

Herausforderungen bei der Übernahme von Unterlagen aus Fileablagen - **Zum Einsatz des PreIngestToolsets (PIT) im Bundesarchiv**

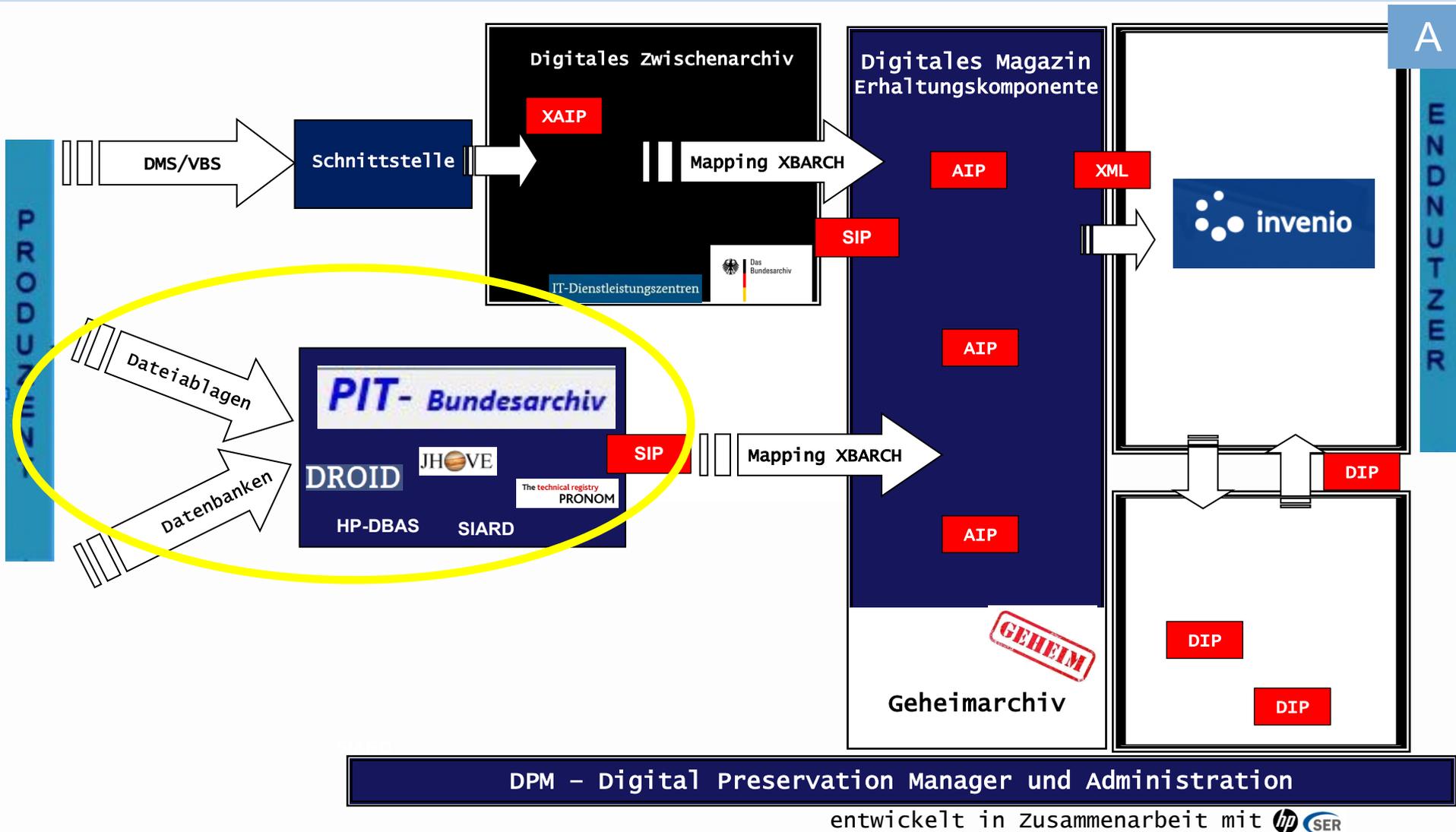
20. Jahrestagung des Arbeitskreises
„Archivierung von Unterlagen aus digitalen Systemen“ (AUdS)
Heike Simon

01. März 2016

Gliederung:

- A Einsatz des PIT
- B PIT-Arbeitsschritte
- C Herausforderungen / offene Fragen

Zum Einsatz des PreIngest Toolsets (PIT) im Bundesarchiv



- Die Anwendung dient der protokollierten...
 - Ergänzung von Metadaten
 - Strukturierung (Sortieren, reguläre Ausdrücke, Bildung IEs/Ordner)
 - Formaterkennung und -validierung
 - ‚Kassation‘

Einsatzmöglichkeiten (“Archivalientypen”)

