

Elisabeth Mödden, Christa Schöning-Walter

Erschließung mit automatischen  
Methoden –

Auswirkungen auf die  
bibliothekarische Praxis

# Inhaltsverzeichnis

- 1. Strategische Ziele und organisatorische Rahmenbedingungen**
- 2. Methoden und Verfahren**
- 3. Blick in die Praxis**

# Strategische Prioritäten 2013 -2016

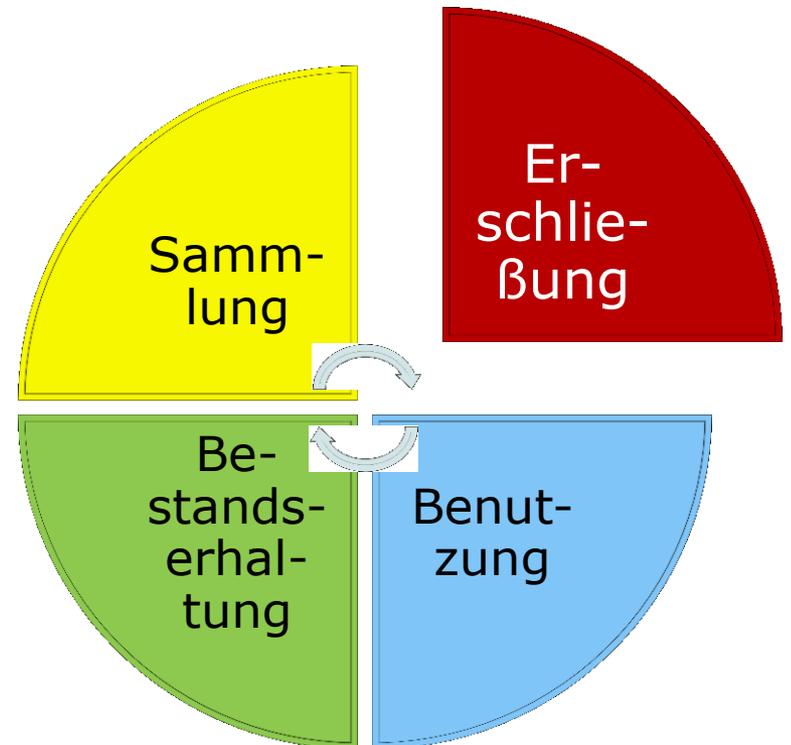
- Ziele und Prioritäten für die Weiterentwicklung in einem Vier-Jahreszeitraum
- Ausgangspunkt für einen fortlaufenden Strategieprozess
- Richtschnur für Planungen und Entscheidungen
- Arbeitsgrundlage und Instrument zur Priorisierung



# Ressort Erschließung

## Aufgaben:

- abteilungs- und standort-  
übergreifende Hinwirkung  
auf eine einheitliche und  
effiziente Erschließungspraxis
- Abstimmung von Regelungen,  
Verfahren und Projekten
  - formale + inhaltliche  
Erschließung
  - intellektuelle + maschinelle  
Erschließung



# Strategische Ziele des Ressorts Erschließung (1/2)

Die Deutsche Nationalbibliothek setzt verstärkt Verfahren zur automatischen Datengewinnung für die Erschließung und Verzeichnung ihrer Bestände ein.

- Ziel 1: Ausbau der Fremddatennutzung (insbesondere für die Bereiche Tonträger und Graue Literatur)
- Ziel 2: Weiterentwicklung der automatischen Erschließung und Ausdehnung auf weitere Publikationstypen und -gruppen

Neu:  
standortübergreifendes  
Referat 2.4  
(Erschließungsverfahren)

# Strategische Ziele des Ressorts Erschließung (2/2)

Die Deutsche Nationalbibliothek setzt verstärkt Verfahren zur automatischen Datengewinnung für die Erschließung und Verzeichnung ihrer Bestände ein.

