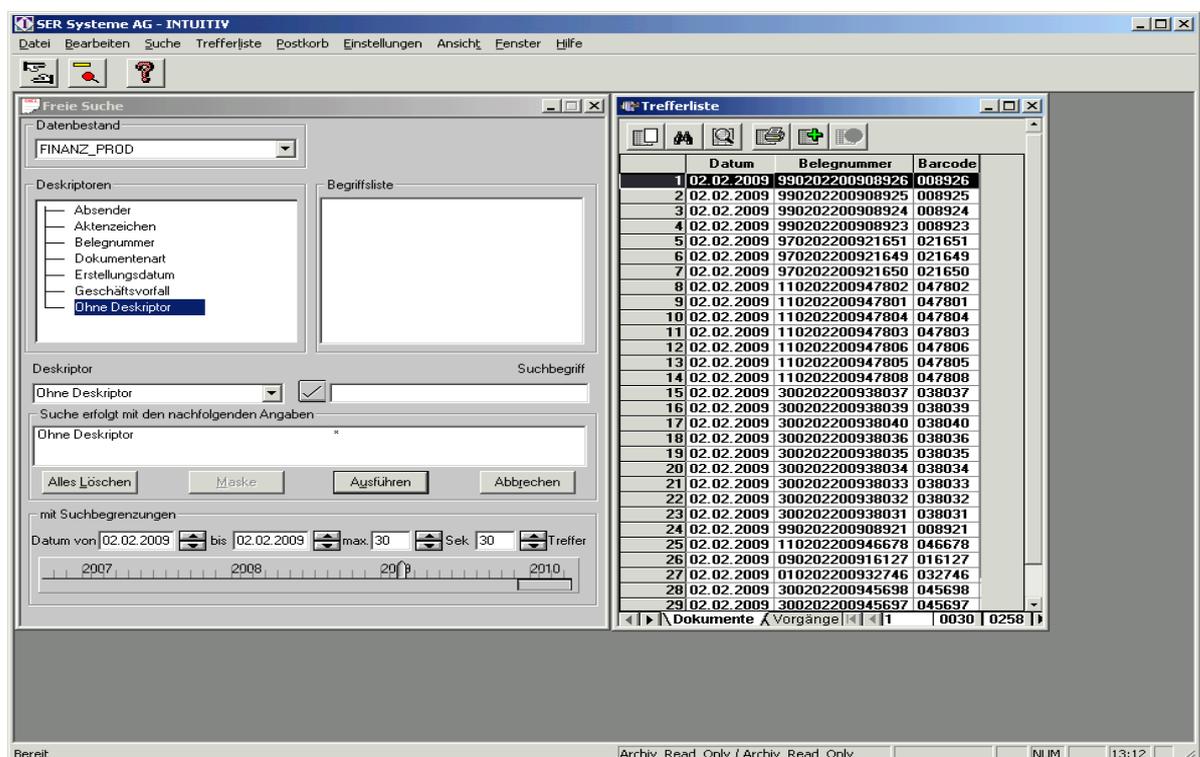


Abbildung 5.1: Direkte Archivsuche im Bereich Finanz

Deskriptor: Archivierungsdatum 02.02.2009 bis 02.02.2009 im Bereich Finanz.
Ergebnis der Trefferliste: 258 Treffer.

Im Bereich UB führt die direkte Archivsuche wie in BG-Standard zu denselben Ergebnissen und Resultaten. Hier stehen ebenso wie in BG-Standard dieselben Deskriptoren zur Verfügung.

