



Catalogue Enrichment Schwerpunkt „Monografien“

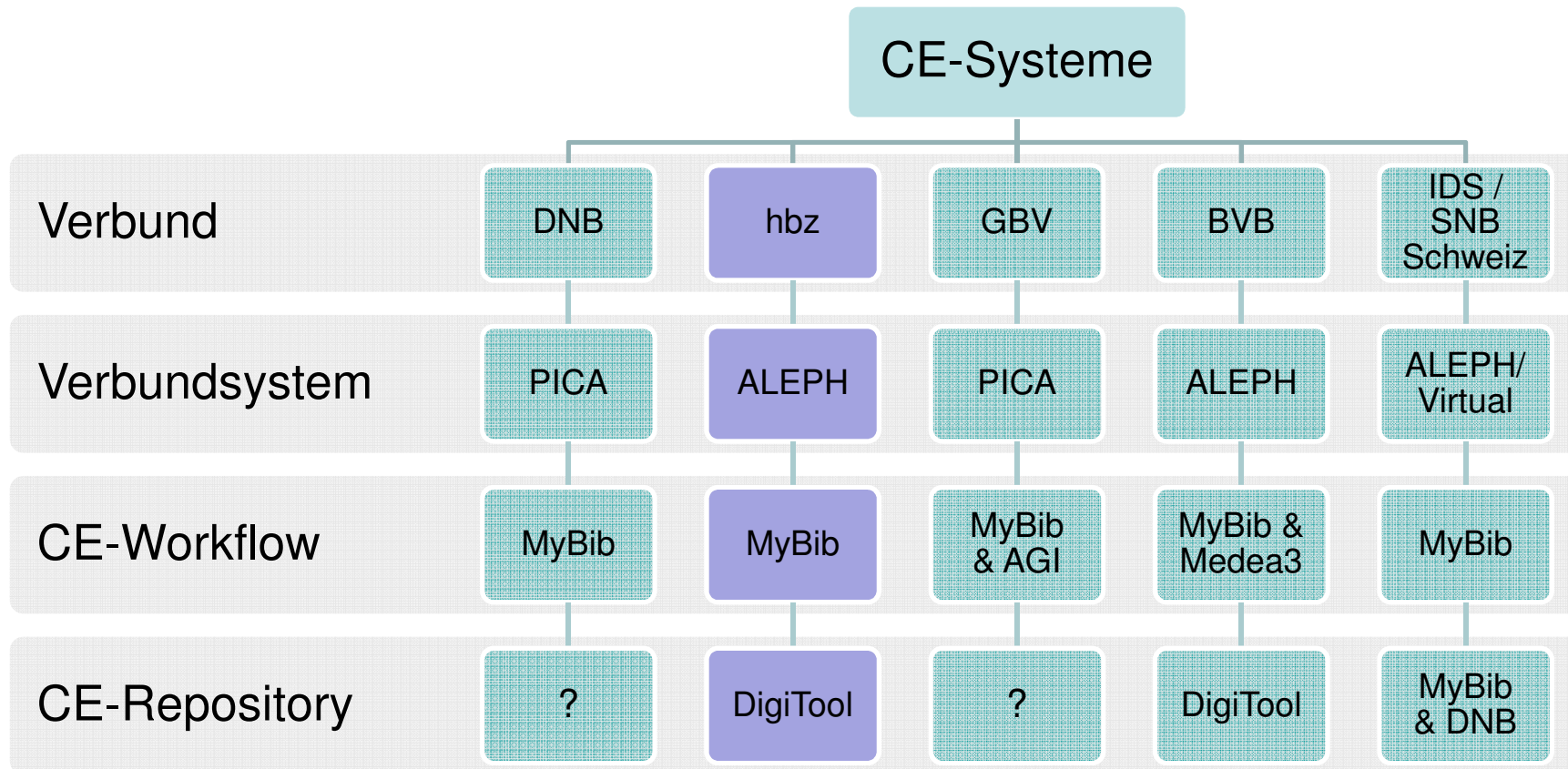
Punktuelle Vorstellung der hbz-Lösung
Michael Nelißen, Rolf Rasche