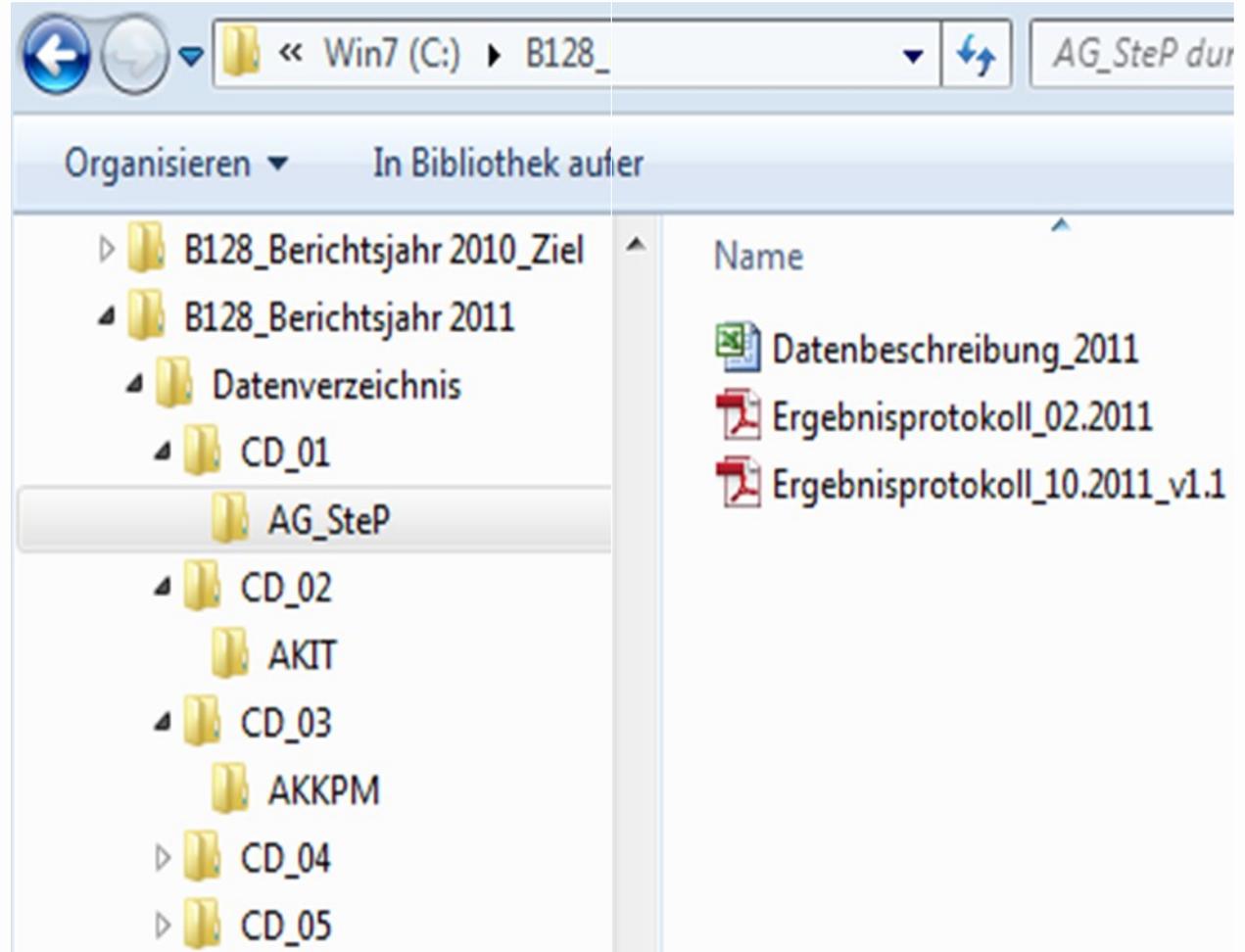
A



Abgabe der Behörde XY

A

- CD-ROM, Datenstick, Laufwerk, Diskette....
- Jeweils spezifische Datenstruktur und Dateiformate