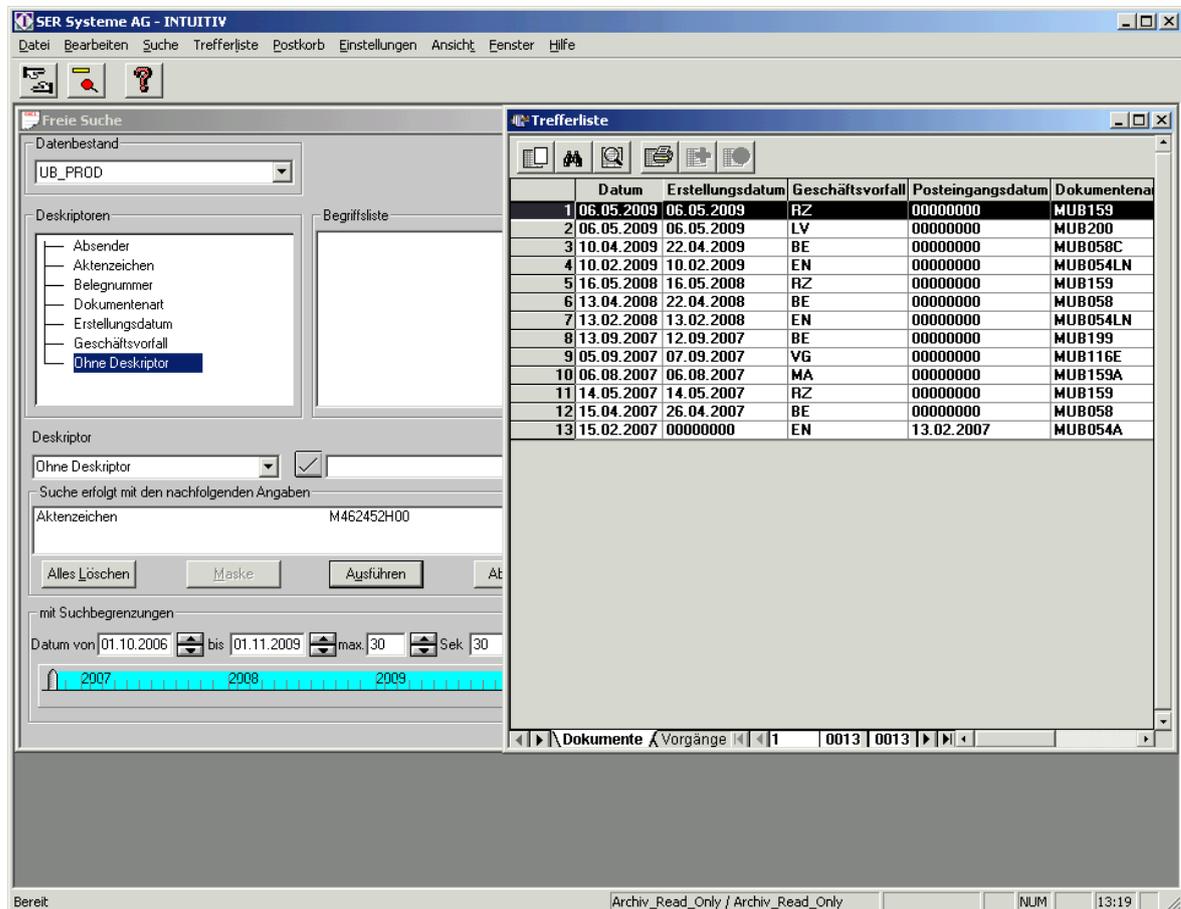


Abbildung 5.2: Direkte Archivrecherche im Bereich UB

Deskriptoren: Aktenzeichen M462452H00,
 Archivierungsdatum: 01.10.2006 bis 01.11.2009 Bereich UB.
 Ergebnis der Trefferliste: 13 Treffer.

Um eine effiziente Recherche auch in den Bereichen Finanz, Prävention und Reha zu ermöglichen, werden folgende Lösungswege untersucht und bewertet:

- Volltextsuche
- NachIndizierung von Dokumenten

5.2 Archivsuche zur Erstellung von heute nicht bekannten Auswertungen über Archivbestände

Anhand von zurzeit gegebenen Möglichkeiten können nur Auswertungen über die Anzahl von Archivadokumenten erstellt werden. Diese Auswertungen können innerhalb des gesamten Archivbestands oder selektiert nach Umgebungen und Bereichen ermittelt werden. Eine andere Art von Auswertung ist zurzeit nicht möglich. Die BGW strebt nach Lösungsmöglichkeiten einer Archivsuche von heute noch nicht bekannten Auswertungen. Auslöser können z. B. Anforderungen seitens der Revision sein, um z. B. Zahlungsströme zu analysieren und die entsprechenden Dokumente bereitzustellen.

Ziel der BGW ist es, die Erstellung von Auswertungen in Unabhängigkeit von Deskriptoren und nicht nur selektiert nach Umgebungen und Bereichen zu ermitteln. Eine Auswertung soll z. B. alle Rechnungen der Firma XY aus dem ganzen Archivgut filtern.

Zur Erstellung solcher Auswertungen wird in den nachfolgenden Kapiteln folgender Lösungsweg untersucht und bewertet:

- Volltextsuche

6 Realisierungsmöglichkeiten

In diesem Kapitel werden die Möglichkeiten zur Informationswiedergewinnung aus dem Archivsystem außerhalb von BG-Standard vorgestellt.

- Volltextsuche
- Schlüsselnacherfassung

6.1 *Volltextsuche*

Im Rahmen des Information Retrieval, also der Informationswiedergewinnung, spielt die Volltextsuche, als immer wichtiger werdendes Instrument in Content-Management-Systemen (CMS), eine zentrale Rolle. Auch in elektronischen Archivsystemen werden die Vorteile der Volltextsuche zur Informationswiedergewinnung übernommen und zusätzlich zur Standardsuche angeboten. Die zwei wesentlichen Komponenten einer Volltextsuche sind die Erstellung eines Index und die eigentliche Suche.

