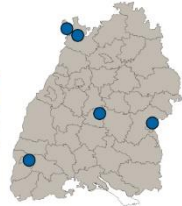




NETZWERK
SELTENE ERKRANKUNGEN
BADEN-WÜRTTEMBERG
KOMPETENZZENTRUM



ulm university universität
uulm

Hochschule Ulm
University of
Applied Sciences



Towards the Automated Generation of Expert Profiles for Rare Diseases through Bibliometric Analysis

Automatische Generierung von Expertenprofilen zu Seltenen Erkrankungen

Andreas Pflugrad^{a,b,c}, Karin Jurkat-Rott^{a,b}, Frank Lehmann-Horn^{a,b} and Jochen Bernauer^{a,c}

a Zentrum für Seltene Erkrankungen Ulm, Kompetenzzentrum Seltene Erkrankungen Baden-Württemberg

b Division of Neurophysiology, Universität Ulm

c Institut für Informatik, Hochschule Ulm




23.05.2014

eHEALTH
2014
www.eHealth2014.at

1. Hintergrund

- Seltene Erkrankungen (SE)
 - *Prävalenz: < 5 : 10.000*
 - *~ 7.000 SE bisher beschrieben*
- Seltene Experten
- Viele Patienten
 - *> 30.000.000 Betroffene in Europa*

Informationsmöglichkeiten

- Lotsen in den Zentren für Seltene Erkrankungen
- Expertenregister in Internetportalen, z.B.

- Unabhängige Patientenberatung
- Selbsthilfegruppen oder SE-Communities

2. Problem: neue Experten identifizieren

Traditionelles Vorgehen:

- Datenerhebung an Kliniken per Fragebogen
- Aktualisierung aufwendig, oftmals nur reaktiv

Alternativer Ansatz:

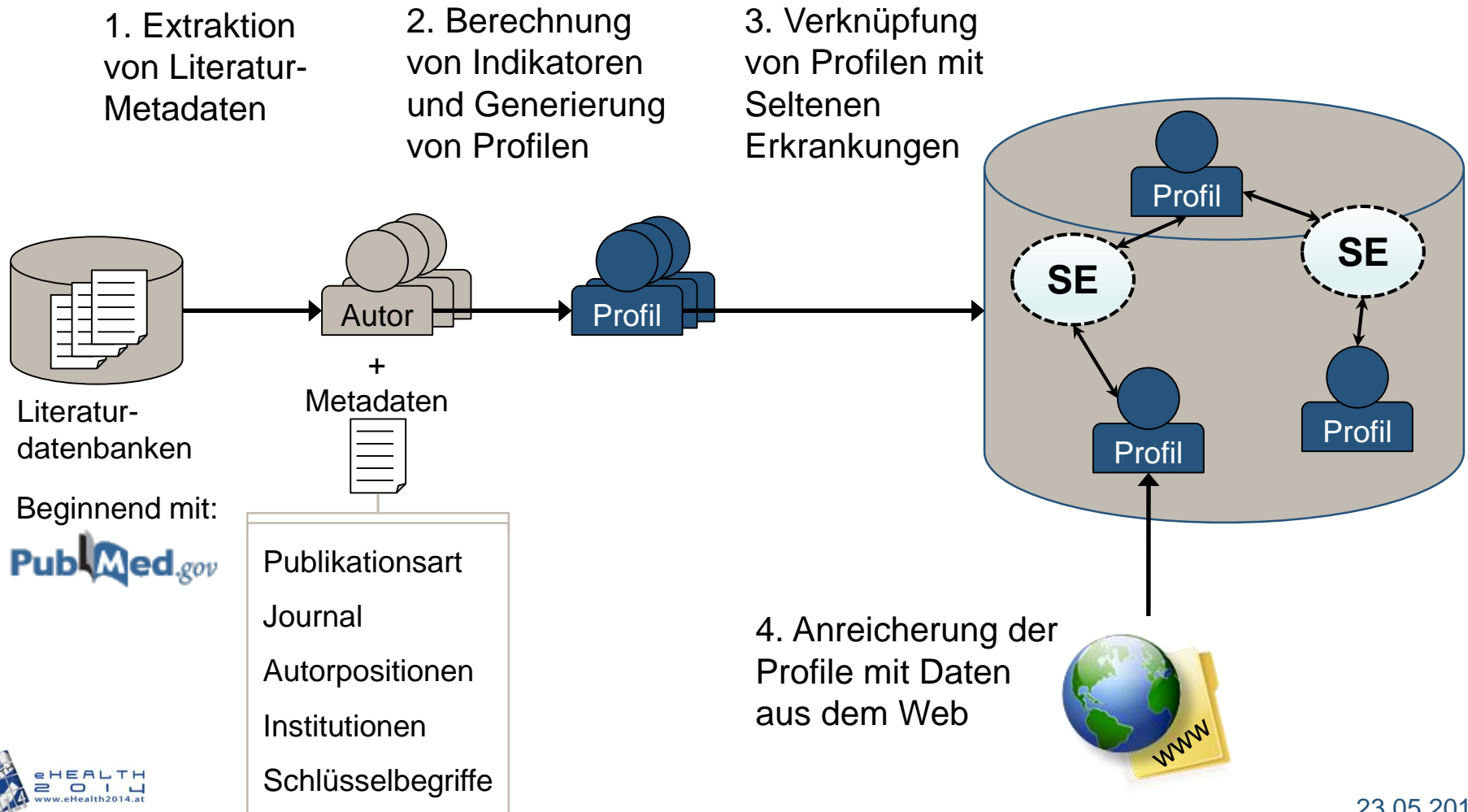
- Automatische Generierung von Expertenprofilen aus bibliografischen Daten
- Kontinuierliche Aktualisierung

