

---

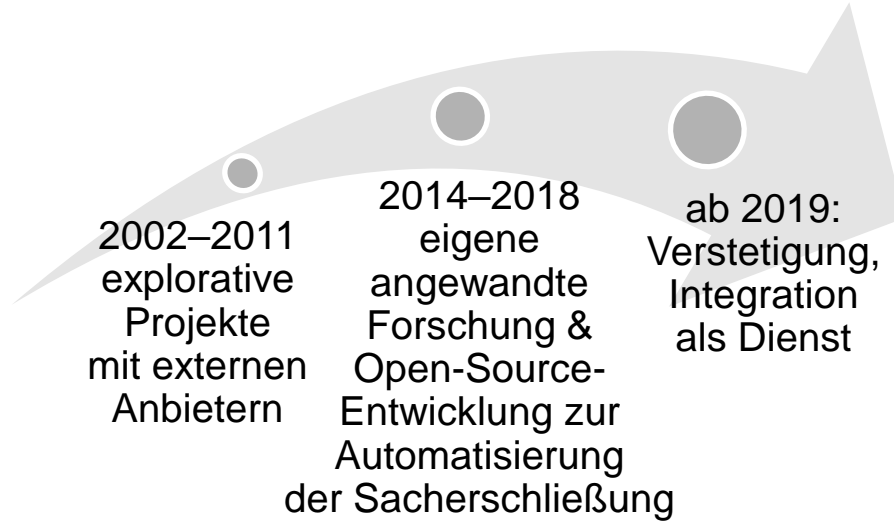
# Automatisierte Inhaltserschließung an der ZBW (AutoSE) und deren Anzeige im DA-3 (*zbwase*)

*Dr. Anna Kasprzik,*

*ZBW – Leibniz-Informationszentrum Wirtschaft*

*Workshop „Computerunterstützte Inhaltserschließung“, 10.+11.11.2021, virtuell*

# AutoSE: Der Weg zum verstetigten Forschungstransfer



# Methodenentwicklung

---

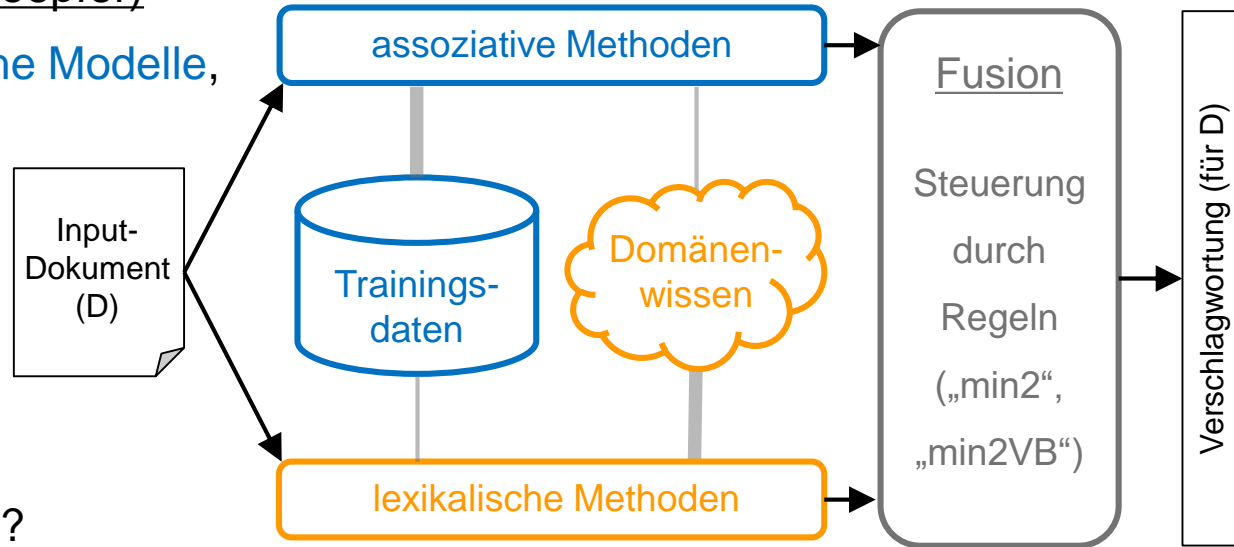
- ab 2016: Doktorand Martin Toepfer entwickelt im Rahmen angewandter Forschung einen **Prototypen für einen regelgesteuerten Fusion-Ansatz** für eine automatisierte SE
  - *meanwhile in Helsinki* ... Team an Finnischer Nationalbibliothek (NLF) entwickelt **Annif**, ein Toolkit mit dem Anspruch, niederschwellig einsetzbar zu sein
- ab 2019:
  - ZBW übernimmt **Annif als „Steckrahmen“** für AutoSECore und **flankiert** diesen mit Mechanismen für das wiss. Experimentieren, Hyperparameteroptimierung, diverse Mechanismen für Qualitätskontrolle, Anschluss an Erschließungsworkflows, etc.
  - ZBW **arbeitet an der Open-Source-Entwicklung von Annif mit**, gibt zusammen mit NLF **Tutorials** zu Annif und **berät** DNB und andere Institutionen zu dessen Einsatz