6.2 *SER-Volltextsuche*

Der aktuelle ITA-Server von SER bietet neben den Standardschlüsseln auch Volltextschlüssel für eine spätere Suche nach individuellen Dokumenten.

Der Nachfolge-Client Evita ist mit zusätzlichen Diensten für eine Volltextarchivierung und -recherche ausgestattet.

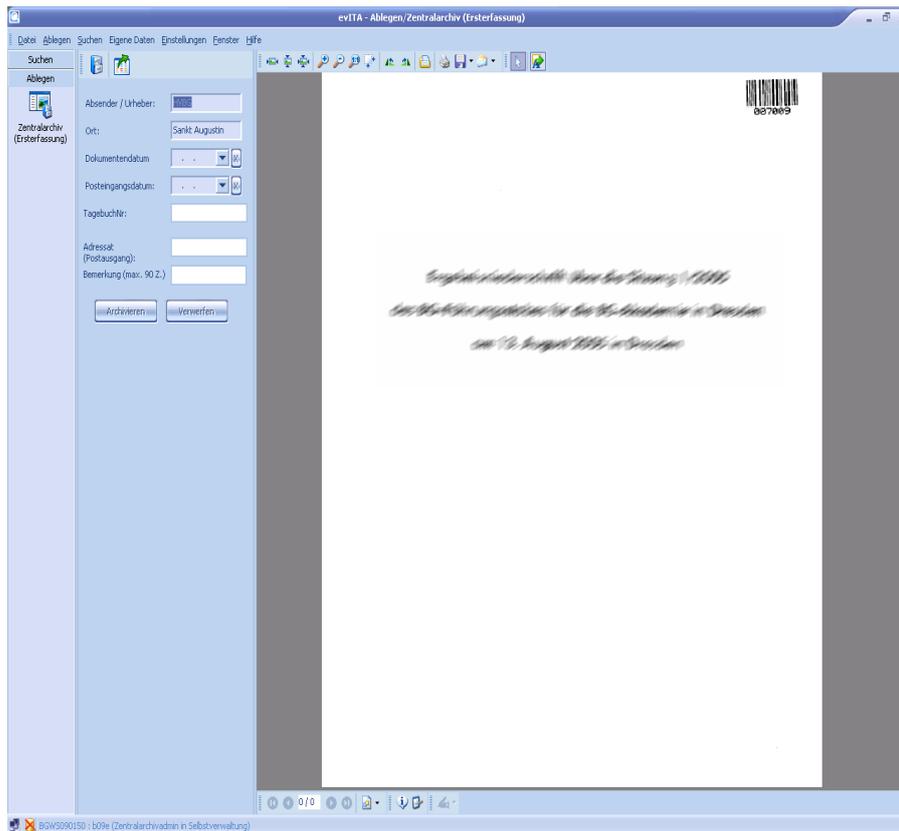


Abbildung 6.1: Volltextsuche mit dem Evita-Client

Die Dokumente werden nicht wie gehabt im Image-Format, sondern in Portable Document Format (PDF/A) im Archivsystem archiviert.

PDF/A ist eine Normreihe der International Organization for Standardization (ISO) für die elektronische Langzeitarchivierung von Dokumenten. Der Vorteil bei der Volltextsuche im PDF/A-Format liegt darin, dass PDF/A voll durchsuchbar ist und das gesuchte Wort markiert angezeigt wird.

Der SER-Archivserver bietet zwei unterschiedliche Varianten zur Volltextsuche: Verity-Volltextsuche und ASSA-Volltextsuche.

Ob eine Verity-Volltextsuche oder ASSA-Volltextsuche verwendet wird, hängt von Einstellungen auf der Datenbankebene im Server ab. Nachfolgend werden die beiden Varianten zur Volltextsuche vorgestellt.

6.2.1 Verity-Volltextsuche

Bei der Volltextindizierung werden alle Worte aus dem Inhalt eines Dokuments in den Index geschrieben. Die Volltextindizes (Schlüssel) stehen nicht im Suchbaum des normalen Index, sondern in einem separaten Volltextindex (Verity) und können dort zur Suche verwendet werden.

Durchsucht man den Index einer Datenbank mit Verity-Unterstützung nach einem bestimmten Wort, erhält man als Suchergebnis alle Dokumente, die dieses Wort enthalten.

Der ITA-Server verwendet zur Volltextindizierung das VDK (Verity Development Kit) der Verity K2-Engine.