3. Grundgedanke

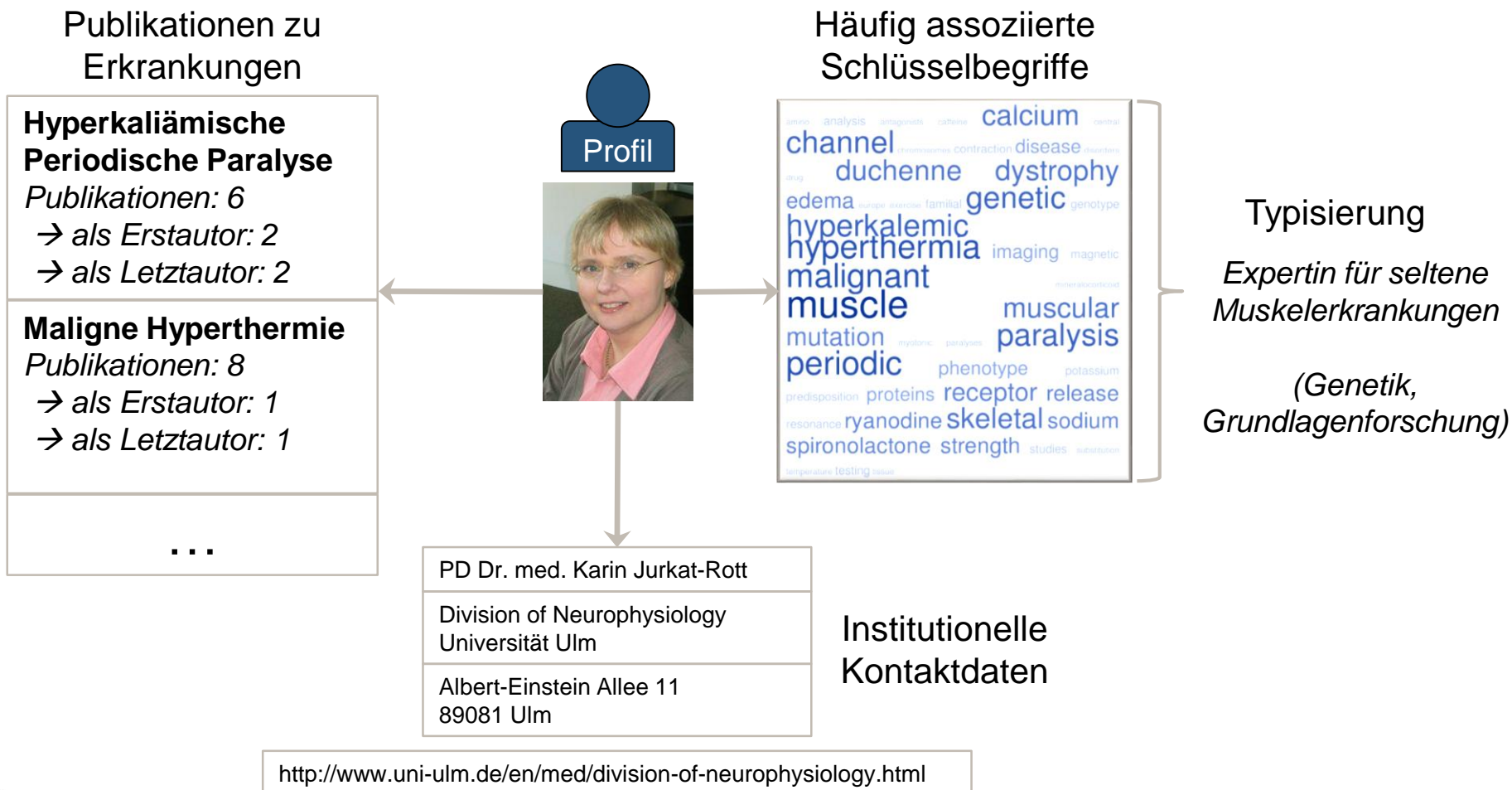
- Autoren von Publikationen zu SE \implies potentielle Experten
- Identifizierbar mittels Auswertung von Literaturdatenbanken
 - MEDLINE
 - Cochrane Library
 - GeneReviews
- Indikatoren zur Typisierung und Validierung der Expertise
 - Anzahl/Art der Publikationen
 - Assoziierte Medical Subject Headings (MeSH)
- Periodische Auswertungen ermöglichen eine zeitnahe Aktualisierung

