



Brigitte Doß

# 1. Community-Workshop

## Community-Aktion: Dubletten- Bereinigung

# Organisatorisches

Im Zoom: Bitte Stummschalten, Video gerne an 😊

Diskutieren erwünscht -> bitte per Handzeichen melden oder Fragen im Chat stellen

Folien werden in Menüpunkt "Erste Schritte" eingestellt

# DBIS – Erste Schritte

## Präsentationen, Anleitungen & Hilfestellungen

The screenshot displays the DBIS Administration interface for the Universitätsbibliothek Regensburg (UBR). The top right corner shows a user greeting: "Guten Tag, Name". The main content area is titled "DBIS-Administration" and "Universitätsbibliothek Regensburg (URL-BIBID: UBR)".

**DBIS-Ankündigungen**

Wir laden Sie ganz herzlich zur EZB-DBIS-Konferenz am 12.11.2024 in Regensburg ein. Sie können sich bis Ende Oktober über folgenden Link anmelden:  
[Programm und Anmeldung](#)

Erhalten Sie Infos über unsere DBIS-Mailinglisten. Hier können Sie sich dafür eintragen:  
[Zur DBIS-Info-Mailingliste](#)  
[Zur DBIS-Update-Mailingliste](#)

**Lizenzadministration**

[→ Lizenzen bearbeiten](#)

**Datenbanktitel**

[→ Titel neu eingeben](#)

The sidebar on the left contains the following navigation items:

- ORGANISATION UBR
  - Universitätsbibliothek Regensburg
  - Organisationsdaten bearbeiten
- DATENBANKEN
  - Datenbanken suchen
  - Datenbank erstellen
  - Datenbanken verknüpfen
- FACHGEBIETE
  - Top-Datenbanken verwalten
  - Fachgebiete verwalten
- SAMMLUNGEN
  - Sammlung erstellen
  - Sammlungen verwalten
- QUALITÄTSMANAGEMENT
  - Datenbereinigung
  - Schlagwörter bearbeiten
- INFORMATION & AUUSTAUSCH
  - DBIS-Profil
  - Erste Schritte
  - Kontakt

Alle Schulungsunterlagen (und auch die Workshops) verlinken wir unter „Erste Schritte“ in der Admin der neuen DBIS-Version

# Community-Aktion: Dubletten-Bereinigung

Neue Struktur von DBIS

Arten von Dubletten

Diskussion: EINE oder ZWEI Datenbankaufnahmen für freie und lizenzierte Datenbanken?