Für jeden Index einer Datenbank mit Verity-Unterstützung wird eine sogenannte Verity Collection angelegt. Um auch bei großem Dokumentendurchsatz besser indizieren und suchen zu können, sind Verity Collections wiederum in Partitionen unterteilt. Jede Partition enthält Volltextinformationen für maximal 64.000 Dokumente. Diese Informationen setzen sich aus Wortindizes und Attributtabelle zusammen.

Der ITA Server befüllt die Attributtabelle zusätzlich zu den Feldern, die von der Verity Engine standardmäßig verwendet werden, mit den Internen Primärschlüsseln und dem Dokumentendatum der volltextindizierten Dokumente.⁸

6.2.1.1 Ablauf der Volltextindizierung mit Verity

Der Archivierungsauftrag wird erteilt, und im Anschluss an die Indizierung mit Suchschlüsseln übergibt der ITAV-Prozess des ITA-Servers einen Verweis auf das zu verarbeitende Dokument an ein ITA-spezifisches Verity Gateway.

Mithilfe der Stellent OutsideIn Content Access Filter wird aus den Dokumentendaten ein Textextrakt erzeugt. Diese Filter unterstützen eine große Anzahl von Formaten, jedoch nicht Image-Formate. Daher archiviert der Evita-Client die Dokumente zusätzlich in ein durchsuchbares Format, um eine nachträgliche OCR-Lesung zu vermeiden. Das Textextrakt wird im Unicode-Format an die Verity-Engine übergeben.

⁸ Vgl. SER Solutions(2007a).

Nun wird festgelegt, in welcher Collection das entsprechende Dokument geschrieben werden soll. Ein entsprechender VDK-Funktionsaufruf erzeugt unter Verwendung des Textextrakts den Wortindex und die Attributtabelle. Das Textextrakt wird dann verworfen. Der Wortindex wird ohne die sogenannten Stoppwörter (z. B. der, die, das) erzeugt, da diese bei der Suche zu übermäßig vielen Treffern führen würden.⁹

6.2.2 ASSA-Volltextsuche (assoziative Suche)

Die ASSA-Volltextsuche ist eine Variante der Volltextsuche, bei der Treffer aufgrund statistischer Ähnlichkeiten von Dokumenten mit einem Beispieltext gefunden werden.

Die ASSA-Suche kann in unterschiedlichen Suchmodi betrieben werden. Über den Such-Operator lässt sich zum einen zwischen fehlertoleranter Suche und exakter Suche umschalten. Standard ist die fehlertolerante Suche. Bei der exakten Suche werden nur Wörter berücksichtigt, die genauso geschrieben sind wie im Beispieltext.

Für jeden Index einer Datenbank mit ASSA-Unterstützung wird eine eigene ASSA Knowledge Base angelegt.

Auch hier stehen die Volltextindizes nicht im Suchbaum des normalen Index, sondern in einem separaten Volltextindex (ASSA) und können dort zur Suche verwendet werden.¹⁰

6.2.2.1 Ablauf der Volltextindizierung mit ASSA

Auch im Falle von ASSA erfolgt zunächst eine Textextraktion nach der Indizierung mit Suchschlüsselwörtern. Hierzu werden wiederum die Steller Filter verwendet. Die ASSA-Engine benötigt das Textextrakt im ANSI-Format. ASSA-intern wird daher eine Unicode-ANSI-Konvertierung durchgeführt. Auch bei der ASSA-Indizierung können keine Image-Formate verarbeitet werden.

Jedes Dokument erhält eine Art Fingerabdruck. Die Fingerabdrücke der Dokumente werden im jeweiligen ASSA-Index gespeichert.

⁹ Vgl. SER Solutions(2007a).

¹⁰ Vgl. SER Solutions(2007a).

ASSA ermittelt für jedes Dokument die Häufigkeitsverteilung der Worte. In einem sogenannten Dictionary werden die statistischen Worthäufigkeiten verwaltet. Jede Datenbank mit ASSA-Unterstützung besitzt ein eigenes Dictionary. Das Textextrakt wird zur Ermittlung des Fingerabdrucks und für den Dictionary-Eintrag in einem eigenen ASSA-Zeichensatz konvertiert. Dabei werden Kleinbuchstaben auf Großbuchstaben und Sonderzeichen auf passenden Standardzeichen abgebildet.

Das Textextrakt selbst wird im jeweiligen ASSA-Index gespeichert. Es enthält den Internen Primärschlüssel, das Dokumentendatum der volltextindizierten ITA-Dokumente sowie die ID des Volltext-Deskriptors.¹¹

6.3 Realisierungswege

6.3.1 Die Eigenentwicklung eines Recherche-Tools

Die API (Application Programming Interface) des ITA-Servers bietet eine Vielzahl von Diensten an, die auch von BG-Standard benutzt werden. Diese Dienste beinhalten die Archivierung, Recherche, Bearbeitung von zu archivierenden Dokumenten und vieles mehr. Durch ein Update des ITA-Servers auf eine neue Version wurde zusätzlich die Volltextindizierung der Dokumente erreicht.