Ziel: Unterstützung von Patienten, Ärzten und Lotsen bei der Suche nach Experten für Seltene Erkrankungen

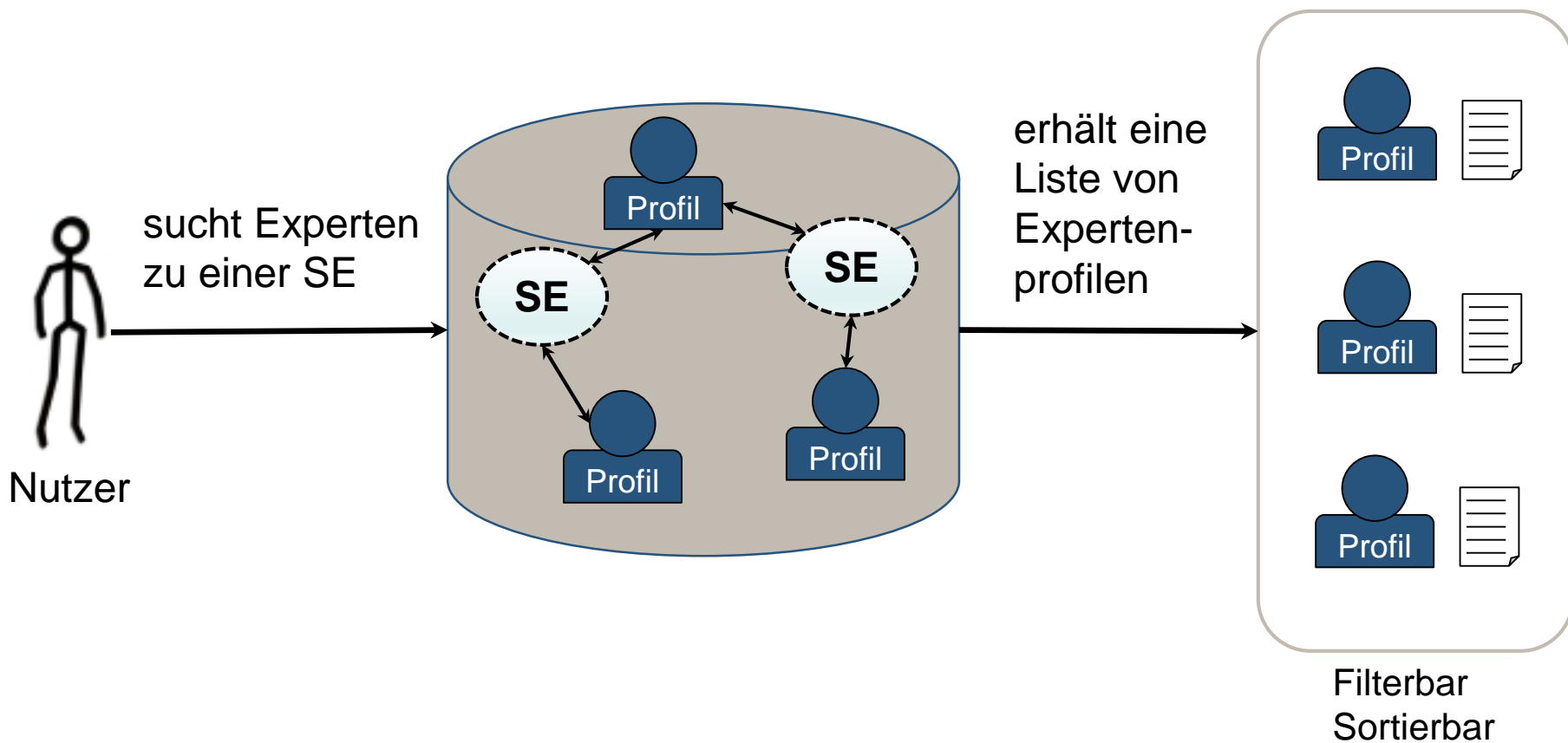
4. Methode – Generierung von Expertenprofilen