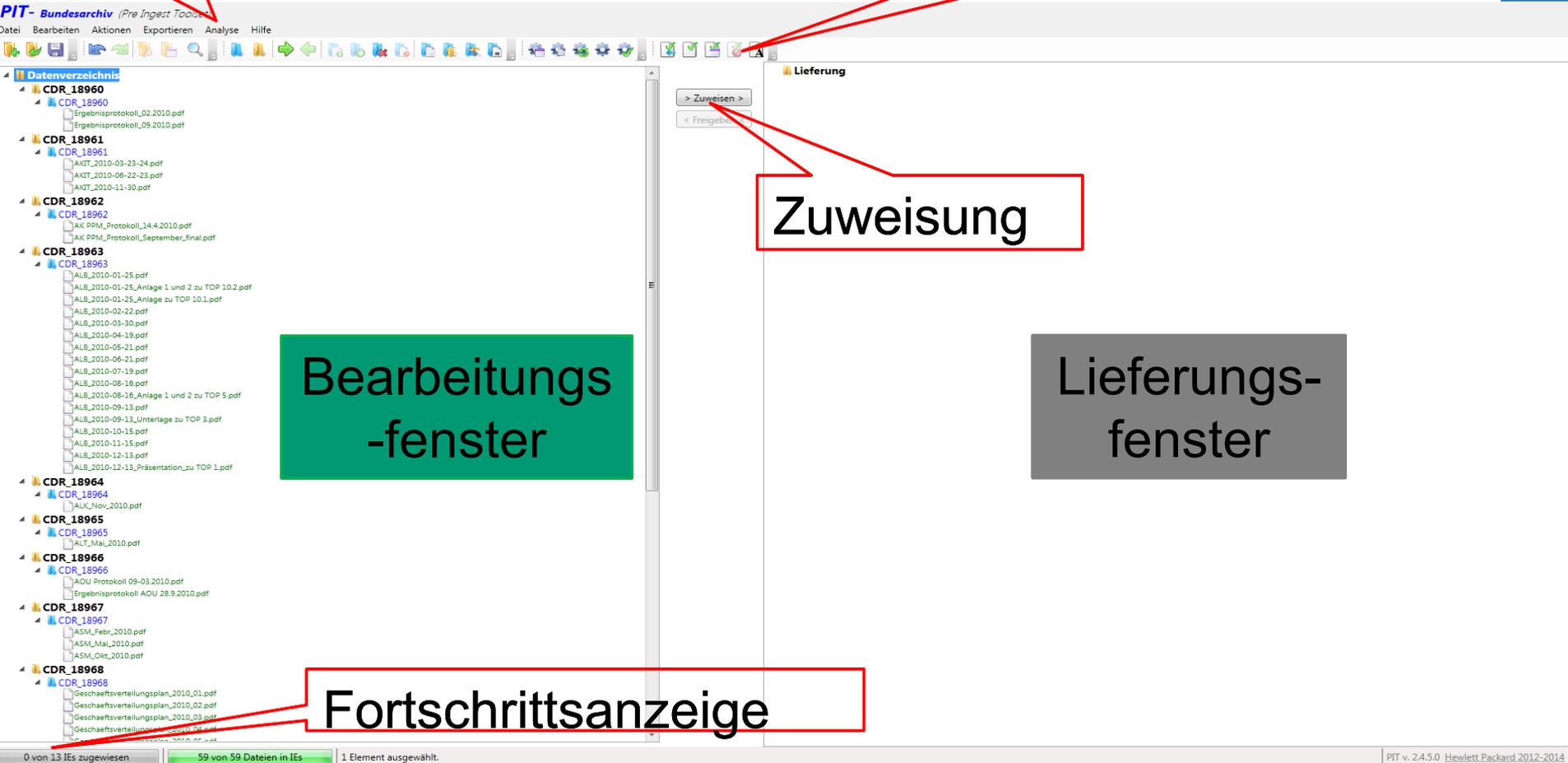


Arbeitsoberfläche

Menü-Befehle

Icon-Befehle

B



0 von 13 IEs zugewiesen

59 von 59 Dateien in IEs

1 Element ausgewählt.

PIT v. 2.4.5.0 Hewlett Packard 2012-2014

Bearbeitungs-
fenster

Zuweisung

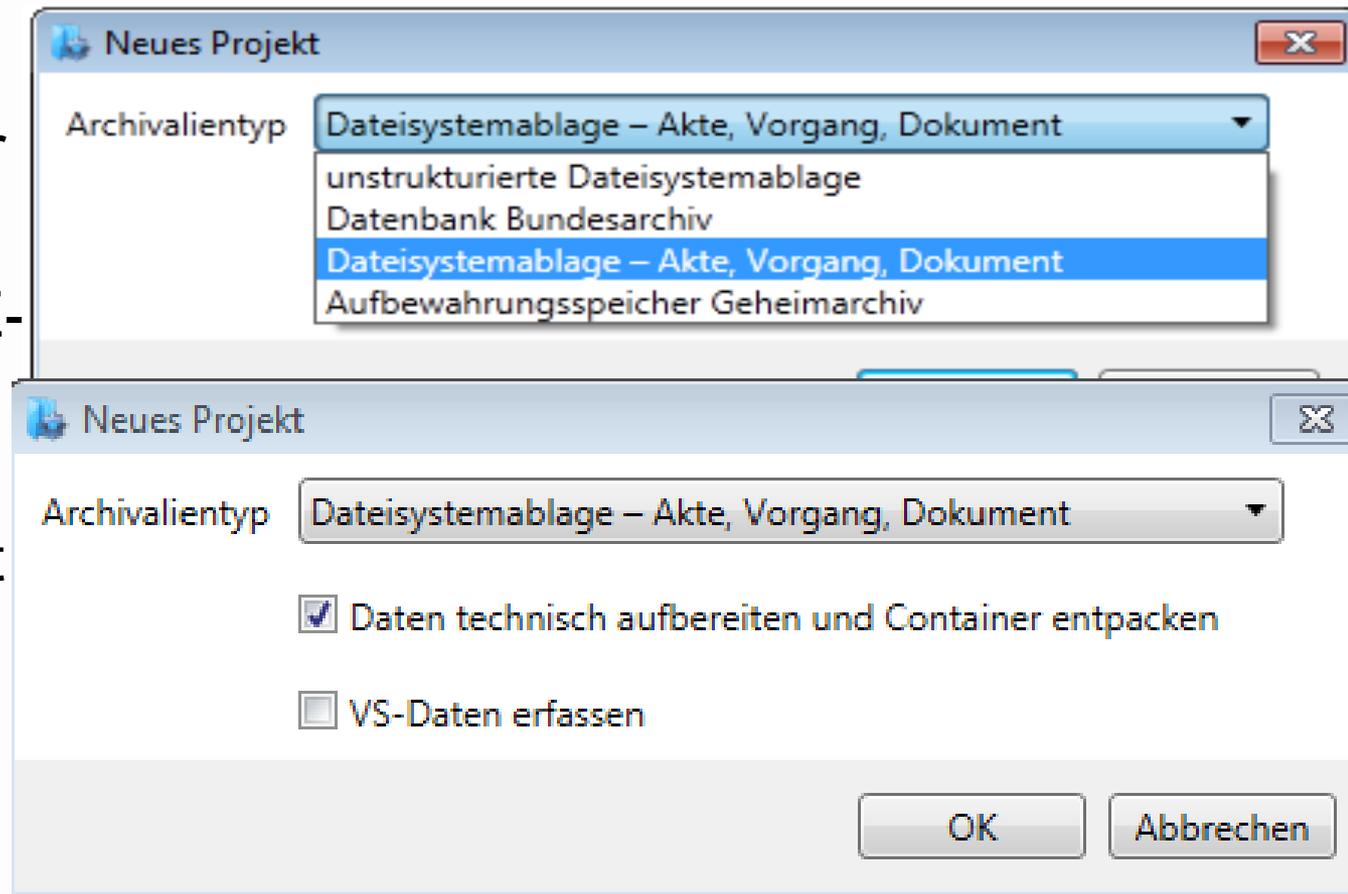
Lieferungs-
fenster

Fortschrittsanzeige

Auswahl: Akte, Vorgang, Dokument

B

- Archivalientyp erzeugt dreigliedrige Struktur
- Entpacken der Container (Zip, E-Mail) in einzelne Dateien
- Jede Datei erhält zusätzlich einen eindeutigen technischen Namen