Um eine Volltextsuche für bestehenden Archivbestand zu ermöglichen, müssen die Dokumente zusätzlich in einem Format archiviert werden, das vom Stellent OutsideIn Content Access Filter unterstützt wird, wie z. B. eine Textdatei oder ein PDF/A-Format.

Der ITA-Server ermöglicht es, dass Daten unterschiedlichen Formats zu einem Dokument zusammengefasst werden. Für jedes Format wird eine sogenannte Repräsentation angelegt. Mehrere Repräsentationen, also PDF und TIFF, werden dann zu einem Dokument zusammengefasst.

Das Recherche-Tool soll über die API die Dienste des ITA-Servers benutzen, um die Archivdokumente für die Konvertierung bereitzustellen. Dabei soll das Tool jedes Dokument nur einmal zur Konvertierung bereitstellen. Für Recherche nach Archivgut zur Konvertierung soll der Deskriptor „Archivierungsdatum“ dienen.

¹¹ Vgl. SER Solutions(2007a).

Die Konvertierung von Archivdokumenten kann vom OCR-Konverter des EDM-Servers durchgeführt werden. Der OCR-Konverter kann aus TIFF-Bildern bloße Textdateien konvertieren, um eine Volltextsuche zu ermöglichen. Um die Textdateien zum ursprünglichen Dokument als eine Repräsentation zu archivieren, soll die Belegnummer (Externer Primärschlüssel), die durch die Recherche ermittelt wurde, benutzt werden.

ITA-COLD kann zu bestehenden Dokumenten eine weitere Repräsentation hinzufügen. In diesem Fall wäre dies also das OCR-Ergebnis, in Form einer Textdatei, ergänzend zum bestehenden TIFF-Dokument. Auf dieser Basis kann dann ein Volltextindex aufgebaut werden.¹²

Anschließend wird die Volltextindizierung, wie im Fall ASSA oder Verity erklärt, durchgeführt.

6.3.2 Der Rendition-Server

Der Rendition-Server der Firma SER ist eine alternative Lösung für die Konvertierung von Archivdokumenten ins PDF/A-Format.

Der Server wurde entwickelt, um aus unterschiedlichen Quelldateien ein Standard-Langzeit-Format (PDF und PDF/A) zu konvertieren.

Außerdem wurde der Server für Massenkonzertierung und Speicherung von mehreren Repräsentationen in ein Dokument entwickelt.

Ein spezielles Tool des Servers durchsucht den Archivserver und konvertiert TIFF-Bilder in durchsuchbare PDFs und ermöglicht so die Volltextsuche.

Im Anschluss wird das konvertierte Dokument zu dem ursprünglichen Format als eine Repräsentation archiviert. Dieser Vorgang verläuft automatisch und benötigt keine Anweisungen.

Auch hier wird im Anschluss die Volltextindizierung durch den Server durchgeführt.

¹² Vgl. SER Solutions(2009).

6.3.3 ITA-COLD-Schlüsselnacherfassung

Mithilfe einer Schlüsselnacherfassung können bereits archivierten Dokumenten nachträglich weitere Schlüssel zugeordnet werden. Am Dokument bereits vorhandene Schlüssel können jedoch weder geändert noch gelöscht werden. Bei der Schlüsselnacherfassung werden Dokumente mithilfe der folgenden Schlüsseltypen gesucht:

- Teilbestandsname
- Zeitraum
- Retrieval-Schlüssel

Der Retrieval-Schlüssel muss ein Suchschlüssel sein. Mit der Kombination der genannten Schlüssel muss ein Dokument eindeutig identifizierbar sein, ansonsten ist die Schlüsselnacherfassung nicht möglich.

Die Schlüsselnacherfassung in ITA-COLD erfolgt über eine Textdatei, die wie CI-Eingangsdaten behandelt wird. Die Datei beinhaltet Schlüssel zur eindeutigen Identifikation von archivierten Dokumenten. Außerdem beinhaltet die Datei Schlüsselwerte, die den gefundenen Dokumenten neu zugeordnet werden sollen.¹³

6.4 *Bestätigung durch die Hamburger Geschäftsstelle der Firma SER*

In dieser Arbeit wurde die Umsetzung der Lösung nicht angefordert, somit entfällt die darauffolgende Testphase und Bestätigung der Realisierbarkeit.

Dennoch wurde der Lösungsweg „Die Eigenentwicklung eines Recherche-Tools“ durch die Hamburger Geschäftsstelle der Firma SER auf Realisierbarkeit geprüft und dessen Umsetzungsmöglichkeit bestätigt.