5. Expertenprofil – Beispiel

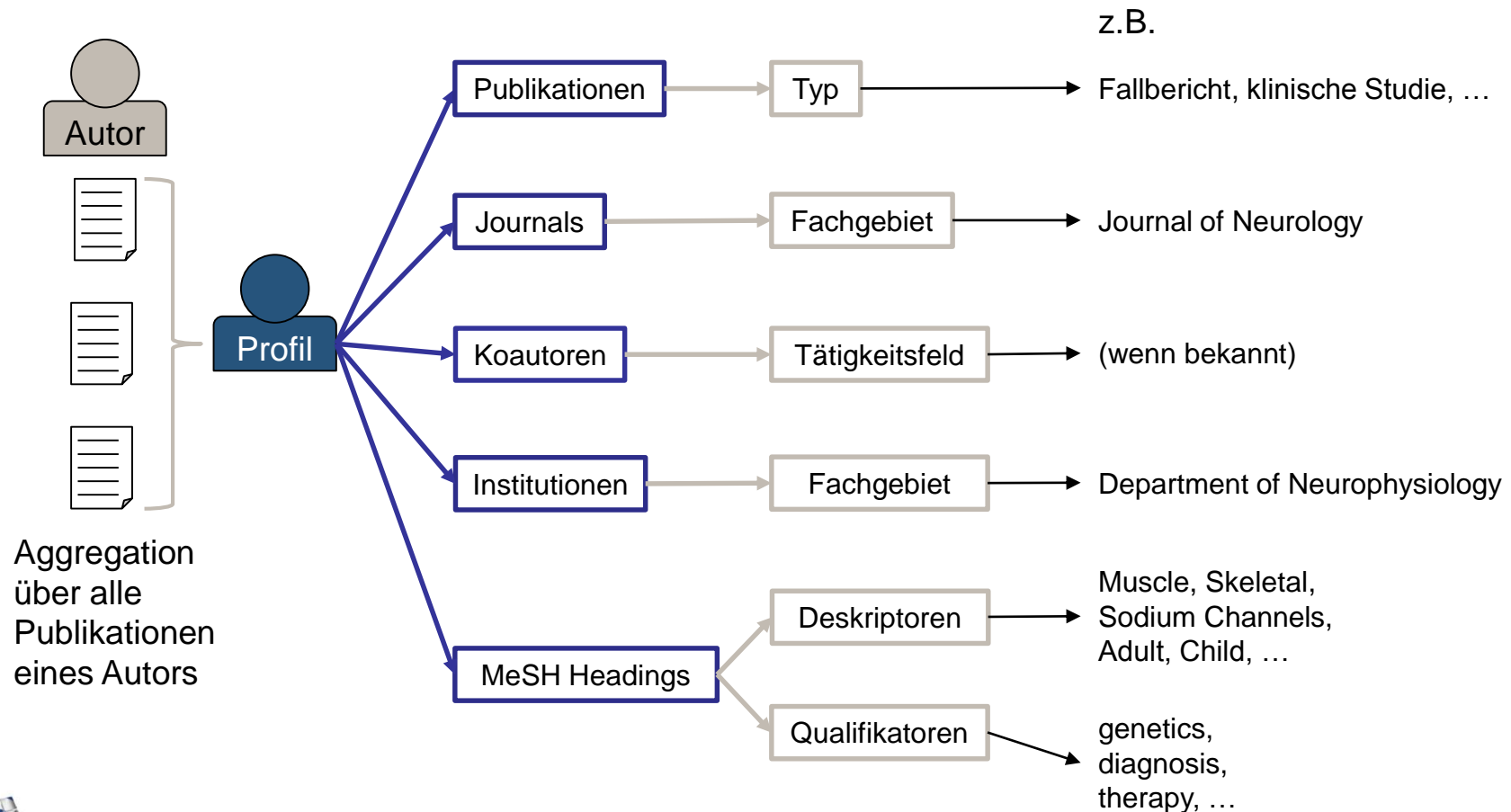


6. Ziel – Information für