- Ziel 3: führende Mitarbeit bei der Anpassung der Standards an die Erfordernisse der digitalen Produktion (Regelwerke, Erschließungsgrundsätze, Datenformate etc.)
- Ziel 4: aktive Kommunikation

# Methodenbeispiel 1: Strukturerkennung

maschinelle Titelblattanalyse und Übertragung formaler Erschließungsdaten in die Katalogisierungsmaske

## Anwendungspotential:

- einheitlich strukturierte, gedruckte Monografien
- Aktuell: modellhafte Implementierung und Evaluierung am Beispiel der Universitätsdissertationen

## Merkmale:

- regelbasiertes Analyseverfahren
- spezifische Regeln und Wortlisten für jeden Publikationstyp
- spezifischer Workflow (incl. Digitalisierung der Titelblätter)

# Methodenbeispiel 2: Austausch von Erschließungsdaten

## Anwendungsfälle:

1. Personennormdatenverknüpfung
2. parallele Online- und Printausgaben (Monografien)
3. Inhaltserschließungsdaten aus anderen Ausgaben eines Werkes (geplant)

## Anforderungen:

- hinreichende Metadatenbasis und Kriterien für den maschinellen Abgleich (Clustering-Verfahren)
- intellektuelle Erfassung/Bearbeitung der Entitäten in der GND

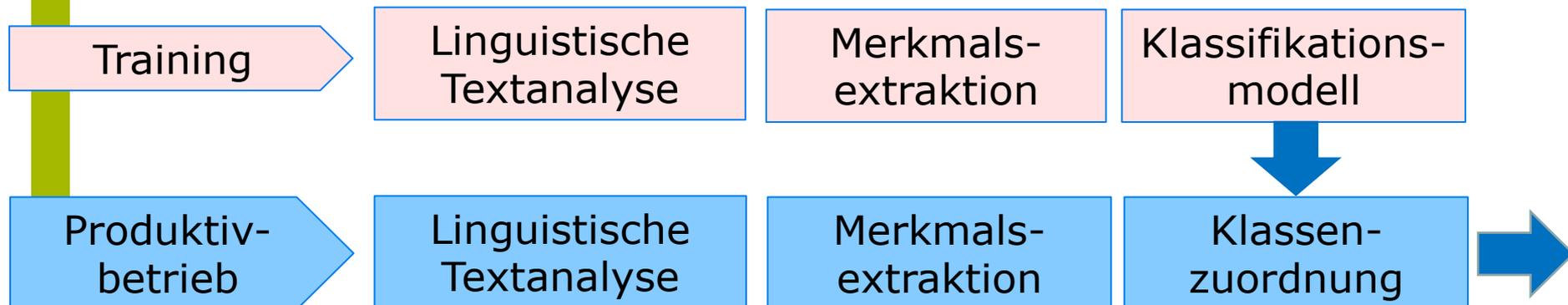
# Methodenbeispiel 3: Maschinelle Klassifikation

## Anwendungsfälle:

1. DDC-Sachgruppen
2. Medizinische Kurznotationen

Hier: Systematiken mit jeweils ca. 100 Klassen + Volltexte oder TOCs als Analysebasis

Modellbildung mit statistischen Lernverfahren:



# Methodenbeispiel 4: Maschinelle Beschlagwortung

Aktuell: praktische Erprobung anhand der Hochschulschriften

maschinelle Zuordnung von  
 bis zu 10 Schlagwörtern  
 aus dem Vokabular der GND  
 (Teilbestand s, Qualitätslevel 1)

- ca. 370.000 Personen
- ca. 205.000 Geografika
- ca. 185.000 Sachschlagwörter
- ca. 143.000 Körperschaften
- ca. 88.000 Werke
- ca. 12.000 Kongresse
- ca. 4.700 Hinweissätze

Linguistische Textanalyse

Abgleich der Terme  
 (Segmente) mit dem  
 Wörterbuch: Identifikation  
 der „passenden“ GND-  
 Entitäten, Disambiguierung  
 mehrdeutiger Terme etc.

Gewichtung und Selektion

Ausgabe der Schlagwörter

# Aufbau und Pflege der Wörterbücher

- regelmäßiges Einspielen der GND-Terminologie, wahlweise als Voll-Update oder als inkrementelles Update
- Pflege der Wörterbuch-Profile (Terminologie-Plattform): bevorzugte Benennungen, Synonyme, komplette Datensätze oder auch ganze Teilbäume der GND können hinsichtlich der gewünschten Verarbeitungsform markiert werden