Arbeitsschritte

B

Vergabe

Metadaten



Analyse

Dateiformate



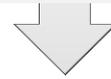
„Kassation“

Dokumente



Strukturierung

Verzeichnisse

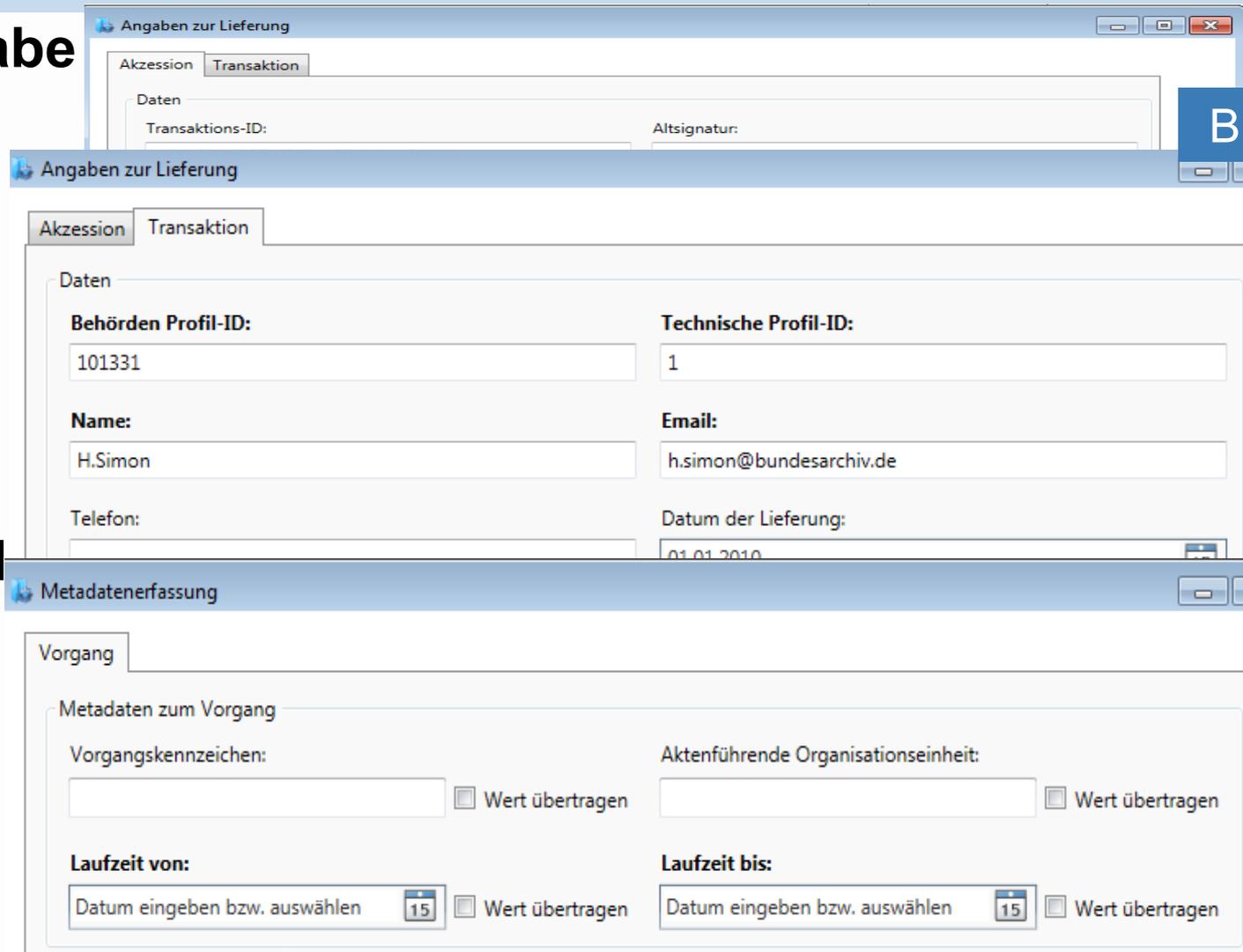


Zuweisung

Vorgänge

Metadatenvergabe

- Angaben zur abgebenden Stelle
- Technische Angaben aus Behördenprofil
- Angaben zu Laufzeiten

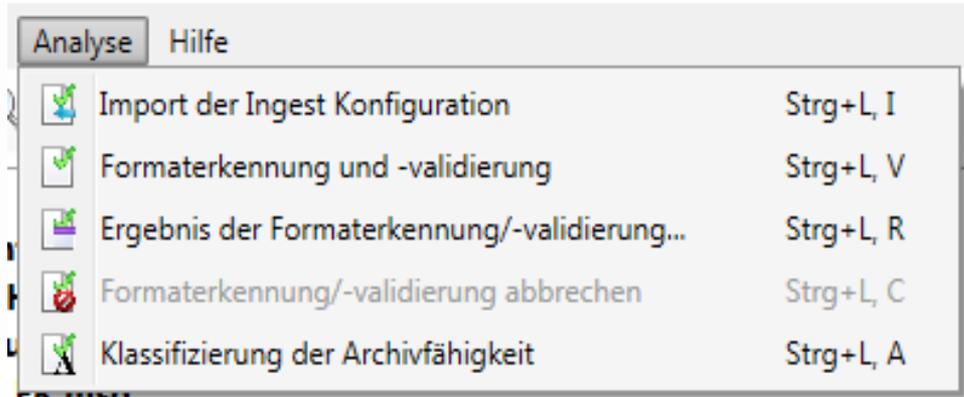


The screenshot displays three overlapping windows from the PreIngest Toolset (PIT) interface:

- Angaben zur Lieferung (top window):** Shows tabs for 'Akzession' and 'Transaktion'. The 'Daten' section includes fields for 'Transaktions-ID:' and 'Altsignatur:'.
- Angaben zur Lieferung (middle window):** Shows the 'Akzession' tab. The 'Daten' section includes fields for:
 - Behörden Profil-ID:** 101331
 - Technische Profil-ID:** 1
 - Name:** H.Simon
 - Email:** h.simon@bundesarchiv.de
 - Telefon:** (empty)
 - Datum der Lieferung:** 01.01.2010
- Metadatenerfassung (bottom window):** Shows the 'Vorgang' tab. The 'Metadaten zum Vorgang' section includes:
 - Vorgangskennzeichen:** (empty) with a checkbox for 'Wert übertragen'.
 - Aktenführende Organisationseinheit:** (empty) with a checkbox for 'Wert übertragen'.
 - Laufzeit von:** Datum eingeben bzw. auswählen [15] with a checkbox for 'Wert übertragen'.
 - Laufzeit bis:** Datum eingeben bzw. auswählen [15] with a checkbox for 'Wert übertragen'.

Analyse der Dateiformate I

- Zeitintensiv je nach Umfang



B

- Reiter „Tech. Eigenschaften“: u.a. Information über identische Dateien (Dubletten)

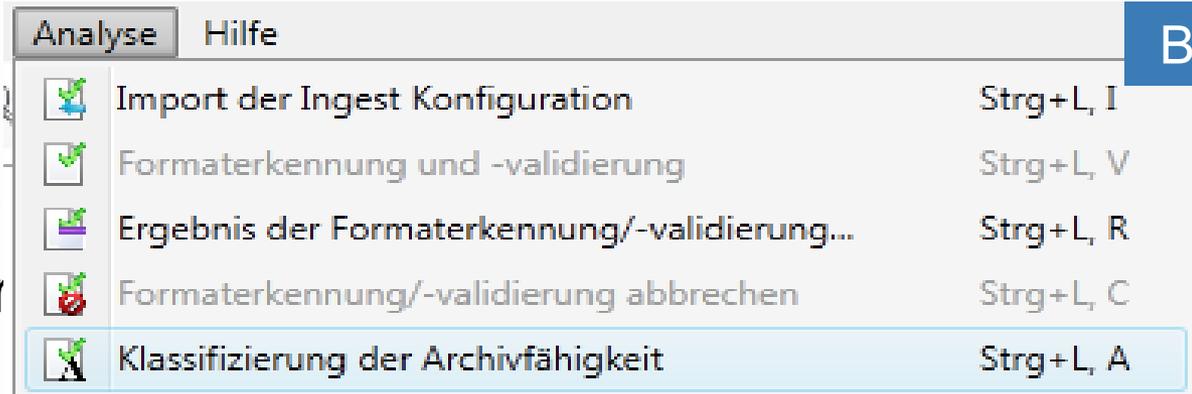
Ergebnis der Formaterkennung/-validierung

Technische Eigenschaften Pronom / Droid JHove

Name	Pfad	SHA1	Anzahl der Dubletten	Status
E_5.5 Änder.vorschlag 1	C:\gmk_2009_Ziel_Test\	9e1d1a1fbdbb2cc005ba9c82eb8f8f7	1	✓
E_GMK_10.1_dioxPCB.d	C:\gmk_2009_Ziel_Test\	ee375d546613826784a85c6a59ff088	4	✓
E_GMK_10.2 A_Lärm ur	C:\gmk_2009_Ziel_Test\	998a1ce1be8c00b0571850ab4ef6a99	1	✓

Analyse der Dateiformate II

- Techn. Analyse Dateiformate
- Reiter „Pronom/Droid“
Abgleich vordefinierter Archivfähigkeit



B

Ergebnis der Formaterkennung/-validierung

Technische Eigenschaften Pronom / Droid JHove

Name	formatName	puid	formatVersion	FormatRegistryName	mime	Archivfähigkeit
Mail CSL 20091029 Name.doc	Microsoft Word Document	fmt/40	97-2003	PRONOM	application/msword	Zulässiges Dateiformat (mögliche Konvertierung im Ingest)
Packungsgröße.doc	Microsoft Word Document	fmt/40	97-2003	PRONOM	application/msword	Zulässiges Dateiformat (mögliche Konvertierung im Ingest)
Panvax press release.pdf	Acrobat PDF 1.4 - Portable Do	fmt/18	1.4	PRONOM	application/pdf	Zulässiges Dateiformat (mögliche Konvertierung im Ingest)
Panvax Produktinfo.pdf	Acrobat PDF 1.4 - Portable Do	fmt/18	1.4	PRONOM	application/pdf	Zulässiges Dateiformat (mögliche Konvertierung im Ingest)
Panvax CSL.doc	Microsoft Word Document	fmt/40	97-2003	PRONOM	application/msword	Zulässiges Dateiformat (mögliche Konvertierung im Ingest)
...