# Methodenentwicklung

bis 2018 (wiss. MA Martin Toepfer)

- assoziative und lexikalische Modelle, inkl. prototypischer Eigenentwicklung: *stwfsa*
- kombiniert in einem sog. **Fusion-Ansatz** – nachgeschaltete **Regeln**
- Forschungsfrage: wie mit *concept drift* umgehen?
- automatisierte Qualitätsabschätzung (prototypische Eigenentwicklung: *qualle*)



# Methodenentwicklung

ab 2019 (wiss. MA Moritz Fürneisen)



- Fusion-Ansatz abgelöst: Einsatz des Open-Source-Tools **annif** (*ensembles*), weiterhin ergänzt durch nachgeschaltete Anwendung von Regeln und Filtern
- separat durchgeführte **Hyperparameteroptimierung** (bietet Annif aktuell nicht)
- Einsatz einer Kombination von **state-of-the-art-Algorithmen** und maßgeschneiderter Eigenentwicklung (reimplementiertes **stwfsa**) *omikuji*  
*parabel* *bonsai*
- mittlerweile auch Experimente mit Ansätzen aus dem **Deep Learning**, *fastText*  
insbesondere mit **Transformermodellen** (Google BERT & Co.)
- Ansätze zur automatisierten Qualitätskontrolle (reimplementiertes **qualle**)

# Informationen zu Daten aus der Quelle „zbwase“

---

aktuell trifft Folgendes zu (kann sich in Zukunft ändern/erweitern):

- nur Publikationen mit (ausschließlicher!) Publikationssprache „**englisch**“
- nur auf der Basis des **Titels**, und, wenn vorhanden, der **Autoren-Keywords**
- wir prüfen die EconBiz-Datenbasis **stündlich** auf neue Publikationen im Bestand der ZBW und verschlagworten sofort  
(EconBiz: Rechercheportal der ZBW)  
(EconBiz-Datenbasis wird aus der WinIBW gespeist)