DEFAULT – Term wird segmentiert

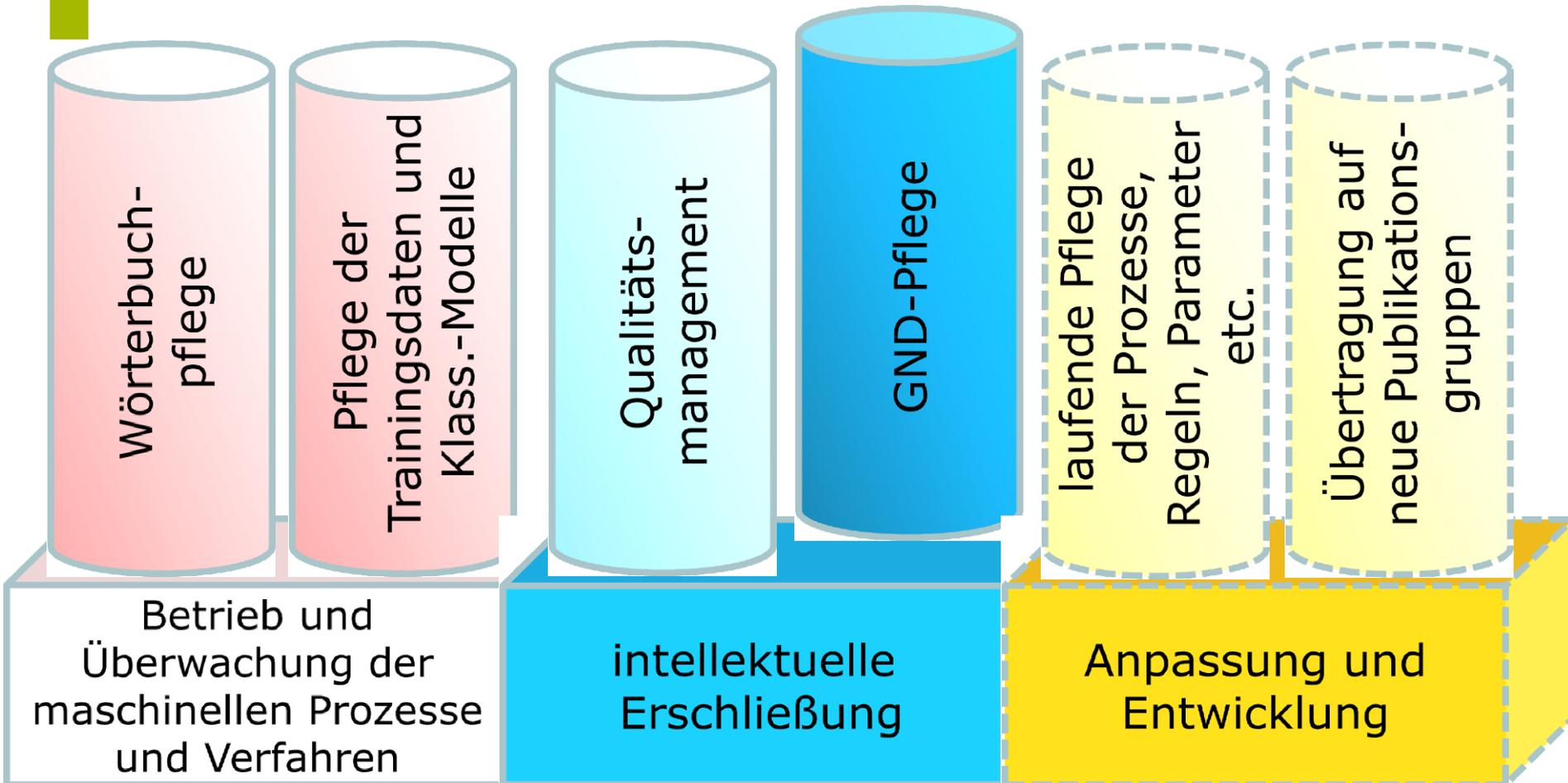
EXACT – Term wird nicht segmentiert (also in der exakten Schreibweise prozessiert)

IGNORE – Term wird „stillgelegt“  
(also nicht in das Wörterbuch eingelesen)

d.h.:

Ausschluss  
 problematischer  
 Terme  
 möglich

## Neue Aufgaben auf Dauer



# Blick in die Praxis



Gescanntes  
Titelblatt und  
Inhaltsverzeichnis

Online-Publikation

Fremddaten

Maschinelle Erschließungsverfahren



# Blick in die Praxis

Maschinelle Erschließungsverfahren

Strukturerkennung



Titelaufnahme



Maschinelle Sachgruppenvergabe

Maschinelle Vergabe mediz. Kurznotationen

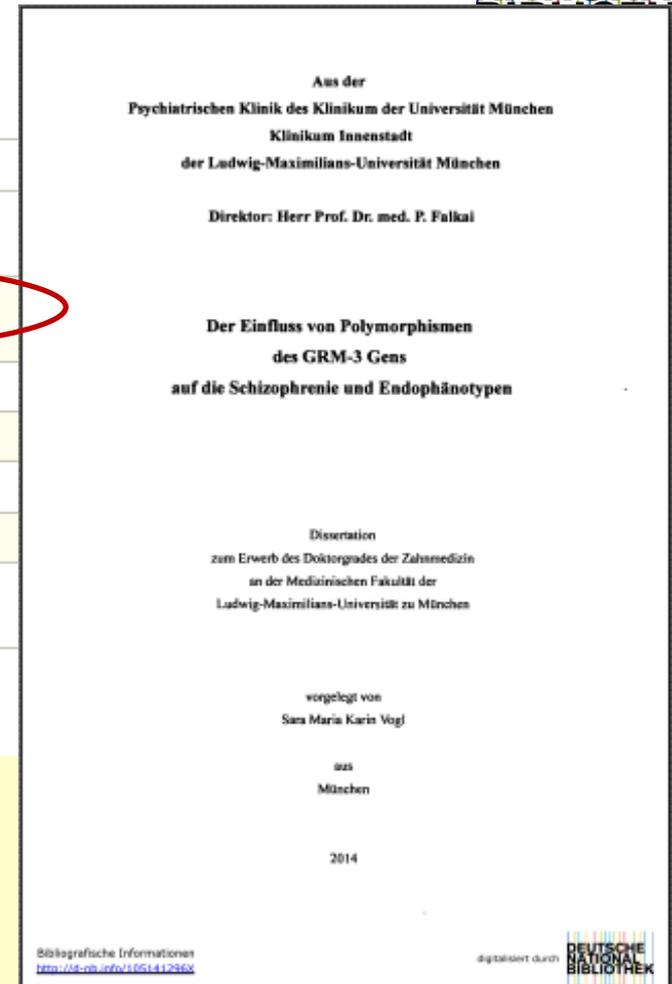
Maschinelle Beschlagwortung



Inhaltliche Erschließung

# Strukturerkennung – Titelaufnahme

	
<b>Link zu diesem Datensatz</b>	<a href="http://d-nb.info/105141296X">http://d-nb.info/105141296X</a>
<b>Titel/Bezeichnung</b>	Der Einfluss von Polymorphismen des GRM-3 Gens auf die Schizophrenie und Endophänotypen / Sara Maria Karin Vogl
<b>Person(en)</b>	Vogl, Sara Maria Karin
<b>Erscheinungsjahr</b>	2014
<b>Hochschulschrift</b>	München, Univ., Diss., 2014
<b>Sprache(n)</b>	Deutsch (ger)
<b>Weiterführende Informationen</b>	<a href="#">Inhaltsverzeichnis</a>



0500 Aaf  
0599 14-05-22 : f  
0600 rh  
0701 /f/#2  
1100 2014  
1500 /1ger  
2240 DNB:105141296X  
3000 Vogl, Sara Maria Karin  
4000 Der Einfluss von Polymorphismen des GRM-3 Gens auf die Schizophrenie und Endophänotypen / Sara Maria Karin Vogl  
4204 München, Univ., Diss., 2014  
4715 =u \$=c 04=d DNB=e 1

# Strukturerkennung – Titelaufnahme

0500 Aaf  
 0599 14-05-20 : f  
 0600 rh  
 1100 9999  
 1500 /1eng  
 2240 DNB:1050907108  
 3000 Mishra, Jayanta Kumar  
 4000 / Jayanta Kumar Mishra  
 4062 21 cm  
 4204 Diss., 9999  
 4715 =u \$=c 04=d DNB=e 1

1 - 2 von 2 Datum (neuestes zuerst) sortieren →

«« « »»

1 Datensatz kann nicht angezeigt werden. L F

GND 2 Mishra, Jayanta Kumar  
 1981- / Physiker

1 - 2 von 2 Gehe zu →

«« « »»

# Strukturerkennung – Titelaufnahme

Alternative Einsatzmöglichkeiten:

1. Unterstützung  
regelwerksgerechter Titelaufnahmen 
2. Basistitelaufnahme  
(Ergänzung und Verbesserung falls notwendig)
3. Stichproben  
zur Überwachung und technischen Verbesserung  
der Strukturerkennung

# Maschinelle Beschlagwortung

- 4000** Klinische Studie zur Überprüfung der Effektivität verschiedener Zahnbürsten im Rahmen der Initialbehandlung parodontal erkrankter Patienten
- 5050** **610**\$Em\$Hdnb\$K1,000 ...
- ...
- 5540** ...Zahnbürste\$K0,939 ...
- 5540** ...Klinisches Experiment\$K0,392 ...
- 5540** ...Patient\$K0,199 ...
- 5540** ...Initiale\$K0,041 ...
- 5540** ...Parodontium\$K0,033 ...
- 5540** ...Mikrobiologische Diagnostik\$K0,031
- 5540** ...Alveolarknochen\$K0,030 ...
- 5540** ...Zahnpflege\$K0,030 ...
- 5540** ...Zahnfleisch\$K0,029 ...
- 5540** ...Parodontitis\$K0,028 ...

