

Mödden, Elisabeth; Poley, Christoph

Künstliche Intelligenz in der Erschließung

Schlaglichter Automatische Erschließung in der DNB

- 2006: Novelle DNBG für unkörperliche Medienwerke – Sammelauftrag
 - Seit 2010 Reihe O
- Projekt Petrus (2009-2011)
 - Seit 2012: Maschinelle DDC-Sachgruppenvergabe
 - Seit 2014: Maschinelle GND-Schlagwortvergabe
- Seit 2014 Bündelung Aktivitäten für Automatische Erschließung und Sammlung von Netzpublikationen im Referat AEN

Ziele KI bzw. automatischer Verfahren

- Verarbeitung / Anreicherung von Massendaten
- Qualitätsverbesserung der Daten
- Vernetzung von Informationen
- Gewinnung neuer Zusammenhänge
- Politische Ziele, z.B. Einsparung von Personal



...

Welche Ziele sind realistisch?

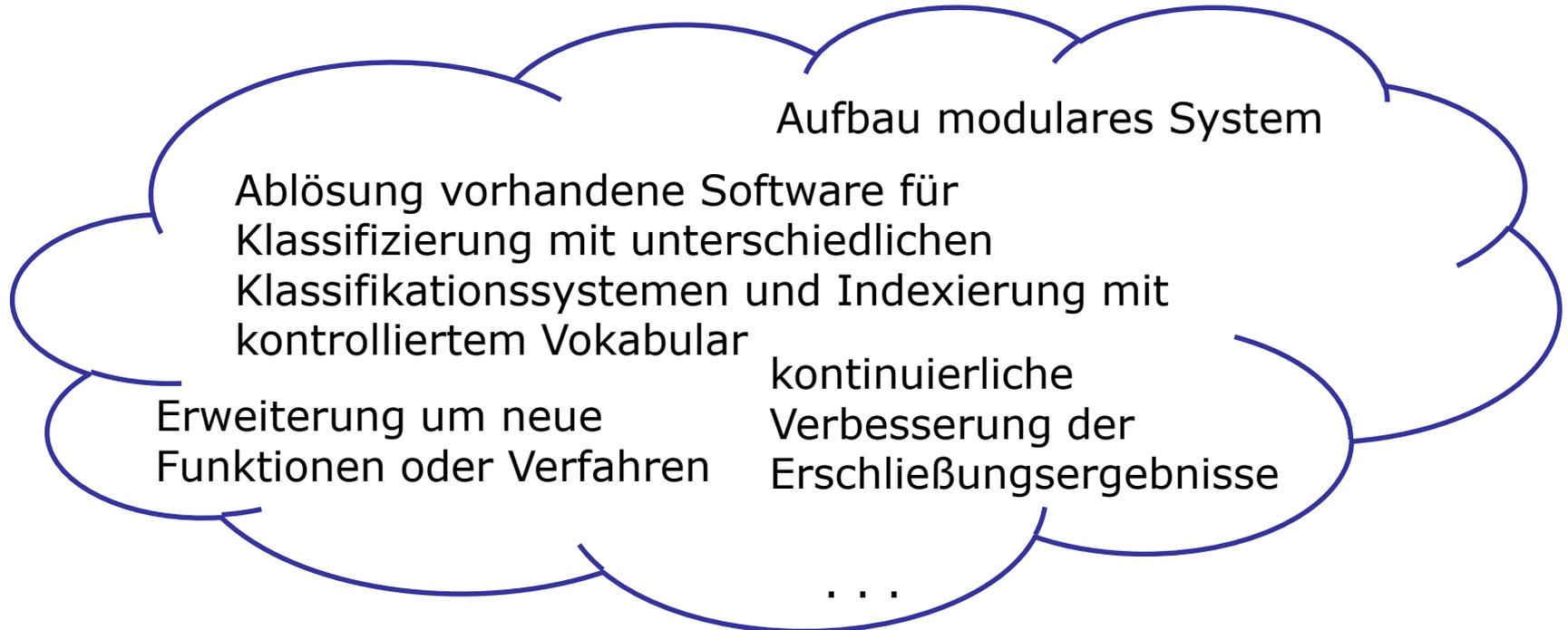
Halbautomatische Erschließung von gedruckten Universitätsdissertationen