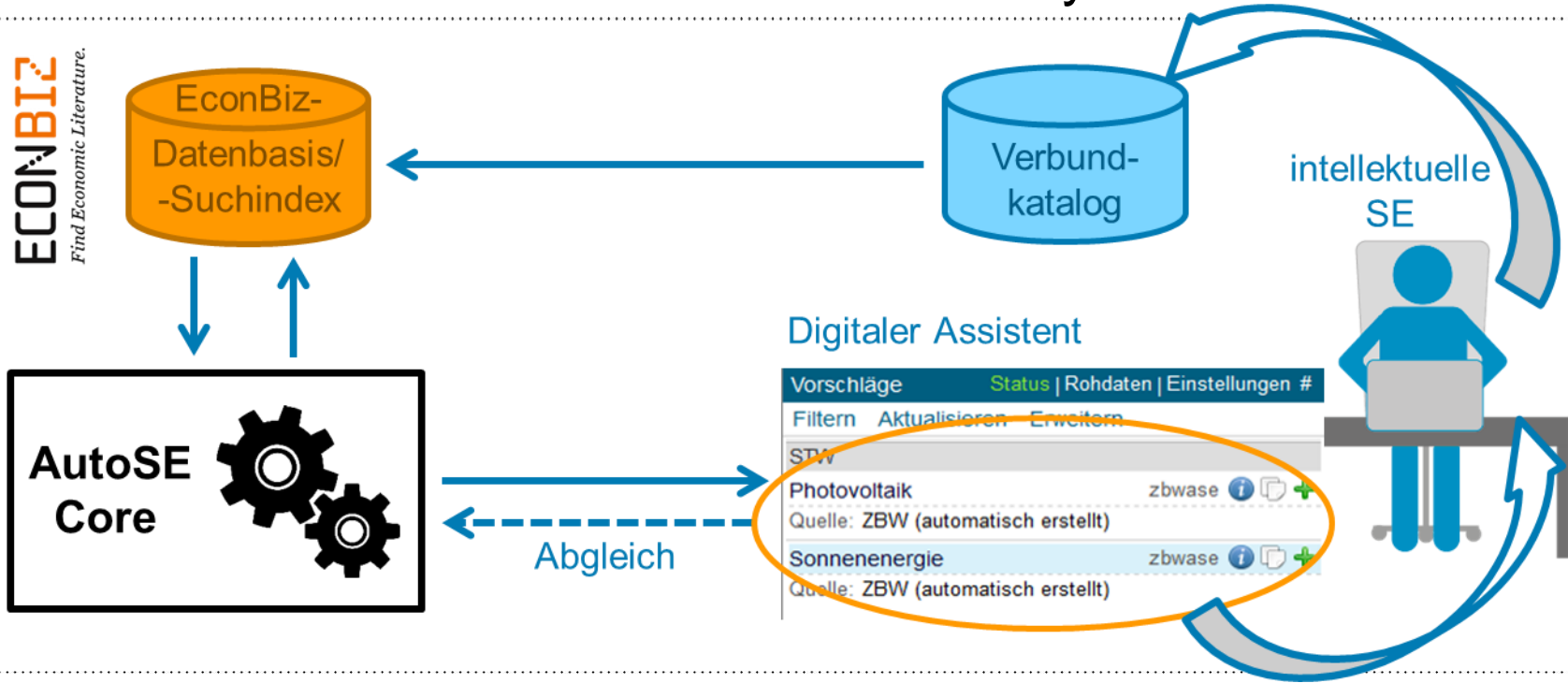
# Informationen zu Daten aus der Quelle ‚zbwase‘

---

...

- die einzelnen Vorschläge unterliegen einer **Konfidenzwertschwelle** von 0.23
- außerdem wenden wir aktuell noch zwei **Filter** an (ad-hoc-Anforderung aus dem Fachbereich), die bei den zwei häufigsten in den Trainingsdaten vergebenen Deskriptoren („Theorie“ und „USA“) gegensteuern
- die Vorschläge, die das erfüllen, **legen wir in einem Speicher ab**
- dort **kann der DA-3 sie sich über eine API abholen**, wenn im DA-3 eine Publikation aufgerufen wird und ‚zbwase‘ als Vorschlagsquelle ausgewählt ist