¹³ Vgl. SER Solutions(2007b).

7 Bewertung der Realisierungswege

Anhand der folgenden Kriterien sollen die Realisierungswege bewertet werden:

- Wie gut werden die Anforderungen der BGW erfüllt?
- Mit welchen Kosten muss bei der Umsetzung der Lösung gerechnet werden?
- Wie lange wird die Realisierbarkeit dauern?

Das erste Kriterium soll die Abdeckung der Anforderungen abbilden. Zu den Anforderungen gehören gezielter Informationsgewinn durch direkte Archivsuche und Erstellung von bestimmten Auswertungen. Bewertet wird, wie gut die Anforderungen erfüllt werden.

Die Kostenbewertung setzt sich aus verschiedenen Bestandteilen zusammen. Neben den Softwareanschaffungskosten können zusätzliche Kosten durch z. B. Hardwareanforderungen oder Beratungen entstehen. Je kostengünstiger eine Lösung ist, desto besser wird sie bewertet.

Ein weiteres Kriterium ist die Dauer der Realisierbarkeit. Sie ist von Bedeutung, da eine baldige Lösung gewünscht wird.

7.1 Bewertung

Nachfolgend wird für jeden Realisierungsweg erläutert, wie es zur Bewertung kommt. Anschließend zeigt die Tabelle die Bewertung der drei vorgestellten Realisierungswege.

7.1.1 Die Eigenentwicklung eines Recherche-Tools

Die Anforderungen werden durch die Volltextindizierung erfüllt. Die Volltextsuche ermöglicht eine gezielte Suche unabhängig von Deskriptoren. Auch eine Auswertung kann durch die Volltextsuche erfolgreich abgeschlossen werden.

Die Kosten der Softwareentwicklung sind im Vergleich zu den Anschaffungskosten des Rendition-Servers günstiger. Es besteht keine Hardwareanforderung, da der EDM-

Server als Konverter benutzt werden kann. Die Kosten werden anhand des Kostenschätzverfahrens Constructive Cost Model (COCOMO) berechnet.¹⁴

Die Kostenberechnung ergibt sich aus der Schätzung der Produktgröße. Die Produktgröße wiederum ergibt sich aus dem geschätzte Anzahl der Codezeilen. Für die Berechnung wurde die folgende Formel benutzt:

$$MM = 2,4 \times KDSI^{1,05} \quad (1)$$

KDSI: Kilo delivered source instructions (Anzahl ausgelieferte Codezeilen in 1000).

MM: Man month (Personenmonat).

Die Produktgröße wird auf ca. 600 Codezeilen (KDSI) geschätzt.

Für die Personalkosten wurde das monatliche Durchschnittsgehalt eines Softwareentwicklers als Vorlage genommen (3.400 EUR).¹⁵

$$MM = 2,4 \times 0,6^{1,05}$$

$$MM = 1,4$$

$$\text{Geschätzte Kosten} = 1,4 \times 3.400$$

Daraus ergeben sich geschätzte Gesamtkosten von ca. 4.800 €.

Internes Know-how für die Integration einer Software (z. B. BG-Standard) in das Archivsystem ist vorhanden.

Das Archivsystem beinhaltet eine Reihe von Clients, die bereits viele Aufgaben übernehmen und viele Dienste anbieten. Daher besteht intern noch kein Know-how bezüglich der Entwicklung einer Software, die mit dem Archivserver kommunizieren und dessen Dienste nutzen kann. Eine Realisierung ohne Fachwissen der Firma SER kann zeitaufwendig werden.

Aus der oben genannten Berechnung ergibt sich eine geschätzte Entwicklungszeit von etwa 8 Wochen, vorausgesetzt es existiert eine vollständige Beschreibung der API des ITA-Servers.

7.1.2 Der Rendition-Server

Auch durch den Rendition-Server wird die Volltextindizierung möglich. Die Anforderungen werden auch hier durch die Volltextsuche komplett erfüllt. Hier wird zusätzlich

¹⁴ Vgl. Ian Sommerville(2001).

¹⁵ Vgl. Gehaltsvergleich.

durch die Konvertierung in PDF/A das gesuchte Wort in dem Dokument markiert angezeigt. Dies kann in bestimmten Recherchevorgängen von Vorteil sein.

Die Beschaffung und der Einsatz des Servers sind kostenintensiv. Die Kosten liegen bei 19.000 €. ¹⁶ Da aber der Rendition-Server schon bei der BGW für andere Aufgaben erworben wurde, kommen keine Zusatzkosten hinzu. Die Kostenbewertung fällt positiv aus.

Der Server ist eine Lösung der Firma SER und wurde als Erweiterung des Archivsystems entwickelt. Die Einführung und Integration des Servers wird auf 2 Wochen geschätzt und ist daher wenig zeitaufwendig.