Hilfesuchende



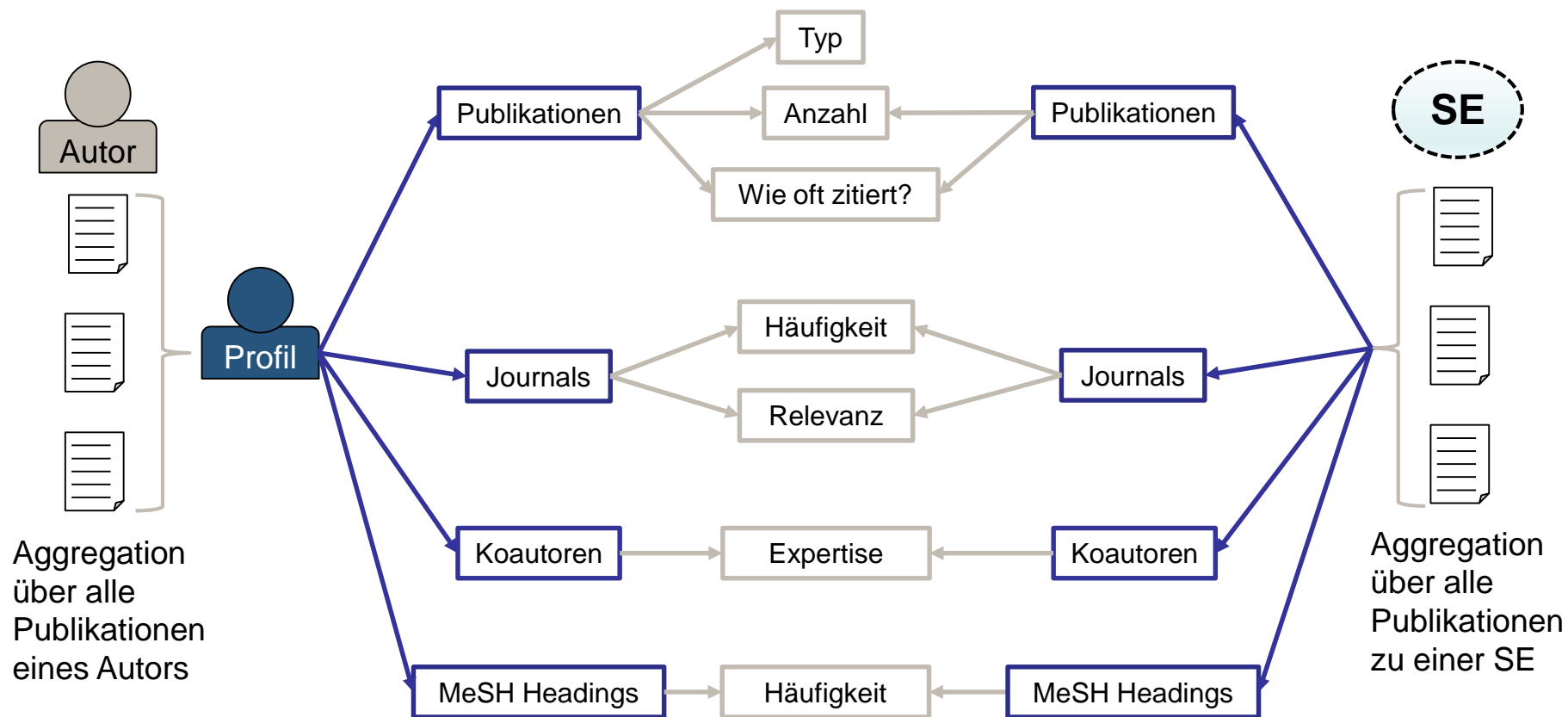
7. Umsetzung – Typisierung eines Experten