-
- Historie & Zahlen
 - Systemvielfalt & hbz-System
 - Das Verfahren macht den Unterschied
 - „OCR ist nicht gleich OCR“
 - Dubletten-Prüfung
 - Ausblick
 - Anhang

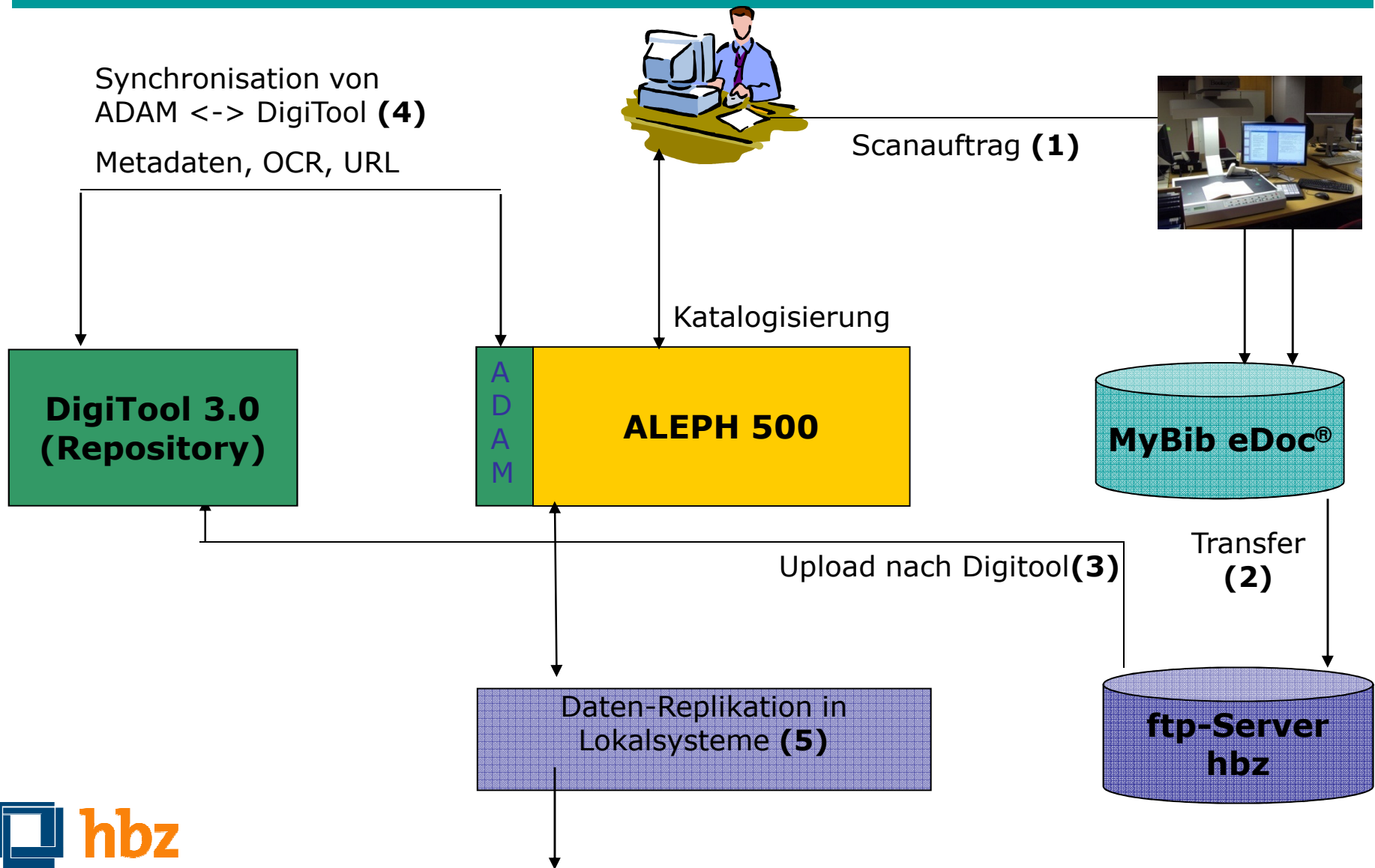
-
- 2005 hbz: Scanning von 180.000 Titeln (Retro)
 - 2006 Start im hbz: Scannen der Neuzugänge
 - 2007 GBV: diverse Retro-Projekte

