

DIE ÖSTERREICHISCHE BIBLIOTHEKENVERBUND
UND SERVICE GMBH

obv sg



eDOC-Workshop 2007

V. Babitchev, J. Brandauer, U. Krabo

UB der Wirtschaftsuniversität Wien, 13. Juni 2007

Funktionen - 1

Datensammlung

- Batch-Client für regelmässige Datenübermittlung
 - alternativ in Ausnahmefällen: E-Mail für kleine Datenmengen
- URL-Objekte in eDOC – echte Objekte liegen auf lokalen Servern (Volltexte etc.)
- Sonderverfahren für OPUS, IVSCAN, ANNO
- **NEU: Datengenerierung (Objekte) aus ACC01**
 - E-Books, Volltexte

Datenkonversion / Datenspeicherung

- Konversion der bibliographischen Metadaten nach Dublin Core (Normalisierung)
- Speicherung von Meta- und Objektdaten in MySQL-Repository
 - Textextraktion aus externen Objekten bei Remote-Indexierung
- Extraktion der ‚Nutzerinformation‘ aus dem Verbundkatalog und Speicherung im eDOC-Repository

Funktionen - 2

Datenindexierung

- Indexierung von Meta- und Objektdaten
 - kombinierte Suche möglich
- Eigene Indizes für lokale Sichten und Sammlungen (Collections)
 - Anreicherung der Indizes mit Nutzerinformation (Gesamtsicht der Einrichtung)
- Remote-Indexierung speziell für Volltexte
- Indexierung erfolgt 1 x täglich (keine Online-Indexierung)

Recherche / Anzeige

- Einfache Suche in zentralen und lokalen Sichten sowie Sammlungen
 - Zentrale Sichten: eDOC-Suche und Österreichische Dissertationsdatenbank
- **NEU: Grundsätzliche Überarbeitung der Suchmaske!**
 - Bereitstellung einer ‚Erweiterten Suche‘
 - Liste weiterer eDOC-Opacs
- **NEU: Grundsätzliche Überarbeitung der Kurzlisten-Anzeige**
- **NEU: Individuelles ‚Branding‘ sowie Unterstützung lokaler Style-Sheets**

eDOC – Aktueller Status 2007 - 3

eDOC-Teilnehmer

Am eDOC-Service der OBVSG nehmen mit Stand Juni 2007 folgende Einrichtungen teil:

Scanning – Datenlieferung über eDOC-Batchclient [8 (9)]

- Oberösterreichische Landesbibliothek
- Universitätsbibliothek der Universität für Bodenkultur Wien
- Universitätsbibliothek der Montanuniversität Leoben
- Universitätsbibliothek der Technischen Universität Wien
- Universitätsbibliothek Graz
- Universitätsbibliothek Innsbruck
- Universitätsbibliothek der Wirtschaftsuniversität Wien
- Universitätsbibliothek der Medizinischen Universität Wien
- Universitätsbibliothek der Veterinärmedizinischen Universität Wien [Herbst 2007]

URL- bzw. Objekt-Übermittlung per Mail [2 (3)]

- Arbeiterkammer Wien
- Universitätsbibliothek der Universität für angewandte Kunst Wien
- FH der Wirtschaftskammer Wien [geplant]

IVSCAN - Lieferung fertiger Katalogisate

- Österreichische Nationalbibliothek

eDOC – Aktueller Status 2007 - 4

Statistik - Objekte

Stand 2007-06-11: **105.441 Objekte**

Objekt-Typ	Anzahl
Abstracts	49.849
Inhaltsverzeichnisse IVSCAN **	30.588
Inhaltsverzeichnisse	17.589
Umschlagbilder	3.162
Klappentexte	1.912
Volltexte	1.076
Rezensionen	272
Aggregate (Inhaltsverzeichnisse)	123
Titelseiten	36
Auszug	2
ONB-ANNO	832

** 2 Objekte für jedes Inhaltsverzeichnis: Image und html

Objekt-Zuwachs (exkl. IVSCAN u. DissDB)