Vorschlag zum Vorgehen

Diskussion des Vorschlags

# Community-Aktion: Dubletten-Bereinigung

## Neue Struktur von DBIS

Arten von Dubletten

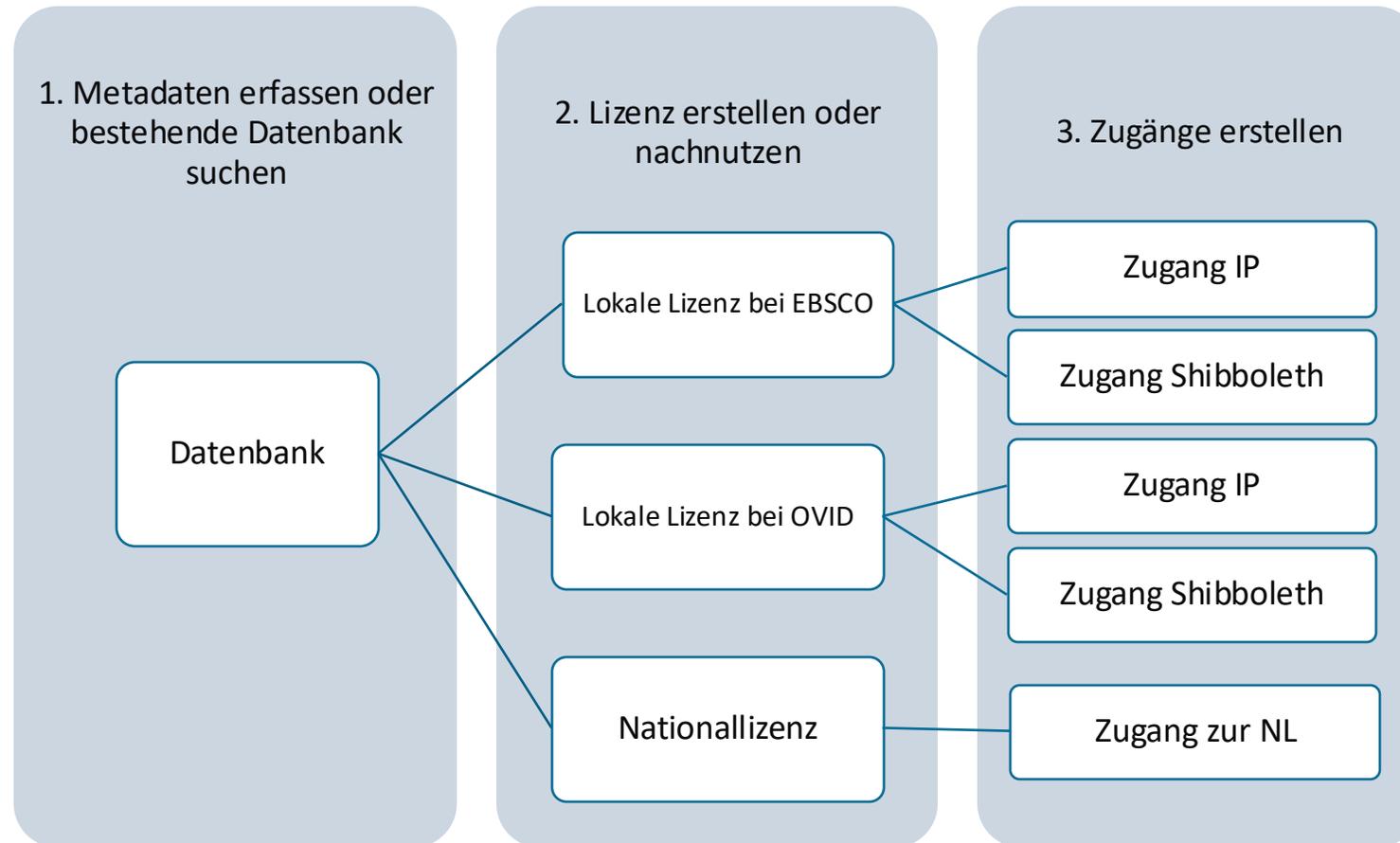
Diskussion: EINE oder ZWEI Datenbankaufnahmen für freie und lizenzierte Datenbanken?

Vorschlag zum Vorgehen

Diskussion des Vorschlags

# Datenbanken erstellen

## Neue Logik



# Community-Aktion: Dubletten-Bereinigung

Neue Struktur von DBIS

## Arten von Dubletten

Diskussion: EINE oder ZWEI Datenbankaufnahmen für freie und lizenzierte Datenbanken?

Vorschlag zum Vorgehen

Diskussion des Vorschlags

# Dubletten aus dem alten DBIS

„echte“ Dubletten

## ●○○ Die digitale Bibliothek des Kirchenkampfes (Dublette)

Die Digitale Bibliothek des Kirchenkampfes ist ein Projekt der Arbeitsgemeinschaft der Archive und Bibliotheken in der evangelischen Kirche (AABevK). Sie weist rund 15.000 Publikationen aus evangelischen Bibliotheken und Archiven nach, die in Zusammenhang mit dem sog. Kirchenkampf in der Zeit des Nationalsozialismus stehen. Die Webseite ermöglicht die Suche nach Druckschriften aus der Zeit des ...

[Mehr Informationen](#)

[→ Zur Datenbank](#)

## ●○○ Digitale Bibliothek des Kirchenkampfes, Die

Die Digitale Bibliothek des Kirchenkampfes ist ein Projekt der Arbeitsgemeinschaft der Archive und Bibliotheken in der evangelischen Kirche (AABevK). Sie weist rund 15.000 Publikationen aus evangelischen Bibliotheken und Archiven nach, die in Zusammenhang mit dem sog. Kirchenkampf in der Zeit des Nationalsozialismus stehen. Das primäre Ziel des Projektes ist ein vollständiger Online-Bestandsnac...