Analyse der Dateiformate III

B

- Reiter „JHove“ Anzeige Validierungsergebnisse

Ergebnis der Formaterkennung/-validierung

Technische Eigenschaften Pronom / Droid **JHove**

Name	format			
Gremienstruktur DOL.ppt	Microsoft Office Powerpoint			
20090630 Bertelsmannstiftung	version	valid	state	mime
FMK und Deutschland-Online.p	97-2003	True	valid	application/vnd.ms-powerpoint
6. 07-10-17 eGov-StS Personen	97-2003	True	valid	application/msword
E Vermerk.doc	97-2003	True	valid	application/vnd.ms-powerpoint
	97-2003	True	valid	application/vnd.ms-powerpoint
	97-2003	True	valid	application/msword

„Kassation I“

- Ausschluss von Dateien, IEs und Verzeichnissen möglich
- Dubletten in ihrem Umfeld auswählbar
- Sonderfunktion: Datei ersetzen

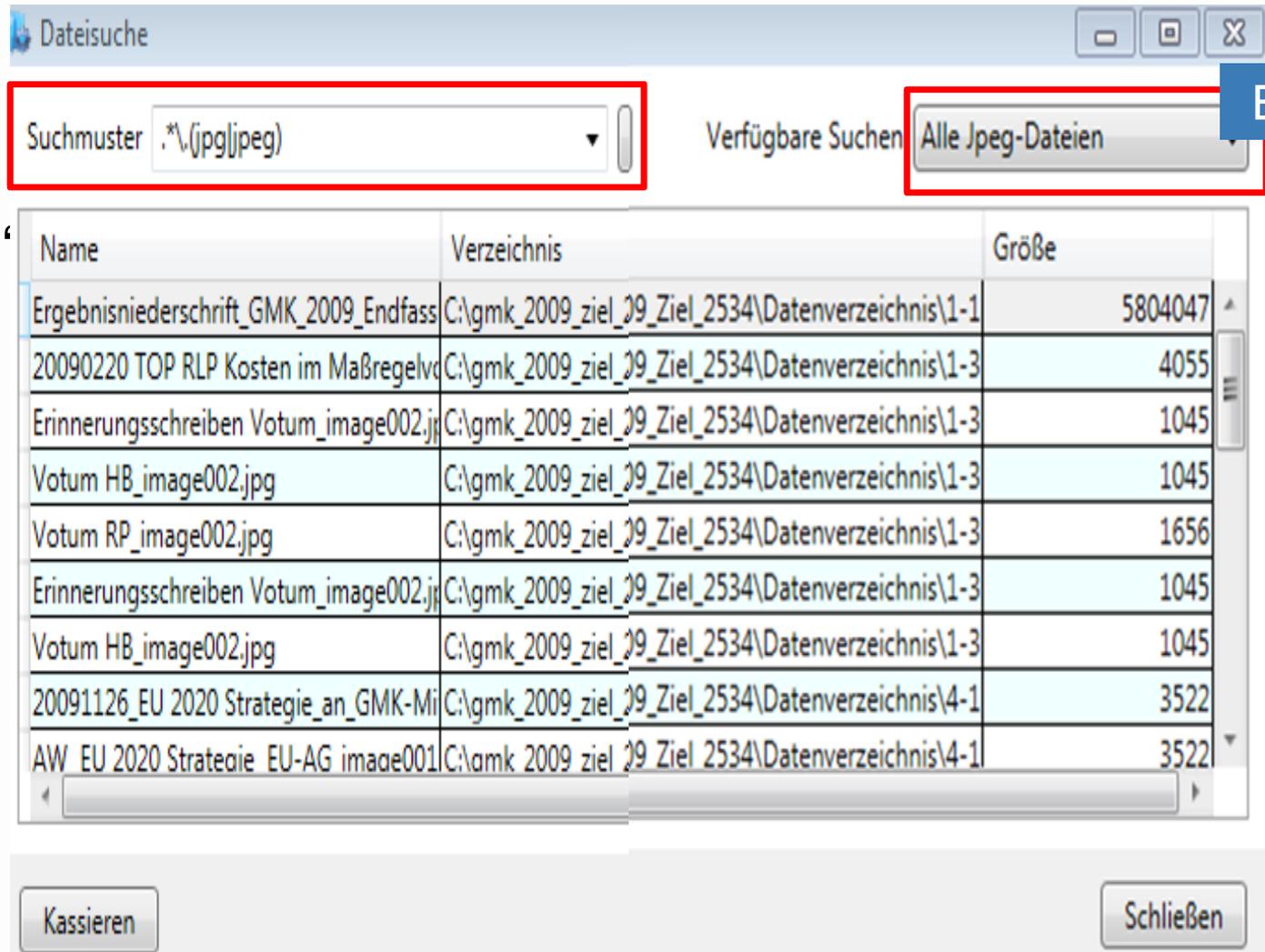
B

Ergebnis der Formaterkennung/-validierung

Technische Eigenschaften Pronom / Droid JHove

Name	Pfad	Anzahl Bytes	Änderungsdatum	SHA1	Anzahl der Dubletten	Status
20090218 HB TOP BÄO ▾	C:\gmk_2009_ziel_2534 ▾	31744	18.02.2009 15:49:57	7e40c3a95f10de2b0db4f32bf7e7d18	3	✓
1401_22AOLG_Verschie ▾	C:\gmk_2009_ziel_2534 ▾	27818	27.11.2008 14:52:10	37b5c4671befdfcf5c63f4699cc946d8	3	✓
By_Änd Länderabgestir ▾	C:\gmk_2009_ziel_2534 ▾	890371	17.06.2009 12:01:21	346021a00c80252ab230ee071adfb51	3	✓
E TOP 9.4 Beschl_Begr ▾	C:\gmk_2009_ziel_2534\gmk2009_Ziel_2534\Datenverzeichnis\Pandemie\Sonderkonferenzen_Pandemie\Sondersitzungen AOLG\Sondersitzung 10.06.09 TLVB\By_Änd Länderabg					
1401_22AOLG_Verschie ▾	C:\gmk_2009_ziel_2534\gmk2009_Ziel_2534\Datenverzeichnis\Pandemie\Sonderkonferenzen_Pandemie\Sondersitzungen AOLG\Sondersitzung 10.06.09 TLVB\Länderabgestimmt					
E TOP 9.3 Beschl_Begr ▾	C:\gmk_2009_ziel_2534 ▾	27648	03.03.2009 08:58:36	279d0327ad17f183cde9ae5583adff1	3	✓