Jahr	Objekte
2004	3.618
2005	5.153
2006	11.489
Bis 2007-06-11	4.745

eDOC – Aktueller Status 2007 - 5

Statistik – Speicherbedarf für Volltext-Objekte (VT) im Repository

Stand 2007-05-31

Inst.	Repository MB	VT intern MB	VT extern MB	Anzahl VT	Objekte ges.
AKW	16,32	16,32	90,64	88	88
BOK	131,82	-	-	-	3.315
LBO	21,39	-	-	-	1.520
MUL	960,22	571,26	-	141	6.200
TUW	386,79	219,66	5.122,40	776	10.645
UBG	272,21	-	-	-	12.105
UBI	114,99	-	-	-	12.016
VUW	1,11	0,49	53,62	5	1.764
WUW	78,99	-	-	-	5.393
ZBM	38,29	13,57	-	6	1.693

Repository Monthly Data Checks

From 01.2007 each months the following data checks performed

- results from 2007-06-01, 83.500 meta data records, 105.000 objects

- Meta data records without linked objects – **9**
- Meta records with more than 20 linked objects – **9**
 - E.g. Aggregat-Sätze
- AC-NRs not found in ACC01 – **61**
 - Deleted records in ACC01 etc.
- Non-accessible URL's – **14**

Results discussed with respective eDOC participants

Wguard - Werkzeug zum Schutz gegen unerlaubte Zugriffe

Background

In fall 2006 misuses in accessing eDOC were registered!

- multiple repeated downloads of selected objects
- download of a large number of objects in short time intervals

Wguard is a new tool running daily from Jan. 2007 providing following features:

- limits daily download access per IP-address (currently to 250 documents)
- registers and blocks for the rest of the day “robots” (access speed plus daily limit)
- supports special limits and time slots for privileged sites (e.g. partners working on documents exchange)
- supports black-listing (single IP or IP-masks)
- Produces daily statistics on
 - Total nr. of accesses
 - Frequent accesses
 - Blocked sites during the day
 - Accesses from black-listed sites
- Provides a new basis to restrict access to eDOC repository

e-DOC – Neue Funktionalitäten - 3

Neues Web-Design – Öffnung ‘versteckter’ eDOC-Funktionen

Neues Web-Design

- Flexibilität durch Verwendung von ‘templates’ und ‘style-sheets’
- Neue Suchmöglichkeiten durch
 - neue ‘Einfache Suchmaske’ und ‘Erweiterte Suchmaske’
 - Liste weiterer eDOC-Opacs

Grundsätzliche Überarbeitung der Swish-Kurzliste

- Neues Layout (neuer Link zur Vollanzeige, Thumbnails und Icons)
- Deduplizierung von Titeln mit mehreren Objekten
- Gruppierung und Sortierung aller Objekte zu einem Titel

Lokale Sichten

- Erweiterte ‘Branding’- und Layout-Gestaltungsmöglichkeiten (eigene style-sheets)
- Möglichkeiten zur Verwendung von ‘Einfacher’ und ‘Erweiterter’ Suche
- Möglichkeit der Bildung lokaler Collections

Online-Demo

Logische Sichten und Sammlungen in eDOC - 1

- Jedes Objekt in eDOC ist einem Owner zugeordnet = Liefernde Einrichtung
- Optional können die Objekte einer Sammlung = Collection oder Teilsammlung zugeordnet werden
- Alle diese Elemente sind in eDOC kodiert und durch Delimiter getrennt – ‚Ownerstring‘

Beispiele:

TOC der TU Wien:	‚TUW‘
Scan der ONB (IVSCAN):	‚ONB-IV‘
Hochschulschrift der UBG:	‚UBG-DISSDB-THD‘
E-Book von Springer:	‚SPR-EBOOK-STI‘

- Die Bildung von logischen Sichten bzw. Sammlungen entsteht durch Zuordnung bzw. Kombination einzelner Stringelemente zu den Sicht- bzw. Sammlungsindizes.

Logische Sichten und Sammlungen in eDOC - 2

Index für logische Sichten = eDOC-Sucheinstiege

- Sicht auf Einrichtungen:
Alle von TUW gelieferten und genutzten Objekte: → ‚TUW*‘
- Sicht auf Sammlungen:
Österreichische Dissertationsdatenbank:
Zuordnung der Objekte der UBG (UBG-DISSDB-THD) durch Stringpermutation:
→ ‚DISSDB-UBG*‘

Index für Collections

- Dieser Index definiert Sammlungen (Collections) bzw. Teilsammlungen innerhalb einer logischen Sicht
- Dient zur Einschränkung (Filterung) der Abfrage

Alle Springer E-Books der TUW:

Logische Sicht: → ‚TUW*‘

Collection: → ‚EBOOK*‘ oder

→ ‚EBOOK-STI‘ (wenn mehrere Packages vorhanden)