- Unterstützung der Erschließung durch maschinelle Methoden unter Einbindung des Geschäftsgangs zur Kataloganreicherung zwecks Vereinfachung und Beschleunigung der Bearbeitung
- Modellprojekt mit Universität Innsbruck 2012 - 2015:
 - Digitalisierung der Titelblätter
 - Texterkennung, Strukturerkennung
 - Metadatenextraktion
 - Automatische Übertragung in die Katalogisierungsmaske

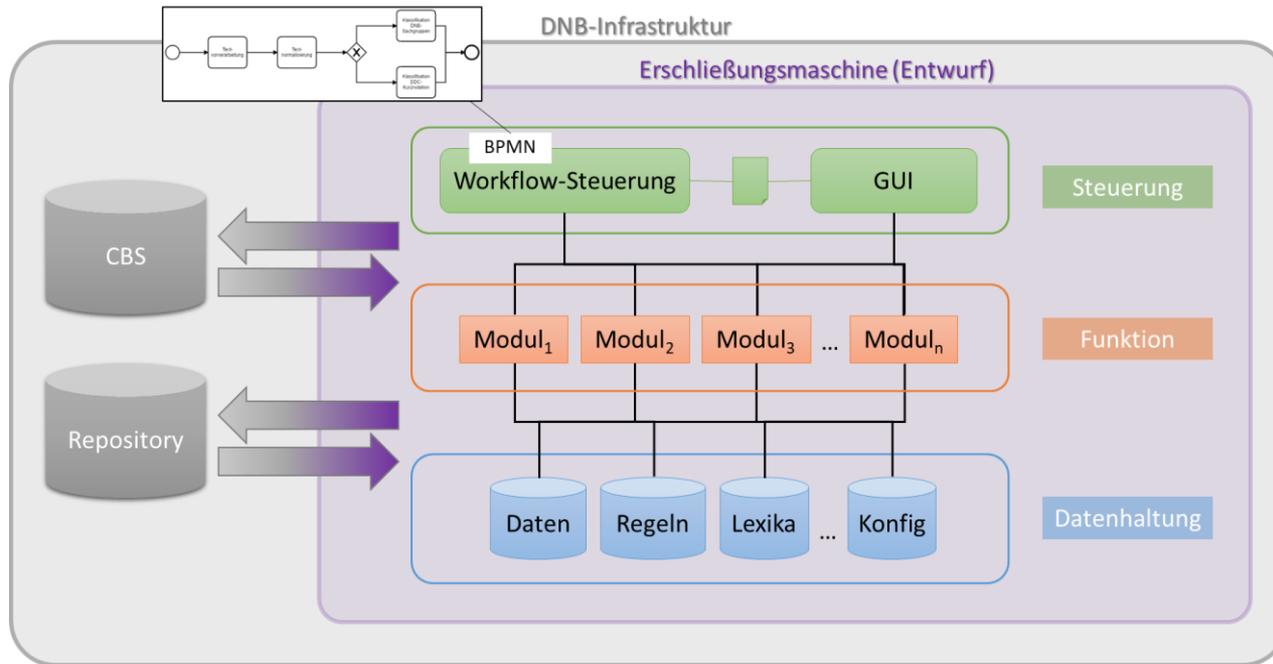
Halbautomatische Erschließung von gedruckten Universitätsdissertationen

- Zeitersparnis bei guten Extraktionsergebnissen 4 vs. 7 min bei konventioneller Katalogisierung
- Risiken: Mangelnde Texterkennung -> Nacharbeitung -> Optimierung Verfahren -> Zeitersparnis?
- Ergo: Automatismen helfen nicht automatisch zu kostengünstigeren Verfahren oder besseren Ergebnissen