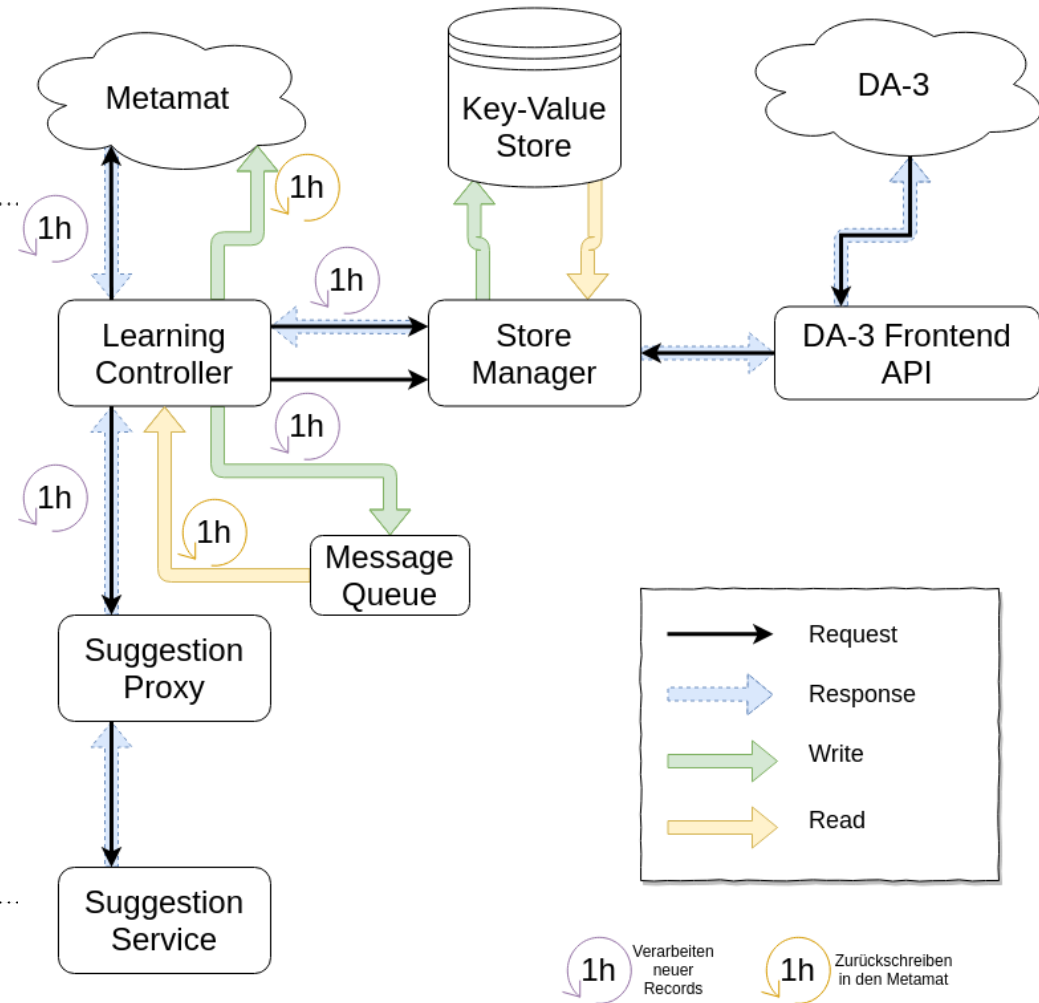
# Interaktion der verschiedenen Produktivsysteme





# Unter der Haube ...

- Metamat ~ EconBiz-Datenbasis
- Suggestion Service: da sitzen die trainierten Backends und verschlagworten
- Key-Value-Store: hier werden Verschlagwortungen abgelegt, die die Qualitätsschwellen und -filter passiert haben



---

# Live-Demonstration im Produktivsystem des DA-3 für K10plus

Einstellungen übernehmen

Quellen

- bne (Biblioteca Nacional de España)
- bnf (Bibliothèque nationale de France)
- bvb (Bibliotheksverbund Bayern)
- dnb (Deutsche Nationalbibliothek)
- hbz (Hochschulbibliothekszenrum NRW)
- hebis (Hessisches BibliotheksInformationsSystem)
- loc (Library of Congress)
- obv (Österreichischer Bibliothekenverbund)
- rebiun (Red de Bibliotecas Universitarias Españolas)
- slsp (Swiss Library Service Platform)
- sudoc (Système Universitaire de Documentation)
- zbwase (ZBW (automatisch erstellt))**

Vorschlaggeneratoren

- STW zu GND (Exakt)
- STW zu GND (Weit)