Ownerstring-Bildung

Einfache eDOC-Objekte

- Aus Objektnamen: tuwAC12345678n01in.pdf → ‚TUW‘

Österreichische Dissertationsdatenbank

- Ownerstring-Bildung aus MAB Kat. 519 → ‚TUW-DISSDB-THM‘

NEU: Objekt im bibliographischen Satz - 655e SFx

- E-Books, Volltexte (ALO) etc.
- Für jeden 655e-Eintrag wird bei definierter Subfeld-x-Struktur ein Objekt erzeugt

655e SFu http://
SFx SPR-EBOOK-STI → SPR-EBOOK-STI

655e SFu http://
SFx ZMB-ALO → ZBM-ALO

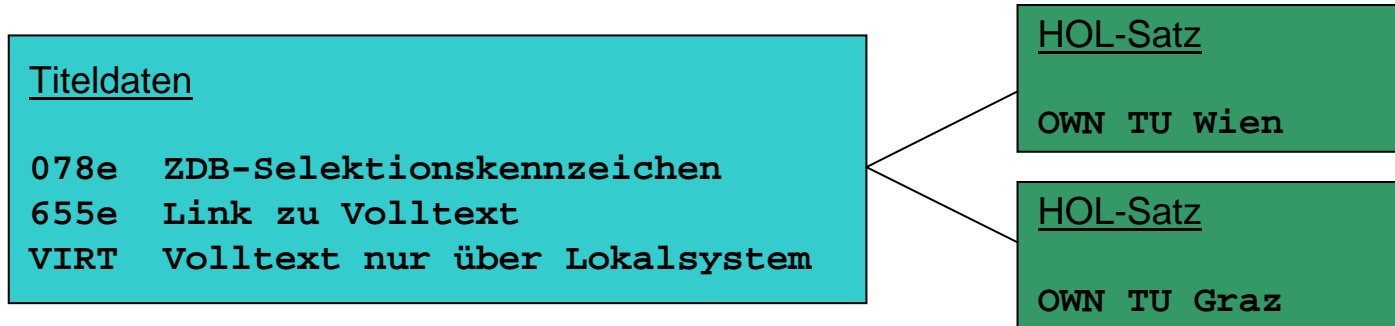
eDOC – Unterstützung für E-Books - 1

Ziel: Nachweis aller vorhandenen Volltexte in eDOC

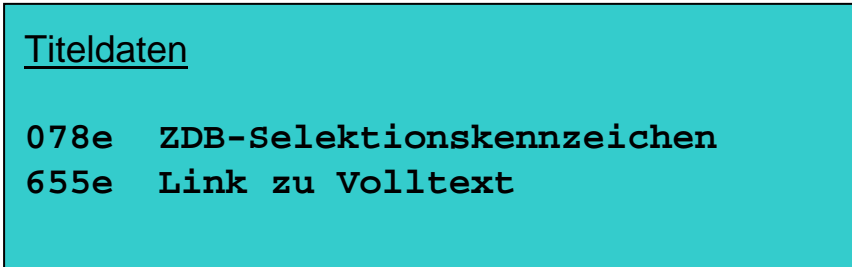
- E-Book-Pakete werden als Sammlungen definiert
- Frei zugängliche Objekte der Verleger-E-Books werden zusätzlich über eDOC indexiert !
- Front Matter = Inhaltsverzeichnis
- Durch den Besitzvermerk in Form von HOL-Sätzen können die E-Books auch in den lokalen Sichten angeboten werden
- Mit eDOC ist es möglich, auf die lokale Speicherung der Metadaten zu verzichten !!