[Mehr Informationen](#)

[→ Zur Datenbank](#)

# Dubletten aus dem alten DBIS

## Verschiedene Anbieter

### ●●○ Medline

Die Online-Datenbank weist, zum Teil mit kurzen Zusammenfassungen, medizinische Fachpublikationen aus ca. 4.500 Zeitschriften nach, beginnend mit dem Jahr 1966. Berücksichtigt wird der Gesamtbereich der Medizin, einschließlich Krankenpflege, Dental- und Veterinärmedizin und Randbereiche wie z.B. Biologie, Biochemie und Psychologie. Der Datenbestand entspricht inhaltlich u.a. den gedruckten Dien...  
[Mehr Informationen](#)

[→ Zur Datenbank](#)

### ○○○ MEDLINE

MEDLINE ist die wichtigste bibliographische Datenbank mit dem Schwerpunkt Medizin. Sie weist Publikationen nach, die seit 1946 in ca. 5.200 Zeitschriften weltweit in verschiedenen Sprachen erschienen sind, wobei die englischsprachigen Quellen überwiegen. Die Dokumente sind durch Schlagwörter (Medical Subject Headings MeSH) erschlossen, wobei aktuelle Nachweise noch keine MeSH-Schlagwörter besit...  
[Mehr Informationen](#)

[→ Zur Datenbank](#)

### ●●○ Medline Complete

Diese Volltextdatenbank ist die größte Ergänzung zum MEDLINE-Index und bietet Zugang zu renommierten Fachzeitschriften in der Biomedizin und Gesundheitswissenschaft. MEDLINE Complete deckt ein breites Themenspektrum ab und ist deshalb eine unverzichtbare Ressource für Ärzte, Pflegekräfte und medizinisches Fachpersonal sowie für Forscher, die sich mit der klinischen Versorgung, Public Health und...  
[Mehr Informationen](#)

[→ Zur Datenbank](#)

### ●●○ Medline with Full Text

Medline with Full Text weist Artikel aus circa 4.800 medizinischen Fachzeitschriften nach, beginnend mit dem Jahr 1949. Im Unterschied zur Datenbank Medline enthält Medline with Full Text zusätzlich Volltexte aus 1.470 dieser Zeitschriften. Die Volltexte beginnen mit dem Jahr 1965. Zu den Artikeln ohne Volltext existieren zum großen Teil Abstracts. Berücksichtigt wird der Gesamt...  
[Mehr Informationen](#)

[→ Zur Datenbank](#)

### ●●○ MEDLINE über Web of Science

Die Online-Datenbank weist, zum Teil mit kurzen Zusammenfassungen, medizinische Fachpublikationen aus ca. 4.500 Zeitschriften nach, beginnend mit dem Jahr 1950. Berücksichtigt wird der Gesamtbereich der Medizin, einschließlich Krankenpflege, Dental- und Veterinärmedizin und Randbereiche wie z.B. Biologie, Biochemie und Psychologie. Der Datenbestand entspricht inhaltlich u.a. den gedruckten Dien...  
[Mehr Informationen](#)

[→ Zur Datenbank](#)

Dublette?

Keine Dubletten,  
da Name von  
Datenbank  
anders

# Dubletten aus dem alten DBIS

## Anderer Formaltyp: CD-ROM vs. online

### ●○○ British National Corpus

The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century, both spoken and written.

[Mehr Informationen](#)

[→ Zur Datenbank](#)

### ●○○ British National Corpus

Textkorpus des modernen British English, enthält ca. 4000 Ausschnitte aus gesprochenen und geschriebenen Texten (insgesamt über 100 Millionen Wörter)

[Mehr Informationen](#)

[→ Zur Datenbank](#)

### ○○● British National Corpus

Textkorpus des modernen British English, enthält ca. 4000 Ausschnitte aus gesprochenen und geschriebenen Texten, insgesamt über 100 Millionen Wörter.

[Mehr Informationen](#)

[→ Zur Datenbank](#)

### ○○● British National Corpus

Textkorpus des modernen British English, enthält ca. 4000 Ausschnitte aus gesprochenen und geschriebenen Texten (insgesamt über 100 Millionen Wörter)

[Mehr Informationen](#)

[→ Zur Datenbank](#)

### ○○● British National Corpus online

Textkorpus des modernen British English. Enthält ca. 4000 Ausschnitte aus gesprochenen und geschriebenen Texten, insgesamt über 100 Millionen Wörter.