 - 2008 DNB: Pilotprojekt Kataloganreicherung
 - 2009 DNB: Aufnahme als Standarddienst

 - 2010 IDS: Basel, Bern, Luzern Regelbetrieb
 - 2010 hbz: Übernahme der DNB-Daten
 - 2011 SNB: Einbindung der Schweizer National Bibliothek



Überblick hbz-Lösung



Zahlenüberblick per 30.04.2011



Wer	Was	Titel	Seiten
hbz	Retro & Neuzugänge (98% Nettodaten, zzgl. ca. 120.000 durch Import DNB-Titel)	auf Anfrage	auf Anfrage
GBV	Retro & Neuzugänge	auf Anfrage	auf Anfrage
DNB	Neuzugänge (100% Nettodaten, keine Dubletten!)	auf Anfrage	auf Anfrage
BVB	Neuzugänge	auf Anfrage	
IDS	Neuzugänge (100% Nettodaten, da Dubletten-Prüfung gegen DNB, hbz)	auf Anfrage	auf Anfrage
Sonstige	SWB, HEBIS, KOBV, Einzelbibliotheken Verfahren, „harte Zahlen“ etc. unbekannt	auf Anfrage	
	Gesamt (ca.)	S 1.550.000	4.500.000
	ImageWare Dubletten-Pool (ISBN/Verbund)	910.773	

S = Schätzung, **Titel** = nur Eigentitel ohne Fremddaten, **Seiten** = Zählung ohne Titelblatt!

OCR: Unterschiede in der Qualität



QS-Stufe	Verfahren	Bemerkung	Personal-aufwand für OCR/Seite	Anwender
1	nachgelagerte OCR ohne visuelle Kontrolle	problematisch bei mehreren Sprachen (SSG, Diakritika)	n/a	BVB
2	OCR beim Scannen mit manueller Sprachenauswahl	bessere Qualität bei mehrsprachigen Titeln	ca. 10 sec	
3	OCR beim Scannen mit manueller Sprachenauswahl & visueller Ergebniskontrolle	wie zuvor zzgl. schnellere Erkennung/Korrektur von Fehlern. Vor. 21“-Bildschirm	ca. 12 sec	DNB, hbz SBB-PK, usw.



Aspekt	Anmerkung	Varianten
Logo	Authentizität Ja/Nein	2
Titelblatt	z.B. kein IHV Ja/Nein	2
PDF	Durchsuchbar Ja/Nein	2
Mischbetrieb	Neuzugänge/Retro jeweils Ja/Nein	2 x 2
Dubletten-Prüfung	Ja/Nein und Prüfungskaskade (intern, Verbund, DN)	2 x 3
Massendigitalisierung	Retro im Batch > 10.000 Titel Ja/Nein	2
OCR-Verfahren	Qualitätsstufen 1 - 3	3
Verfahren	mit/ohne QS; impliziter/expliziter Scanauftrag	2 x 2
Konkordanz Buch/OPAC	Barcode/Medien-Nr. =/<> Verbund-ID usw.	ca. 9
Kartesisches Produkt = Anzahl Varianten		41.472

- 2005 bibliotheksinterne Dubletten-Kontrolle
=> Effekt ca. 15% weniger Scanaufträge
- 2007 Einführung hbz-verbundinterne Dubletten-Kontrolle & Bereinigung des Altbestandes aus 2005/2006 (Neuzugänge)
=> ca.15.000 Dubletten = 25% der Neuzugänge wurden gelöscht!
=> Seit 2007 ca. **25%** abgelehnte Scanaufträge, da Dubletten!