- bvb-1
- hebis-1
- rebiun-1
- slsp-1
- slsp-2
- zbwase-1**

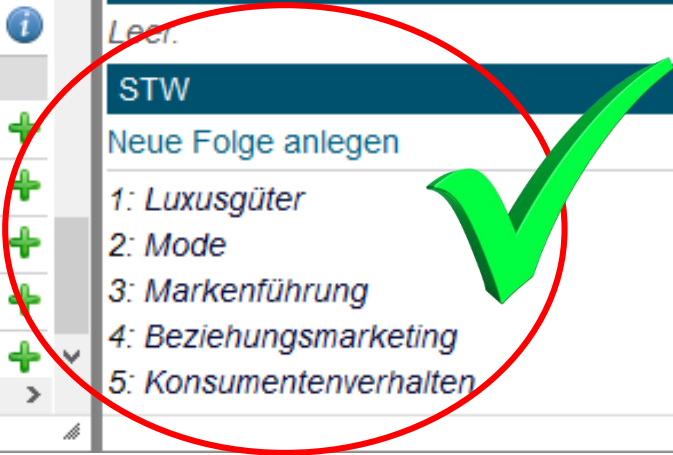
01147cam a2200265 a 4500  
 001 1032536500  
 650 7 \$a Luxusgüter  
 \$0 (DE-2867)10302-6  
 \$2 stw  
 650 7 \$a Mode  
 \$0 (DE-2867)18295-0  
 \$2 stw  
 650 7 \$a Konsumentenverhalten  
 \$0 (DE-2867)10300-3  
 \$2 stw  
 650 7 \$a Beziehungsmarketing  
 \$0 (DE-2867)19133-0  
 \$2 stw  
 650 7 \$a Markenführung  
 \$0 (DE-2867)19129-5  
 \$2 stw

Kurztitel	#
Nummer: 1032536500	
Titel: <b>Signature experience</b> : art and science of customer engagement for fashion and luxury companies / edited by Stefania Saviolo	
Personen: Saviolo, Stefania [HerausgeberIn]	
Ausgabe: First edition	
Publ.: Milano : BUP, August 2018	
ISBN: 978-88-99902-31-5, 978-88-85486-59-1	
Sprache: Englisch [text]	
Weitere Daten	

Vorschläge	Status	Rohdaten	Einstellungen	#
Filtern	Aktualisieren	Erweitern		
Stammkunde [Sach]	@stw-related			
Verbraucheranalyse [Sach]	@stw-related			
Verbraucherforschung [Sach]	@stw-related			
Verbraucherverhalten [Sach]	@stw-exact			
LCSH				
Customer relations – Management.	slsp			
STW				
Beziehungsmarketing	zbwase			
Konsumentenverhalten	zbwase			
Luxusgüter	zbwase			
Markenführung	zbwase			
Mode	zbwase			

Vorschläge	Status	Rohdaten	Einstellungen #
Filtern Aktualisieren Erweitern			
Stammkunde [Sach]	@stw-related		
Verbraucheranalyse [Sach]	@stw-related		
Verbraucherforschung [Sach]	@stw-related		
Verbraucherverhalten [Sach]	@stw-exact		
LCSH			
Customer relations – Management.	slsp		
STW			
Beziehungsmarketing	zbwase		
Konsumentenverhalten	zbwase		
Luxusgüter	zbwase		
Markenführung	zbwase		
Mode	zbwase		

Erschließung	Prüfen	Rohdaten	Einstellungen #
Leer.			
SBBRC			
Leer.			
Sigel			
Leer.			
SSG/FID			
Leer.			
TUM-Systematik			
Leer.			
STW			
Neue Folge anlegen			
1: Luxusgüter			
2: Mode			
3: Markenführung			
4: Beziehungsmarketing			
5: Konsumentenverhalten			



Kurztitel	#
Nummer: 1066736987 	
Titel: <b>Agency, human dignity, and subjective well-being</b> / Daniel A. Hojman, Álvaro Miranda	
In: World development 101(2018) vom: Jan., Seite 1-15 Amsterdam : Elsevier Science, 1973	
Personen: Hojman, Daniel* [VerfasserIn] Miranda, Álvaro [VerfasserIn]	
Publ.: January 2018	
Sprache: Englisch [text]	
<b>Weitere Daten</b>	
DOI: <a href="https://dx.doi.org/10.1016/j.worlddev.2017.07.029">https://dx.doi.org/10.1016/j.worlddev.2017.07.029</a>	