Da der Rendition-Server bereits eine komplette Lösung anbietet, ist keine zusätzliche Softwareentwicklung notwendig. Ein weiterer Vorteil besteht darin, dass internes Know-how aufgrund der jetzigen Administration des ITA-Servers vorhanden ist.

7.1.3 ITA-COLD-Schlüsselnacherfassung

Nicht alle Anforderungen können durch Einsatz der ITA-COLD-Schlüsselnacherfassung erfüllt werden. Ein effizientes Ergebnis einer direkten Archivsuche wird durch die Schlüsselerweiterung in allen Bereichen geliefert. Nicht möglich sein wird weiterhin die Auswertung des Archivguts.

ITA-COLD ist ein Bestandteil des Archivsystems; seine Dienste werden seit dessen Einführung genutzt. Dadurch entstehen keine weiteren Kosten.

Aufgrund der Posteingangarchivierung ist bereits internes Know-how für die Benutzung des ITA-COLD-Clients vorhanden. Die Dienste des ITA-COLD-Clients können ab sofort in Anspruch genommen werden.

Bei der BGW ist die Realisierbarkeit der ITA-COLD-Schlüsselnacherfassung nicht möglich, da die Retrieval-Schlüssel in den betroffenen Bereichen nicht vorhanden sind. Eine eindeutige Identifizierung der Dokumente erfolgt mit Externem Primärschlüssel. Der Externe Primärschlüssel ist kein Suchschlüssel. Durch ihn kann keine Schlüsselnacherfassung erfolgen.

¹⁶ Vgl. BGW(2010).

7.1.4 Bewertungstabelle

In der Tabelle werden die Kriterien folgend bewertet:

„-“ bedeutet eine schlechte Erfüllung des Kriterium,

„+“ eine gute Erfüllung des Kriterium.

Kriterium	Rendition-Server	Eigenentwicklung Recherche-Tool	ITA-Schlüssel- nacherfassung
Anf. direkte Archivsuche	+	+	+
Anf. Auswertungen	+	+	-
Kosten	+	+	+
Dauer	+	-	+

Abbildung 7.1: Bewertungstabelle

7.2 Fazit

Abschließend kann gesagt werden, dass der Rendition-Server eindeutig als Lösungsmöglichkeit empfohlen werden kann. Zusätzlich zur Erfüllung der Anforderungen, die die BGW an den Server stellt, bietet der Server auch eine Erweiterung des Archivsystems, um unabhängig von Applikationen Dokumente zu öffnen.

Unter der Voraussetzung, dass die Posteingangs- und Postausgangsdokumente für die Volltextsuche zukünftig nur im PDF/A-Format archiviert werden, existiert ein einheitliches Volltextsuchformat im Archiv. Dadurch kommen die Vorteile des PDF/A-Formats bei der Volltextsuche im ganzen Archivgut zum Tragen.

Der Rendition-Server soll als Lösung für den aktuellen Archivbestand angewendet werden.

Zukünftig sollen die Dokumente mit dem Evita-Client im PDF/A-Format archiviert werden, um eine nachträgliche Konvertierung und damit zusätzlichen Speicherverbrauch zu vermeiden.

Die Einführung des Rendition-Servers erfordert einige Maßnahmen, die sowohl im ITA-Server, als auch im Rendition-Server vorgenommen werden müssen.

Zuerst wird festgelegt welche Dokumente und Bereiche im ITA-Server für eine Volltextsuche eine Indizierung erhalten. Im ITA-Server wird dann für diese Dokumente die

Volltextindizierung freigeschaltet. Dabei soll auch entschieden werden welche Volltextsuchvariante (Verity-Volltextsuche oder ASSA-Volltextsuche) als Einstellung auf der Datenbankebene vorgenommen werden.

Auch im Rendition-Server müssen Parameter bezüglich der Dokumente und deren Bereiche im Archivsystem eingesetzt werden.

Damit wird erreicht, dass die Dokumente in PDF/A konvertiert werden und als eine Repräsentation archiviert werden. Im Anschluss wird die Indizierung durch den ITA-Server durchgeführt.