- 2008 endgültige Entscheidung für einen Workflow via
 - „impliziter Scanauftrag („das Buch an sich ist der Auftrag“) und gegen den
 - „expliziten Scanauftrag“ (manuelle Erfassung notwendig, wie BVB)

Gründe

- Wegfall Personalaufwand & Fehleingaben
- Wegfall Daten-Synchronisation & Differenzenklärung

-
- 2008** tagesaktuelle Übernahme der Daten
Scanning/OCR plus 1 Tag => Verfügbarkeit im hbz-OPAC
- 2010** verbundübergreifende Dubletten-Kontrolle zur Vermeidung von Doppelarbeiten unter Einbeziehung der Datenbestände
- DNB
 - GBV/SBB-PK (SSG-Recht), Braunschweig (SSG Pharmazie)
 - IDS Basel, Bern, Luzern (französische, italienische Titel)

Service Dubletten-Prüfung



Kundenportal ImageWare

Produkte ▾ Anwendungen & Lösungen ▾ Service ▾

Suche

[Startseite](#) > [Anwendungen & Lösungen](#) > [Kataloganreicherung](#) >

Dublettenprüfung

Dieser Service dient der schnellen Überprüfung, ob es innerhalb eines Verbundes schon Digitalisate zu einem Titel gibt. Die verbundübergreifende Suche ist über die ISBN möglich. Als Suchparameter dient die Verbundkennung (Ident-Nummer) des bibliographischen Datensatzes (Verbund-ID im HBZ, PPN im GBV und IDN in der DNB) bzw. die ISBN.

Zur Überprüfung wählen Sie den zu überprüfenden Verbundkatalog oder ISBN aus der Auswahlliste aus und geben die gesuchte Kennung ein. Die ISBN kann als String (978123456789) oder mit Bindestrichen (978-3-1234-5678-9) gesucht werden. Als Antwort bekommen Sie eine Rückmeldung ob Digitalisate vorhanden sind oder nicht. Suchbar sind nur ISBN-13. Anfrage mit 'alten' ISBN-10 werden ignoriert, da beim Anlegen neuer Scanaufträge alle ISBN-10 normalisiert werden!

Hinweis: Über diese Funktion läßt sich nur feststellen ob Digitalisate vorhanden sind, die im Rahmen des Catalogue Enrichment-Projektes erzeugt wurden. Es können keine anderen Digitalisate (beispielsweise LoC, Thieme, etc.) abgefragt werden.

Verbund/ISBN auswählen: *

Verbundkennung/ISBN-Nummer: *

Bitte geben Sie jeweils nur eine Verbundkennung ein

© 2010 ImageWare Components GmbH [Impressum](#) [Login](#)

-
- Übernahme des IDS-Datenbestandes (ca. 60.000 Titel)
 - ggf. Aufbereitung des Altbestandes von ca. 500.000 Titeln gemäß der DNB-Ansätze
 - durchsuchbare PDF's,
 - hbz-Logo,
 - Rückverweis auf hbz-ID
 - Ausbau in der Fläche durch die „hbz-Servicepakete“, um Bibliotheken mit Spezialbeständen in CE einzubinden
 - Weiteres hängt von der Kooperationsbereitschaft der Verbünde/DNB ab.