# Maschinelle Beschlagwortung

Alternative Einsatzmöglichkeiten:

1. Zur Unterstützung für regelgerechte Beschlagwortung nach RSWK.
2. Maschinelle Beschlagwortung mit Stichprobenkontrollen zur Überwachung und zur Verbesserung der Verfahren.



# Maschinelle Beschlagwortung

**4000** Klinische Studie zur Überprüfung der Effektivität verschiedener Zahnbürsten im Rahmen der Initialbehandlung parodontal erkrankter Patienten

**5050** **610**\$Em\$Hdnb\$K1,000 ...

**5100** Zahnbürste

**5101** Parodontitis

...

**5540** ...Zahnbürste\$K0,939 ...

**5540** ...Klinisches Experiment\$K0,392 ...

**5540** ...Patient\$K0,199 ...

**5540** ...Initiale\$K0,041 ...

**5540** ...Parodontium\$K0,033 ...

**5540** ...Mikrobiologische Diagnostik\$K0,031

**5540** ...Alveolarknochen\$K0,030 ...

**5540** ...Zahnpflege\$K0,030 ...

**5540** ...Zahnfleisch\$K0,029 ...

**5540** ...Parodontitis\$K0,028 ...

Zur Unterstützung für  
regelgerechte  
Beschlagwortung  
nach RSWK.

# Maschinelle Beschlagwortung – Bewertung der Titelstichproben

**4000** Numerische Analyse des Nachstroms und Propellereffektivität  
am manövrierenden Schiff

**5540** [GND]...Schiff\$K0,476... wenig nützlich

**5540** [GND]...Manöver\$gSchiffahrt\$K0,356... sehr nützlich

**5540** [GND]...Wirbelschleppe\$K0,312... falsch

**5540** [GND]...Schiffsantrieb\$K0,041... nützlich

**5540** [GND]...Nachstrom\$K0,003... falsch

Vierstufige  
Bewertungsskala:  
sehr nützlich / nützlich /  
wenig nützlich / falsch

005 Ts1  
006 <http://d-nb.info/gnd/4776222-6>  
011 s  
035 gnd/4776222-6  
039 swd/4776222-6\$vg  
065 31.9a  
083 621.31\$d2\$t2009-09-29  
150 Nachstrom  
550 !040707458!Elektrischer Strom\$4obal  
670 Lex. Elektrotechn.

# Maschinelle Beschlagwortung – Bewertung der Titelstichproben

**4000** Numerische Analyse des Nachstroms und Propellereffektivität  
am manövrierenden Schiff

**5540** [GND]...Schiff\$K0,476... wenig nützlich

**5540** [GND]...Manöver\$gSchiffahrt\$K0,356... sehr nützlich

**5540** [GND]...Wirbelschlepe\$K0,312... falsch

**5540** [GND]...Schiffsantrieb\$K0,041... nützlich

**5540** [GND]...Nachstrom\$K0,003... falsch

**5540** [FA]...Nachstrom\$gStrömungsmechanik

**5540** [FA]...Schiffspropeller

**5540** [FA]...Numerische Strömungssimulation



# Herausforderungen aus der praktischen Anwendung

- GND-Pflege
  - Entwicklung eines GND-Pflege-Tools
- Wörterbuchpflege
- Suchmaschinenoptimierung
- Auslieferung der Daten
  - Kennzeichnung der Daten
- Verbesserung der maschinellen Verfahren durch Qualitätsmanagement
- Entwicklung neuer Komponenten

## Kontakt und weitere Informationen

Christa Schöning-Walter

Referentin im Ressort

Erschließung

Telefon: +49 69 1525 1014

mailto: [c.schoening@dnb.de](mailto:c.schoening@dnb.de)

Elisabeth Mödden

Referatsleitung

Erschließungsverfahren

Telefon: +49 69 1525 1533

mailto: [e.moedden@dnb.de](mailto:e.moedden@dnb.de)

<http://www.dnb.de/DE/Wir/Projekte/Abgeschlossen/petrus.html>