[Mehr Informationen](#)

[→ Zur Datenbank](#)

# Dubletten aus dem alten DBIS

Freie Datenbankaufnahmen neben lizenzierte Datenbankaufnahmen

 **Statista**  
Statista ist ein Statistik-Portal, das statistische Daten verschiedener Institute und Quellen bündelt. Der Volltextzugriff ist abhängig vom Lizenzumfang Ihrer Einrichtung.  
[Mehr Informationen](#) [→ Zur Datenbank](#)

 **Statista Consumer Insights**  
Statistas Consumer Insights (bis 2022 u.d.T.: Statista Global Consumer Survey) untersuchen das Konsumverhalten und die Mediennutzung von KonsumentInnen weltweit bei Online-Und Offline-Aktivitäten. Grundlage sind mehr als 2.000.000 Online-Interviews in 56 Ländern und Regionen. Die Basis-Version der Global Umfrage wird in 35 Ländern und Regionen und die Erweiterte Umfrage in 21 Ländern durchg...  
[Mehr Informationen](#) [→ Zur Datenbank](#)

 **Statista Company Insights**  
Die 2023 von Unternehmensdatenbank zu Company Insights umbenannte Datenbank von Statista bietet Informationen zu über 70 Mio. börsennotierten und privaten Unternehmen aus mehr als 200 Ländern und Regionen.  
[Mehr Informationen](#) [→ Zur Datenbank](#)

 **Statista international**  
Statistiken zu 80.000 Themenbereichen aus über 18.000 Quellen zu 22 Branchenkategorien. Enthalten sind branchenspezifische Makrodaten, Trends und Prognosen. Sie haben direkten Zugang zu den wichtigsten sozialwissenschaftlichen Erhebungen in Deutschland wie ALLBUS (GESIS) und SOEP (DIW Berlin) Der Datenbestand ist international. Rechts oben können Sie über das Weltkugel-Symbol zwischen der...  
[Mehr Informationen](#) [→ Zur Datenbank](#)

 **Statista: das Statistik-Portal**  
Enthält kostenlose Statistiken zu allen Themen- und Wissensgebieten. Es ist allerdings nur ein Teil der Statistiken frei lesbar, und es sind keine Downloads möglich. Über einen kostenlosen Basis-Account mit einmaliger persönlicher Registrierung können etwa 10% der Daten umfangreicher genutzt werden; für die umfassende Nutzung aller Daten benötigt man einen sogenannten Premium Account, der mit K...  
[Mehr Informationen](#) [→ Zur Datenbank](#)

# Community-Aktion: Dubletten-Bereinigung

Neue Struktur von DBIS

Arten von Dubletten

**Diskussion: EINE oder ZWEI Datenbankaufnahmen für freie und lizenzierte Datenbanken?**

Vorschlag zum Vorgehen

Diskussion des Vorschlags

# Diskussion: Eigene Aufnahme für freie Lizenz?

## Derzeitige Regel:

### Statista: das Statistik-Portal

Enthält kostenlose Statistiken zu allen Themen- und Wissensgebieten. Es ist allerdings nur ein Teil der Statistiken frei lesbar, und es sind keine Downloads möglich. Über einen kostenlosen Basis-Account mit einmaliger persönlicher Registrierung können etwa 10% der Daten umfangreicher genutzt werden; für die umfassende Nutzung aller Daten benötigt man einen sogenannten Premium Account, der mit K...  
[Mehr Informationen](#)

[→ Zur Datenbank](#)

ist verwandt mit

### Statista

Statista ist ein Statistik-Portal, das statistische Daten verschiedener Institute und Quellen bündelt. Der Volltextzugriff ist abhängig vom Lizenzumfang Ihrer Einrichtung.  
[Mehr Informationen](#)

[→ Zur Datenbank](#)

## Andere Möglichkeit:

auch hier konsequent auf Dubletten verzichten

- Nur eine Aufnahme und 2 Lizenzen
- Bei grün und rot: grüne Ampel wird angezeigt
- Bei grün und gelb: gelbe Ampel wird angezeigt
- Beide Lizenzen und beide Ampeln in Detailsicht

# Community-Aktion: Dubletten-Bereinigung

Neue Struktur von DBIS

Arten von Dubletten

Diskussion: EINE oder ZWEI Datenbankaufnahmen für freie und lizenzierte Datenbanken?