Service-Pakete	Leistungsumfang	Anwendung
Fullservice (SP01)	Scan-Station mit Personal für Scanning, OCR, QS erfolgt durch hbz/IWC. Raum & Bücher werden durch Bibliothek gestellt	Volumen > 2.000 Titel p.m. & Arbeitsplatz (!) z.B. bei Retroprojekten
Selfservice (SP02)	Wie zuvor, jedoch sind alle Aufgaben durch die Bibliothek wahrzunehmen.	Volumen > 1.000 Titel p.m. & Arbeitsplatz (!)
MixedService (SP03)	Wie zuvor, jedoch mit dem Unterschied, dass die Bibliothek (analog der Medea3 Fernleihe) nur scannt. OCR & QS erfolgen zentral bei hbz/IWC.	<ul style="list-style-type: none"> - Volumen < 1.000 Titel p.m. & Arbeitsplatz (!) - oder wenig Personal - oder hohe Personal-Fluktuation
hbz Infrastruktur	Infrastruktur, MyBib eDoc® Workflowsystem, 2nd-Level QS, 1st & 2nd Level-Support werden durch hbz beigestellt.	<ul style="list-style-type: none"> - Workflow-Überwachung - Dubletten-Kontrolle - Reporting

- Die Verfahren sind zwecks besserem Austausch (Titelblatt, Logo, etc.) weiter zu standardisieren.
- Ausbau der verbundübergreifenden Dubletten-Kontrolle
=> **Dubletten = Budgetverschwendung!**
- Reduzierung der Kosten durch bessere Nutzung der vorhandenen Infrastruktur an Scanstationen
- Kooperation DNB und hbz/Verbünde
Die DNB liefert einen wertvollen Basisdienst. In den Verbänden sollte darauf aufbauend folgendes gescannt werden:
 - fremdsprachige Titel
 - Retrokonversion
 - Dissertationen & Kongressbände

Kooperatives Scannen ist sinnvoller denn je!

Dankeschön!



ImageWare Components GmbH
Am Hofgarten 20
53113 Bonn

Telefon + 49 (0) 228 / 969 85 - 0
Telefax + 49 (0) 228 / 969 85 - 84
E-Mail info@imageware.de
Internet www.imageware.de



hbz – Hochschulbibliothekszenrum NRW
Jülicher Straße 6
50510 Köln

Telefon 0221 / 40075- 0
Telefax 0221 / 40075-180
E-Mail info@hbz-nrw.de
Internet www.hbz-nrw.de

-
- Scannen in der Praxis und im hbz
 - CE-Zahlen nach Sachgebieten (Stand 2008)