„Kassation II“



Suchmuster: *.\\.(jpg|jpeg)

Verfügbare Suchen: Alle Jpeg-Dateien

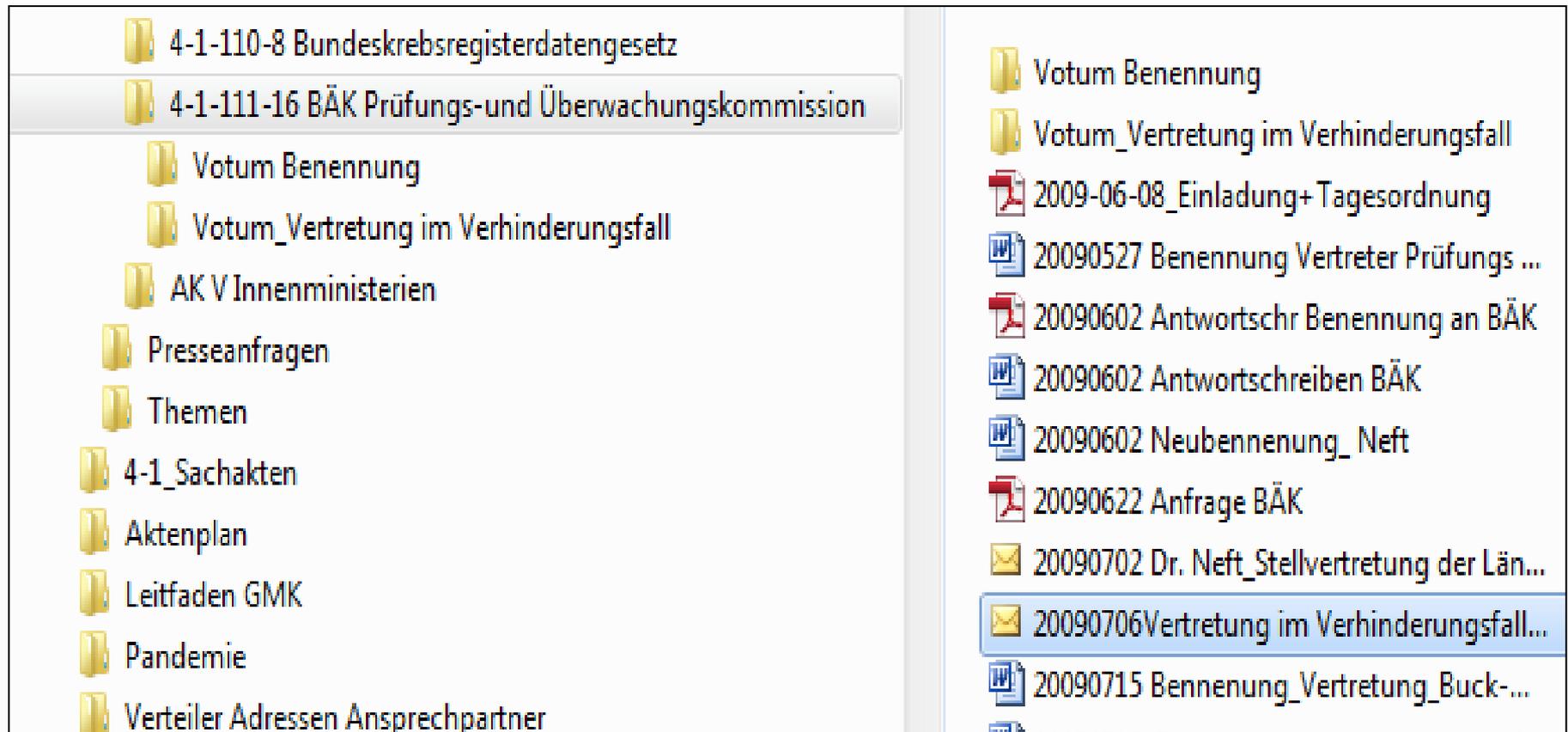
Name	Verzeichnis	Größe
Ergebnisniederschrift_GMK_2009_Endfass	C:\gmk_2009_ziel_09_Ziel_2534\Datenverzeichnis\1-1	5804047
20090220 TOP RLP Kosten im Maßregelver...	C:\gmk_2009_ziel_09_Ziel_2534\Datenverzeichnis\1-3	4055
Erinnerungsschreiben Votum_image002.jpg	C:\gmk_2009_ziel_09_Ziel_2534\Datenverzeichnis\1-3	1045
Votum HB_image002.jpg	C:\gmk_2009_ziel_09_Ziel_2534\Datenverzeichnis\1-3	1045
Votum RP_image002.jpg	C:\gmk_2009_ziel_09_Ziel_2534\Datenverzeichnis\1-3	1656
Erinnerungsschreiben Votum_image002.jpg	C:\gmk_2009_ziel_09_Ziel_2534\Datenverzeichnis\1-3	1045
Votum HB_image002.jpg	C:\gmk_2009_ziel_09_Ziel_2534\Datenverzeichnis\1-3	1045
20091126_EU 2020 Strategie_an_GMK-Mi...	C:\gmk_2009_ziel_09_Ziel_2534\Datenverzeichnis\4-1	3522
AW EU 2020 Strategie EU-AG image001	C:\gmk_2009_ziel_09_Ziel_2534\Datenverzeichnis\4-1	3522

Kassieren Schließen

- In „Dateisuche“ (z.B. Dateien mit definierten Endungen)

Strukturen am Bsp. E-Mail:

I Abgabe Dateiablage – E-Mail-Container



- 4-1-110-8 Bundeskrebsregisterdatengesetz
- 4-1-111-16 BÄK Prüfungs- und Überwachungskommission
 - Votum Benennung
 - Votum_Vertretung im Verhinderungsfall
- AK V Innenministerien
- Presseanfragen
- Themen
- 4-1_Sachakten
- Aktenplan
- Leitfaden GMK
- Pandemie
- Verteiler Adressen Ansprechpartner

- Votum Benennung
- Votum_Vertretung im Verhinderungsfall
- 2009-06-08_Einladung+ Tagesordnung
- 20090527 Benennung Vertreter Prüfungs ...
- 20090602 Antwortschr Benennung an BÄK
- 20090602 Antwortschreiben BÄK
- 20090602 Neubenennung_ Neft
- 20090622 Anfrage BÄK
- 20090702 Dr. Neft_Stellvertretung der Län...
- 20090706Vertretung im Verhinderungsfall...
- 20090715 Bennennung_Vertretung_Buck-...