Projekt EMa: Ziele



Modulare Systeme entwickeln



Transformation KI-Systeme

Ist



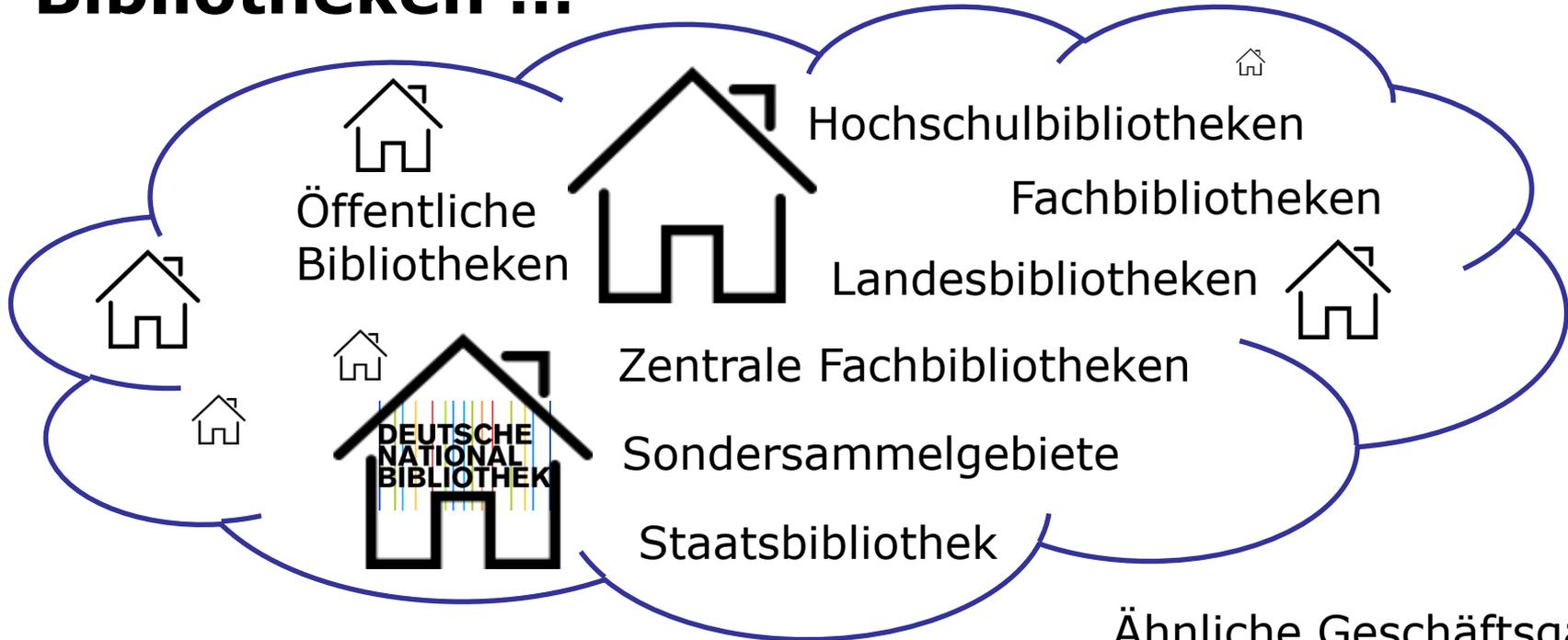
Soll

**Statische
Einzellösungen (für
Sacherschließung)**

**Modulare Lösungen auf
Basis von Kooperation**

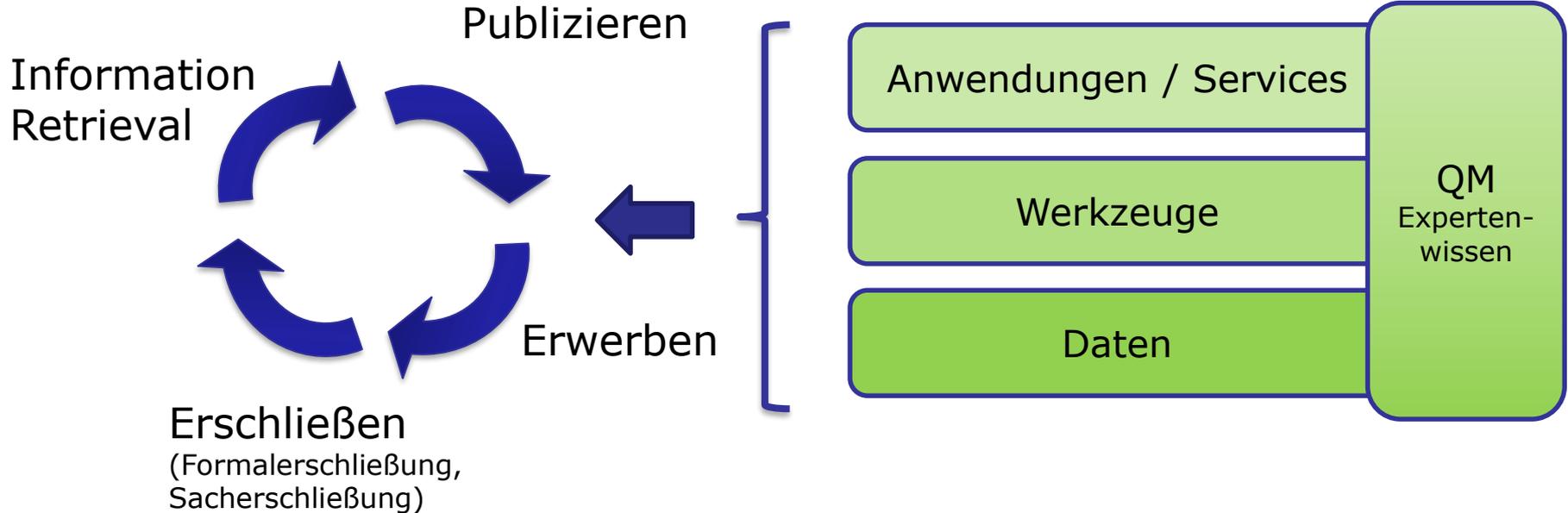
- Bibliotheken
- Kommerzielle/ freie Anbieter
- Forschung

Bibliotheken ...



Ähnliche Geschäftsgänge,
viele Gemeinsamkeiten, verschiedene Schwerpunkte.

Modularität und Kooperation



Daten

- Trainings-/Testdaten
 - Metadaten (PICA, MARC21, JATS, ONIX, DC, ...) + Goldstandard
 - Inhaltsverzeichnisse, Abstracts
 - Freie, lizenzierte Volltexte
 - (anonymisierte) Nutzerdaten ...
- Terminologien, Wörterbücher, ... mit korrespondierenden Metadaten
- Fertig trainierte Modelle
- ...

Werkzeuge

- Textextraktion / OCR
- Spracherkennung
- Klassifikation
- Indexierung

- Anbindung von Katalogsystemen
- Workflow-Engines
- ...

Anwendungen / Services

- Basis: Verzeichnis von Werkzeugen / Daten
- Offline: Bereitstellung einzelner Anwendungen bzw. Komponenten (Open Software(-pakete), Docker, VMs, ...)
- Online: Bereitstellungen fertiger Services (z.B. via ReST bzw. OpenURL, ...)
- QM: Bewertung von Ergebnissen maschineller Erschließung (intern / extern)
- ...

Fazit

- Chance KI:
 - kann (nicht nur in der Sacherschließung) helfen, zusätzliche Aufgaben in einer Bibliothek zu erfüllen
 - Die intellektuell zu leistenden Aufgaben ändern sich stückweise (Expert*innen, QM)
- Voraussetzung: Modulare Lösungen und Kooperationen
- Desiderat: Zugang zu KI-Systemen muss vereinfacht werden
- In eigener Sache: 03./04.12.2020 Online-Workshop: Erfahrungen und Perspektiven mit dem Toolkit Annif: <https://wiki.dnb.de/x/GwfmC>

Herzlichen Dank 

Elisabeth Mödden

Christoph Poley