Abkürzungsverzeichnis

AIT	Advanced Intelligent Tape
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
ANSI	American National Standards Institute
BG	Berufsgenossenschaft
BGW	Berufsgenossenschaft für Gesundheitsdienst und Wohlfahrtspflege
BV	Bezirksverwaltung
CAS	Content Addressed Storage Systems
CD	Compact Disc
CI	Coded Information
CMS	Content-Management-System
COCOMO	Constructive Cost Model
COLD	Computer Output on Laser Disk
DB	Datenbank
DR	Data Retention
DVD	Digital Versatile Disc
EBCDIC	Extended Binary Coded Decimals Interchange Code
EDM	Elektronischer Dokumenten Manager
etc.	Et cetera
FTP	File Transfer Protocol
HSM	Hardware-Security-Module
ID	Identifikator
IDX	Index
KDSI	Kilo delivered source instructions
NCI	Non-Coded Information
NetApp	Network Appliance
MM	Man month
MOD	Magneto Optical Disc
OCR	Optical Character Recognition

PDD	Professional Disc for Data
PDF	Portable Document Format
PEPA	Posteingang Postausgang
PROD	Produktion
RAID	Redundant array of independent disks
RDY	Ready
SQL	Structured Query Language
TIFF	Tagged Image File Format
TSM	Tivoli Storage Manager
UB	Unternehmerbetreuung
UDO	Ultra Density Optical
usw.	und so weiter
VDK	Verity Development Kit
WORM	Write once read many
XML	Extensible Markup Language
z. B.	zum Beispiel

Literaturverzeichnis

- [BGW(2010)] Auskunft durch Herr Ulrich Eilers.
- [BGW(2009)] Berufsgenossenschaft für Gesundheitsdienst und Wohlfahrtspflege.
Schwerpunktprogramm der BGW 2009. 2009
- [BGW(2009a)] Heismann, O.
Schnittstelle BG-Standard - Archiv. 2009
- [Bisges(2009)] Bisges, M.
Urheberrechtliche Aspekte des elektronischen Dokumentenmanagements.
Baden-Baden : Nomos, 2009
- [Blankenberg(1998)] Blankenberg, R.
Implementierung von Dokumentenarchivierungssystemen : eine betriebswirtschaftliche Untersuchung.
Lohmar : Eul, 1998
- [Borghoff et al. (2003)] Borghoff, U.; Scheffczyk, J.; Rödiger, P.; Schmitz, L.
Langzeitarchivierung: Methoden zur Erhaltung digitaler Dokumente.
Heidelberg : dpunkt.verlag GmbH, 2003
- [Gehaltsvergleich] <http://www.gehaltsvergleich.com/berufe-i-2.html>.
Informatiker / Informatikerin (staatl.gepr.).
- [Gugerli(2009)] Gugerli, D.
Suchmaschinen : die Welt als Datenbank.
Frankfurt am Main : Suhrkamp, 2009
- [Gulbins et al.(1999)] Gulbins, J. ; Seyfried, M. ; Strack-Zimmermann, H.
Dokumenten-Management : vom Imaging zum Business-Dokument.
Berlin : Springer, 1999
- [Göker/Davies(2009)] Göker, A. ; Davies, J.
Information retrieval: searching in the 21st century.
Chichester : Wiley, 2009
- [Handelskammer Hamburg] Handelskammer Hamburg.
ABC der Aufbewahrungsfristen von Geschäftsunterlagen in Steuerrecht und Handelsrecht.
http://www.hk24.de/produktmarken/recht_und_fair_play/steuerrecht/abgabenordnung/aufbewahrungsfristen.jsp
- [Ian Sommerville(2001)] Sommerville, I.
Software Engineering.
München : Pearson Studium, 2001

- [**Kampffmeyer/Merkel(1999)**] Kampffmeyer, U.; Merkel, B.
Dokumenten Management: Grundlagen und Zukunft.
Hamburg : Project Consult GmbH, 1999
- [**Meiss(1997)**] Meiss, B.
Information Retrieval und Dokumentenmanagement im Multimedia-Zeitalter.
Frankfurt am Main : Deutsche Gesellschaft für Dokumentation e.V., 1997
- [**SER Solutions(2007a)**] SER Solutions Deutschland GmbH.
Anwenderdokumentation ITA-Server. 2007
- [**SER Solutions(2007b)**] SER Solutions Deutschland GmbH.
Anwenderdokumentation ITA-COLD. 2007
- [**SER Solutions(2009)**] Christian Steinhauer.
Mündliche Auskunft durch Herr Christian Steinhauer. 2009
- [**SER Solutions(2009b)**] SER Solutions Deutschland GmbH.
DOXiS4-Archiv.
http://www.ser.de/ww/de/pub/ser_solutions/produkte/doxis_archiv.cfm
- [**Schwalm/Ullrich(2008)**] Schwalm, S.; Ullrich, R.
Lexikon Dokumentenmanagement und Archivierung.
Berlin : Infora, 2008
- [**Stock(2007)**] Stock, W.
Information retrieval : Informationen suchen und finden
München : Oldenbourg, 2007

Versicherung über Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Arbeit im Sinne der Prüfungsordnung nach §24(5) ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Hilfsmittel benutzt habe.

Hamburg, 30 März 2010

Ort, Datum

Unterschrift