Strukturen am Bsp. E-Mail:

II PIT-Struktur –

Strukturelement E-Mail zur Bearbeitung

AG's u. ä

▶ 2-3 Arbeitsgruppen

▶ 4-1-101-3 DSO

▶ 4-1-102-2 Blutprodukte HIV

▶ 4-1-104-8 Ändervertreter_NKLM

▶ 4-1-110-8 Bundeskrebsregisterdatengesetz

▶ 4-1-111-16 BÄK Prüfungs- und Überwachungskommission

▶ 20090702 Dr. Neft_Stellvertretung der Länder in Prüfungs- und Überwachungskommission.msg

▶ 20090706 Vertretung im Verhinderungsfall in der Prüfungs- und Überwachungskommission.msg

▶ eilt_Benennung_Vertreter_in_Prüfungs_und_Überwachungskommission.msg

Strukturen am Bsp. E-Mail:

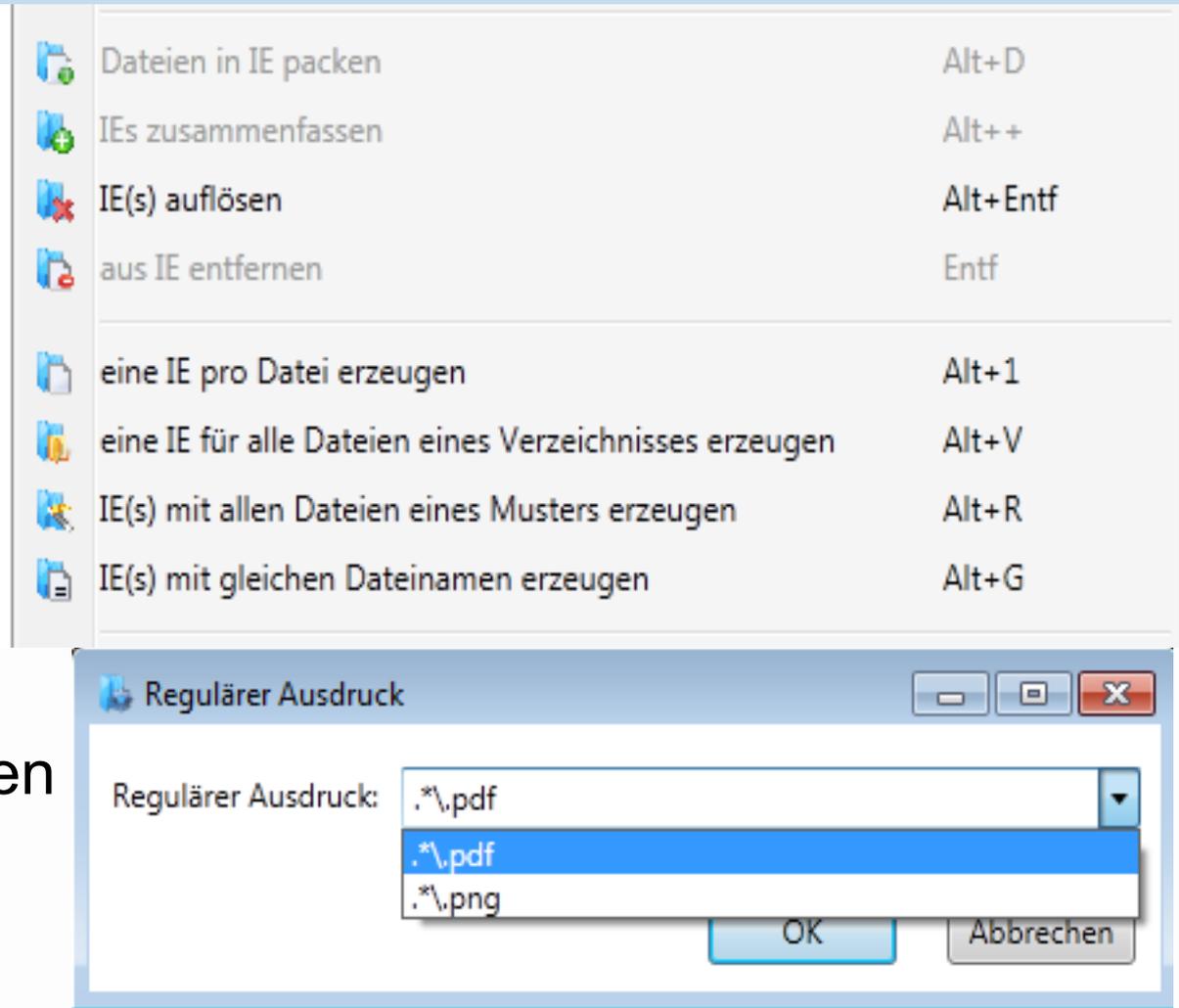
III PIT-Struktur –

Bestandteile des Strukturelementes E-Mail zur Bearbeitung

- ▲  **20090706Vertretung im Verhinderungsfall in der Prüfungs- und Überwachungskommission.msg**
 - ▲  20090706Vertretung im Verhinderungsfall in der Prüfungs- und Überwachungskommission.msg
 -  20090706Vertretung im Verhinderungsfall in der Prüfungs- und Überwachungskommission.htm
 -  20090706Vertretung im Verhinderungsfall in der Prüfungs- und Überwachungskommission.txt
 -  20090706Vertretung im Verhinderungsfall in der Prüfungs- und Überwachungskommission_20090622 Anfrage BÄK.pdf
 -  20090706Vertretung im Verhinderungsfall in der Prüfungs- und Überwachungskommission_header.txt

Strukturierung

- IEs zusammenfassen, auflösen, umbenennen
- IEs nach Regeln formieren (Format u.ä.)
- Dokumentanzeige über ‚Öffnen‘ (je nach installierten Programmen)



The image shows a Windows context menu for Internet Explorer Favorites and a 'Regulärer Ausdruck' (Regular Expression) dialog box.