Datenmodell für E-Book-Daten

Verbundsystem

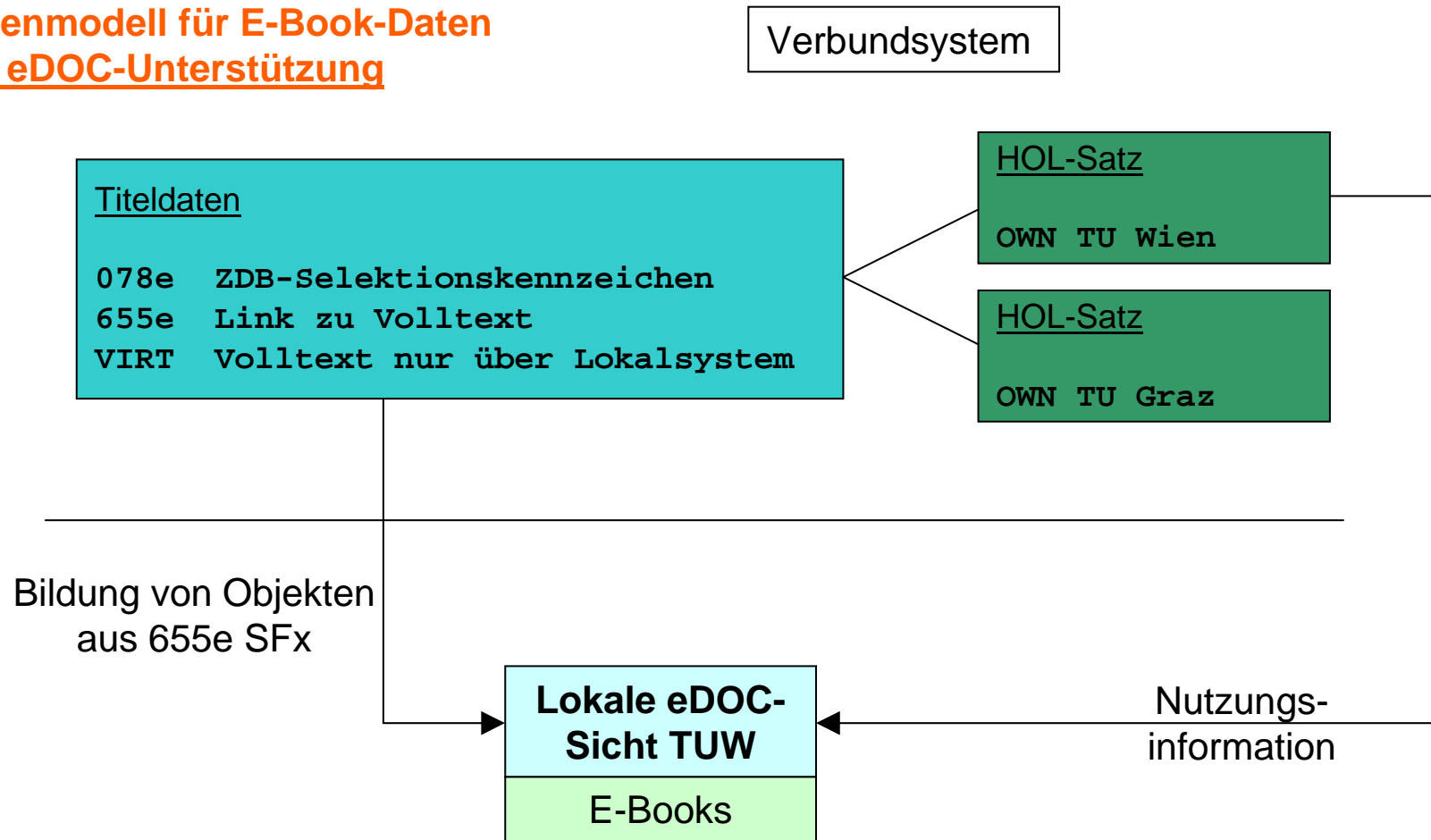


Lokalsystem



eDOC – Unterstützung für E-Books - 3

Datenmodell für E-Book-Daten mit eDOC-Unterstützung



e-DOC – Strategische Überlegungen

eDOC - Strategische Überlegungen

Wir sind der Ansicht, daß mit eDOC die Bedürfnisse vieler Einrichtungen gut und preiswert abgedeckt werden können – eine Ablöse ist noch nicht in Sicht!

Überlegungen für weitere Funktionen

- Lokale Sicht mit Durchschaltung auf lokalen Katalog
- Regelmässiger Update der Nutzungsinformation
- Automatischer eDOC-Update bei Änderungen in ACC01
- OAI-Schnittstelle für Dissertationsdatenbank

Überlegungen für weitere Daten

- OBVSG untersucht den Ankauf / Integration der TOCs von Casalini Libri

eDOC und koop3

- Bei Realisierung des koop3-Projekts sollen alle vorhandenen Anreicherungsverfahren - ADAM, DigiTool und eDOC - erhalten bleiben

eDOC und Primo

- OBVSG wird sehr genau die Primo-Entwicklung verfolgen
- Mit Primo könnte z.B. die lokale eDOC-Sicht abgelöst werden

eDOC – Architektur-Übersicht