- Indikatoren für das Fachgebiet und Tätigkeitsfeld eines Autors



8. Umsetzung – Validierung eines Experten

- Indikatoren für die Validierung der Expertise eines Autors



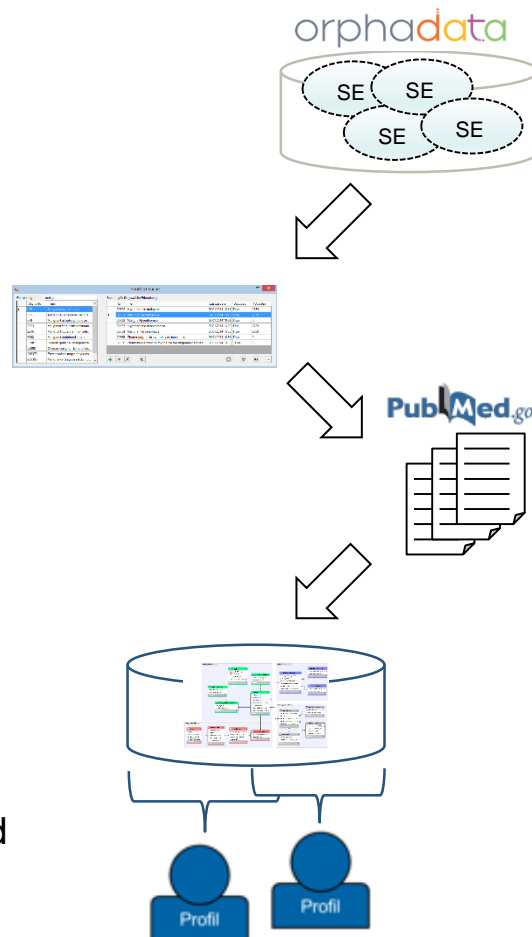
9. Ergebnisse – entwickeltes System

Suchapplikation

- Automatische Extraktion von Publikationsdaten aus PubMed zu den Begriffen des Thesaurus

Profilgenerierung

- Beispielprofile mittels einfacher Aggregation
- Kriterien: Nachname und Initialen übereinstimmend
- Erwartete Verzerrungen durch Namensambiguität