**Vorschlag zum Vorgehen**

Diskussion des Vorschlags

# Vorarbeit für die Dubletten-Suche

## Pivot-Tabelle von Fuzzy-Titelableich

Anzahl von titel_id	titel	formal_type	titel_id	free_or_not	Ergebnis
	Library, Information Science & Technology Abstracts	2	5558	1	1
			8591	2	1
		2 Ergebnis			2
	Library, Information Science & Technology Abstracts Ergebnis				2
	Balance of payments statistics (BOPS)	1	3073	2	1
		1 Ergebnis			1
			9271	2	1
		2 Ergebnis			1
	Balance of payments statistics (BOPS) Ergebnis				2
	Policy Commons	2	104860	1	1
			105297	2	1
		2 Ergebnis			2
	Policy Commons Ergebnis				2
	Berg Fashion Library	2	7125	2	1
			10343	2	1
		2 Ergebnis			2
	Berg Fashion Library Ergebnis				2
	ARTstor - Images for Education & Scholarship	2	7253	2	1
			9156	2	1
		2 Ergebnis			2
	ARTstor - Images for Education & Scholarship Ergebnis				2
	Berliner Wetterkarte	1	8827	2	1
		1 Ergebnis			1
			2192	2	1
		2 Ergebnis			1
	Berliner Wetterkarte Ergebnis				2
	AGRICOLA	2	1076	1	1
			1085	2	1
		2 Ergebnis			2
	AGRICOLA Ergebnis				2

- 14425 Datenbank-aufnahmen aktuell in DBIS
- davon 891 Titel mit Dubletten (2-5) im Fuzzy-Titelabgleich
- Diese könnten reduziert werden auf bis zu 429 Titel.

# Herausforderungen zum derzeitigen Zeitpunkt

- Datenbankaufnahmen müssten angepasst werden, so dass sie lizenzunabhängig sind.
- Felder "Hinweise" und "Anleitungen" sind meistens nicht unabhängig von den Lizenzen. Auch Identifier z.T. nicht. Sollten noch zu Lizenz?!
- Erst dann, wenn Felderzuordnung und -sortierung final sind, macht es Sinn Metadaten umzuarbeiten?!
- Alle sind damit beschäftigt sich einzuarbeiten in die neue DBIS-Version, in der noch technisch nachgearbeitet wird.

# Vorschlag zum Vorgehen

- Abwarten bis Felder final sind
- Dann erst: Sammeln von Dubletten in gemeinsamer Liste oder in dafür eingeführten Feldern (beide ID + Kennzeichnung, welche ID erhalten bleiben soll)
- Markierung der Dubletten (in Titel oder noch besser: suchbares Markierungsfeld für Dubletten)
- Zweistufigen Verfahren bzw. 4-Augenprinzip?
- Zu bestimmen Zeitpunkt: maschinelle Reduktion der Dubletten
- Bitte bis dahin: Dubletten jeder für sich notieren, aber KEINE manuelle Bearbeitung von Alt-Dubletten (neu entstandene Dubletten dürfen natürlich weiterhin sofort eliminiert werden)

# Vorteile des Vorschlags gegenüber manuellem Zusammenlegen

- IDs könnten maschinell mitgeführt werden und gehen nicht verloren (es betrifft viele Links, die dann wieder auf die richtige Aufnahme geleitet werden könnten).

# | Allerdings...

... sind weitere Bearbeitungsschritte wie Verknüpfungen erstellen erst richtig sinnvoll, wenn Dubletten weg sind.

# Community-Aktion: Dubletten-Bereinigung

Neue Struktur von DBIS

Arten von Dubletten

Diskussion: EINE oder ZWEI Datenbankaufnahmen für freie und lizenzierte Datenbanken?

Vorschlag zum Vorgehen

**Diskussion des Vorschlags**

# Diskussion



# Monatliche Community-Workshops

- Dezember Termin entfällt
- 15.01.2025 Workshop “Community-Aktion: Datenbanken verknüpfen” 10:00 – 12:00
- 19.02.2025 Workshop “Community-Aktion: Qualitätsmanagement” 10:00 – 12:00



# Danke!

Dr. Brigitte Doß

[info.dbis@ur.de](mailto:info.dbis@ur.de)