Vorschläge	Status	Rohdaten	Einstellungen	#
Filtern	Aktualisieren	Erweitern		
STW				
Arbeitszufriedenheit	zbwase			
Lebensqualität	zbwase			
Zufriedenheit	zbwase			
GND				
Arbeitszufriedenheit [Sach]	@stw-exact			
Lebensqualität [Sach]	@stw-exact			
Messung [Sach]	@stw-exact			
Zufriedenheit [Sach]	@stw-exact			

Filtern Aktualisieren Erweitern

STW

Arbeitszufriedenheit      zbwase   


Lebensqualität      zbwase   

Zufriedenheit      zbwase   

GND

Arbeitszufriedenheit [Sach]      @stw-exact   

Lebensqualität [Sach]      @stw-exact   

Messung [Sach]      @stw-exact   

Zufriedenheit [Sach]      @stw-exact   

STW

Neue Folge anlegen

1: *Arbeitszufriedenheit*   

2: *Zufriedenheit*

3: *Lebensqualität*

4: *Messung*

5: *Chile*

Autoren-Keywords (ZBW)

*subjective well-being* 

*agency* 

*human dignity* 

*capabilities* 

*development* 

GND

Neue Folge anlegen

Leer.

*Zulkhibri, Turkhan Al*  
*editors*

Vorschläge **Status** | Rohdaten | Einstellungen #

Filtern Aktualisieren Erweitern

Finance, Public-Islamic countries. loc 

Macroeconomics. hbz, loc, slsp 

RVK

Bank-, Geld- und Kreditwesen einzelner Länder (QK 010) bvb   




STW

Finanzmarkt zbwase   

Finanzwissenschaft @gnd-exact   

Islam @gnd-exact, zbwase   

Islamische Staaten zbwase   

Islamisches Finanzsystem zbwase   

Makroökonomik zbwase   

Risiko zbwase   

Theorie zbwase   

Erschließung Prüfen | Rohd

SSG/FID

Leer.

TUM-Systematik

Leer.

STW

Neue Folge anlegen

1: *Islamisches Finanzsystem*

2: *Risikomanagement*

3: *Finanzmarktregulierung*

4: *Bankenregulierung*

5: *Finanztechnologie*

6: *Wirtschaftliche Sicherheit*

7: *Wirtschaftliche Instabilität*

8: *Soziale Stabilität*

9: *Entwicklung*



10: *Islamische Staaten*

11: *Welt*



**Choice Inconsistencies Among****the Elderly** : Evidence from Plan


Choice in the Medicare Part D

Program / *Jason T. Abaluck,**Jonathan Gruber*Wohlfahrtseffekt [Sach] @stw-exact   Wohlfahrtsmessung [Sach] @stw-narrow   Älterer Mensch [Sach] @stw-exact   