SE-Thesaurus

- Liste von Erkrankungen aus Orphadata
- Erweiterung mit Begriffen aus MeSH und UMLS¹

Restriktive Heuristik

- Begriff als Titel oder „Major Topic“
- Keine Interpretation durch PubMed
- Ziel: hohe Spezifität

¹Unified Medical Language System

10. Ergebnisse – Expertenprofil

Author Profile - Hermann Heimpel				
Top 3 Disease Entities by Publication Count	First/Middle/Last Author (Total)			
Chronic myeloid leukemia	1	27	16	(44)
Congenital dyserythropoietic anemia	12	9	8	(29)
Acute myeloid leukemia	2	5	6	(13)
Top 3 MeSH-Descriptors				
Leukemia, Myelogenous, Chronic, BCR-ABL Positive	90			
Anemia, Dyserythropoietic, Congenital	51			
Bone Marrow Transplantation	37			
Top 3 MeSH-Qualifiers		Top 3 Journals		
therapeutic use	115	Ann. Hematol.	14	
pathology	95	Blood	14	
genetics	75	Blut	12	

✓ Typisierung als Experte für seltene Anämien

✓ Validierung der Expertise für eine bestimmte SE durch Betrachtung der Erkrankungsdaten



Prof. Dr. med.
H. Heimpel

„Ich hab‘ aber noch mehr Artikel.“

➔ Notwendigkeit verbesserter Suchstrategien

Disease Data - Congenital Dyserythropoietic Anaemia

Overall Article Count: 688

Top 3 Authors by Publication Count		Top 3 Journals	
A. Iolascon	36	Br. J. Haematol.	295
SN. Wickramasinghe	33	Blood	218
H. Heimpel	29	Eur. J. Haematol.	150
Top 3 MeSH-Descriptors*		Top 3 MeSH-Qualifiers	
Cladribine	727	genetics	3732
Erythroblasts	559	pathology	2528
Bone Marrow	527	blood	2043

11. Rückschlüsse und aktuelle Forschung

- Schwächen der PubMed-Suche
 - Nicht alle relevanten Publikationen zu einer SE werden gefunden
 - Zu vielen SE werden überhaupt keine Publikationen gefunden
 - ➔ Evaluation verbesserter Suchstrategien

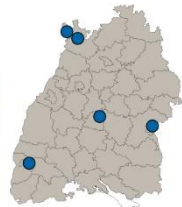
- Schwächen der Profilgenerierung
 - Erwartungsgemäße Verzerrungen durch Namensambiguität
 - ➔ Entwicklung effektiver Methoden zur Namensdesambiguierung

12. Ausblick

- Geplante Nutzung
 - Bereitstellung der Profile für hilfeschende Patienten, Ärzte und Lotsen
 - Bereitstellung der Profile für bestehende Expertenregister zur Ergänzung der vorhandenen Datenbestände

- Geplante Erweiterungen
 - Anreicherung der Expertenprofile mit (institutionellen) Kontaktdaten
 - Reduzierung der Abhängigkeit von der Publikationstätigkeit durch Einsatz von Web-Crawling Mechanismen

Vielen Dank für Ihre Aufmerksamkeit!



ulm university universität
uulm



Kontakt:

Andreas Pflugrad
Division of Neurophysiology
Universität Ulm
Albert-Einstein-Allee 11
89081 Ulm
andreas.pflugrad@uni-ulm.de

www.seltene-erkrankungen.info/Expertensuche

Referenzen

- Orphanet: an online rare disease and orphan drug data base. Copyright, INSERM 1997. Available on <http://www.orpha.net>
- PubMed. Bethesda (MD): National Center for Biotechnology Information (US); 2005-. PubMed. Available from: <http://www.ncbi.nlm.nih.gov/pubmed>
- Tang, Jie, et al. "ArnetMiner: An Expertise Oriented Search System for Web Community." *Semantic Web Challenge*. 2007.
- Torvik, Vetle I., and Neil R. Smalheiser. "Author name disambiguation in MEDLINE." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 3.3 (2009): 11.