Context Menu:

Icon	Text	Shortcut
	Dateien in IE packen	Alt+D
	IEs zusammenfassen	Alt++
	IE(s) auflösen	Alt+Entf
	aus IE entfernen	Entf
	eine IE pro Datei erzeugen	Alt+1
	eine IE für alle Dateien eines Verzeichnisses erzeugen	Alt+V
	IE(s) mit allen Dateien eines Musters erzeugen	Alt+R
	IE(s) mit gleichen Dateinamen erzeugen	Alt+G

Regulärer Ausdruck Dialog:

Regulärer Ausdruck:

Dropdown list:

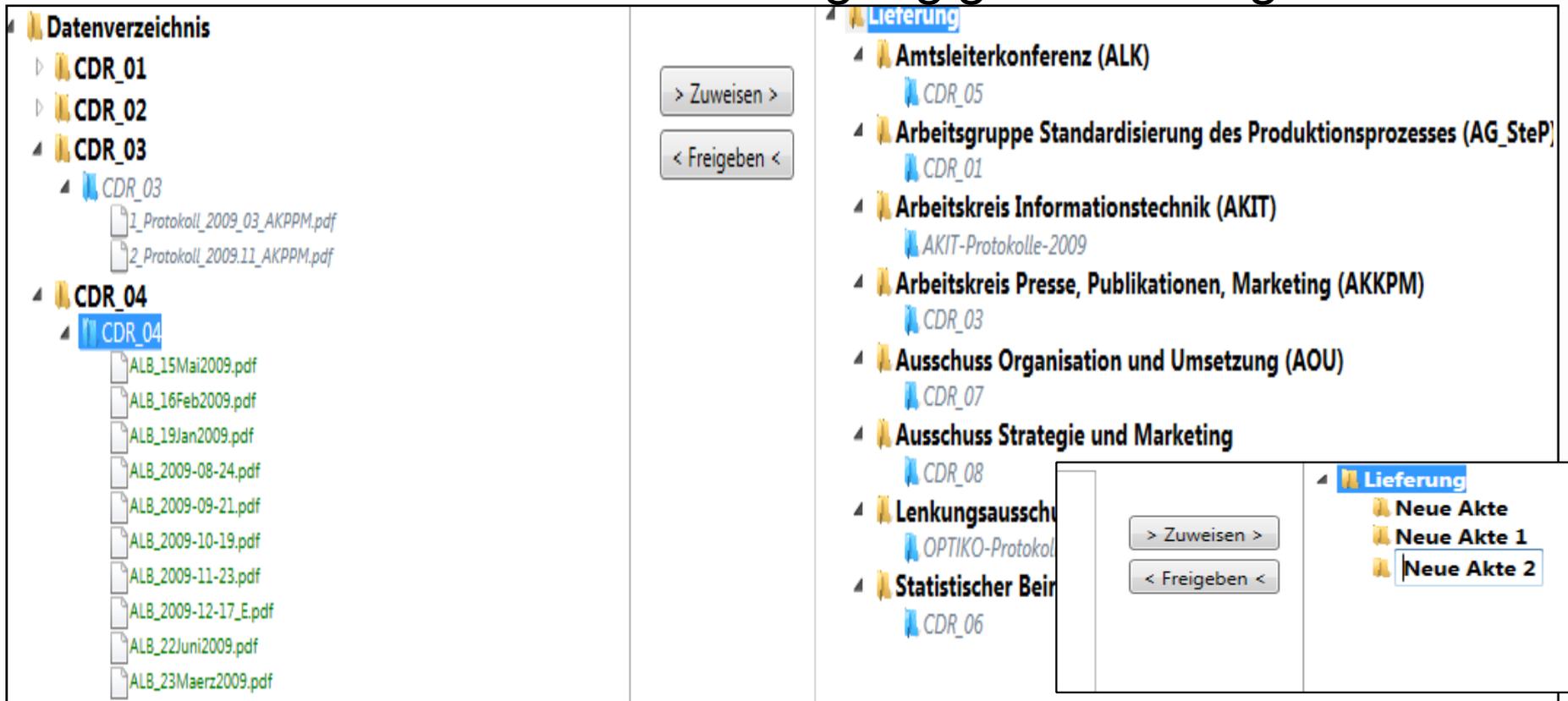
- .*\pdf (selected)
- .*\png

Buttons: OK, Abbrechen

Zuweisung

- IE wird Vorgang
- Einzel oder als Paket
- Vorgänge verschiebbar
- Individuelle Aktentitel möglich
- Rückgängig durch Freigabe

B



The screenshot displays the PreIngest Toolset (PIT) interface. On the left, a tree view shows a 'Datenverzeichnis' folder containing subfolders CDR_01 through CDR_04. CDR_04 is selected, showing a list of PDF files. In the center, there are two buttons: '> Zuweisen >' and '< Freigeben <'. On the right, another tree view shows a 'Lieferung' folder containing several subfolders like 'Amtsleiterkonferenz (ALK)', 'Arbeitsgruppe Standardisierung des Produktionsprozesses (AG_SteP)', etc. A smaller inset window in the bottom right shows a 'Lieferung' folder with three items: 'Neue Akte', 'Neue Akte 1', and 'Neue Akte 2'. The 'Neue Akte 1' and 'Neue Akte 2' items are highlighted with a blue selection box.

Erstellung der Lieferung

möglich, wenn:

- alle Dateien IEs zugeordnet
- alle IEs zugewiesen
- alle Metadatenpflichtfelder befüllt

→ **PIT-Abschluss:
SIPs**

→ *Übergabe an
Ingest*

Pit - Lieferung erstellen



Erstellen der Lieferung erfolgreich.

OK

Pit - Bearbeitung abschließen



Ready-Datei wurde erzeugt.

OK

 B128_2010 Ziel_2532	18.11.2015 13:34	Dateiordner	
 cc329f35-1f4d-4667-b964-df7b59739ea8.r...	18.11.2015 14:17	RDY-Datei	0 KB
 cc329f35-1f4d-4667-b964-df7b59739ea8	18.11.2015 14:17	XML-Datei	2 KB

Inhaltliche Herausforderungen:

- Individualität der Fileablagenstrukturen:
registratorlos / ohne Aktenplan / mit tiefer
Untergliederung / ohne jedwede Gliederung
- Bearbeitungsstand der Dokumente oft nicht zu
ermitteln (Vorgänge?)
- Arbeitsaufwände hinsichtlich Bewertung und
Umstrukturierung beachten
- Archivische Strukturierung als Eingriff in die
Authentizität?

Technische Herausforderungen:

- ‚Museale‘ Datenträgerformen
- Vielfalt von Dateiformaten und Dateifunktionalitäten
- ‚Grenzen‘ der Formaterkennung und -validierung
- Weiterer Umgang mit problematischen Ergebnissen der Formatanalyse (insbes. Ingest)
- Aufwandsberücksichtigung
- Archivwürdigkeit: Archivfachlich contra Technik?

Vielen Dank für
Ihre Aufmerksamkeit !

KONTAKT:
h.simon@bundesarchiv.de