Aktuelle Scanstation für CE

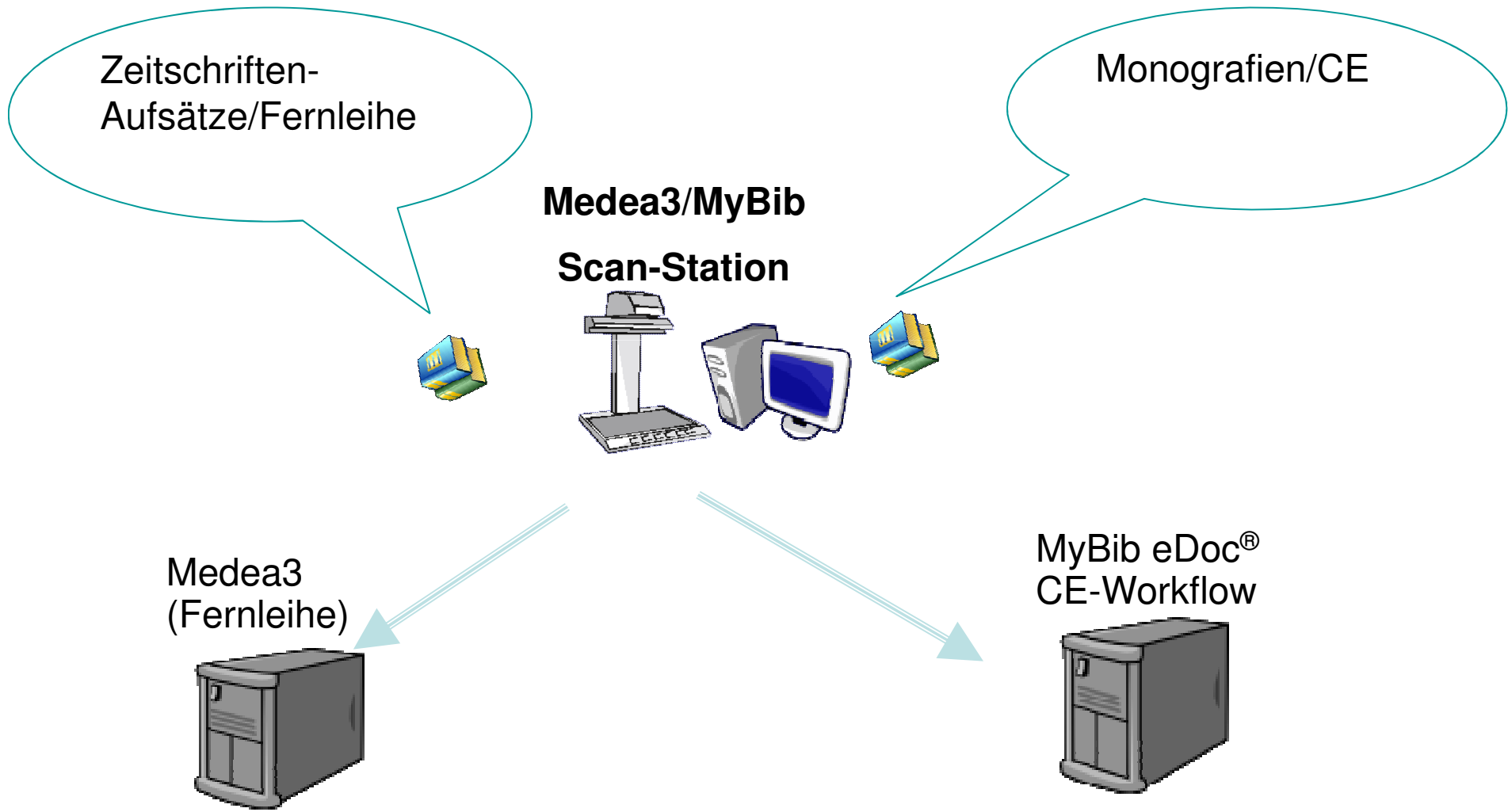


1. Buch per Barcode identifizieren
2. automatische & kaskadierende Dubletten-Prüfung
 - verbundintern
 - verbundübergreifend DNB, IDS
3. Scannen
4. Nachbearbeiten (DNB Bild reinigen)
5. manuelle Sprachauswahl
6. OCR durchführen
7. QS durch visuellen Abgleich von Scan und OCR-Ergebnis



Scannen über Medea3-Scanstation

(Mixed Service 1/2)

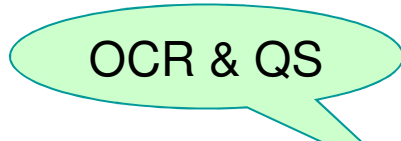
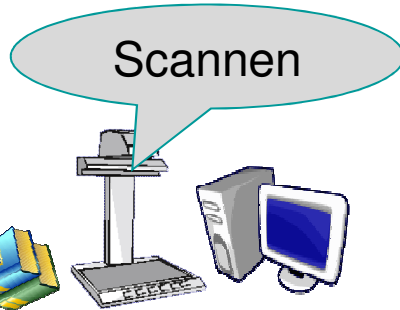


Trennung von Scannen & OCR

(Mixed Service 1/2)



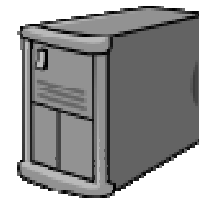
Bibliothek



ImageWare



MyBib



hbz