STW

Gesetzliche zbwase   

Krankenversicherung

Krankenversicherung zbwase   USA zbwase   Ältere Menschen zbwase   

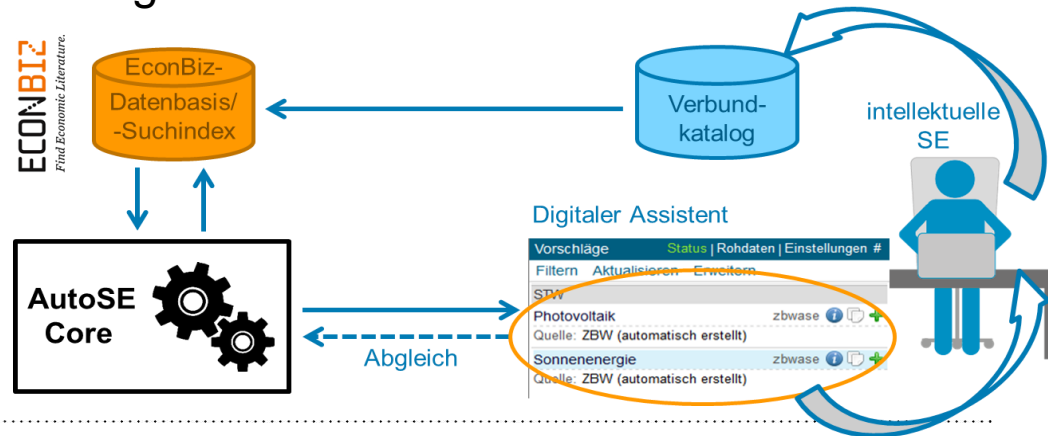
STW

Neue Folge anlegen

1: *Krankenversicherung*2: *Ältere Menschen*3: *Offenbarte Präferenzen*4: *Gesundheitsreform*5: *Wohlfahrtsanalyse*6: *USA*7: *Medicaid*8: *2005-2006 [Zeit]*

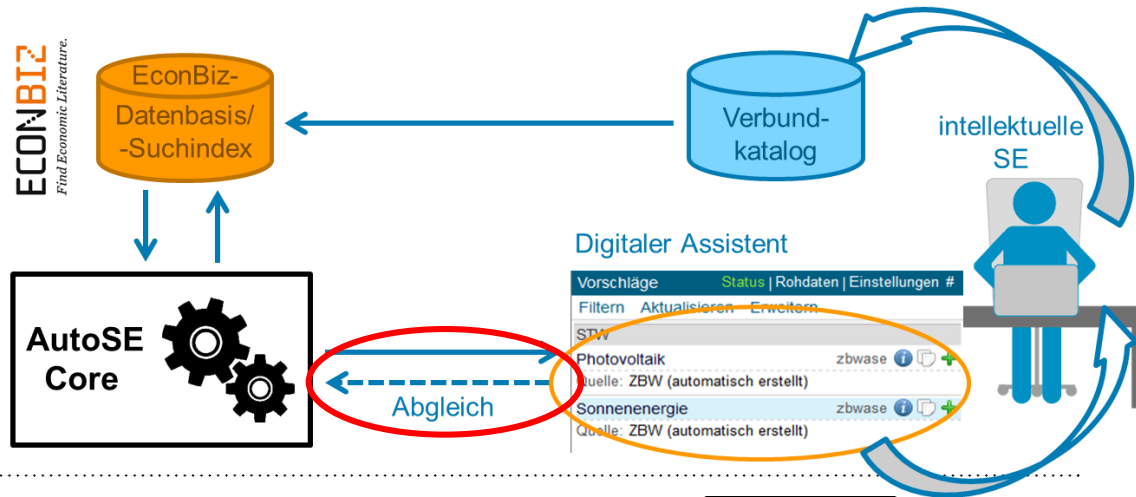
# Unterschiedliche Verwendungszwecke

- **Merke:** Daten, die wir direkt in unsere EconBiz-DB zurückspeichern, werden weiteren Regeln zur Sicherung der Qualität unterworfen – im Gegensatz zu den im DA-3 angezeigten – im DA-3: Vorschlagscharakter!
- Die ZBW plant, AutoSE-Daten zukünftig auch in ein **gesondertes (!)** Feld in die WinIBW einspielen zu lassen, samt Provenienzdaten (Feldschema ist entworfen, wir arbeiten an einem Pilot-Datenbestand zum Testen)



# Performance-Berechnungen über Feedback aus dem DA-3

- Wir greifen neue intellektuelle Sacherschließung ab und vergleichen unsere automatisierten Erzeugnisse damit (Berechnung F1-Wert)
- monatlich aggregierter F1-Wert für aktuelles Backend: 0.55 bis 0.58
- Überlegungen zu einem Open-Source-Modul für abgestufte Bewertungen, das sich an den DA (und andere Systeme) anschließen lässt



# Herzlichen Dank!

---

## Weitere Vorträge und Publikationen zu AutoSE:

siehe Hinweise unten auf der Seite

<https://www.zbw.eu/de/ueber-uns/arbeitsschwerpunkte/automatisierung-der-erschliessung/>

Kontakt: {a.kasprzik,m.fuerneisen,c.bartz,autose}@zbw.eu

---