

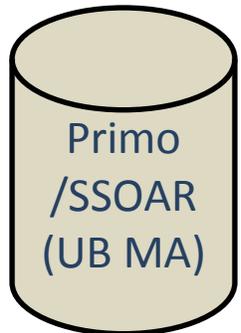
# „Es wächst zusammen, was zusammen gehört“ – Vernetzung von Titel- und Forschungsdaten

Dominique Ritze  
UB Mannheim

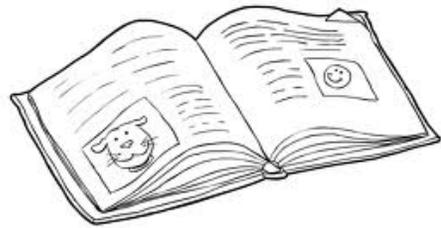
# Motivation



Suche



Resultat



Ref



Resultat



Suche



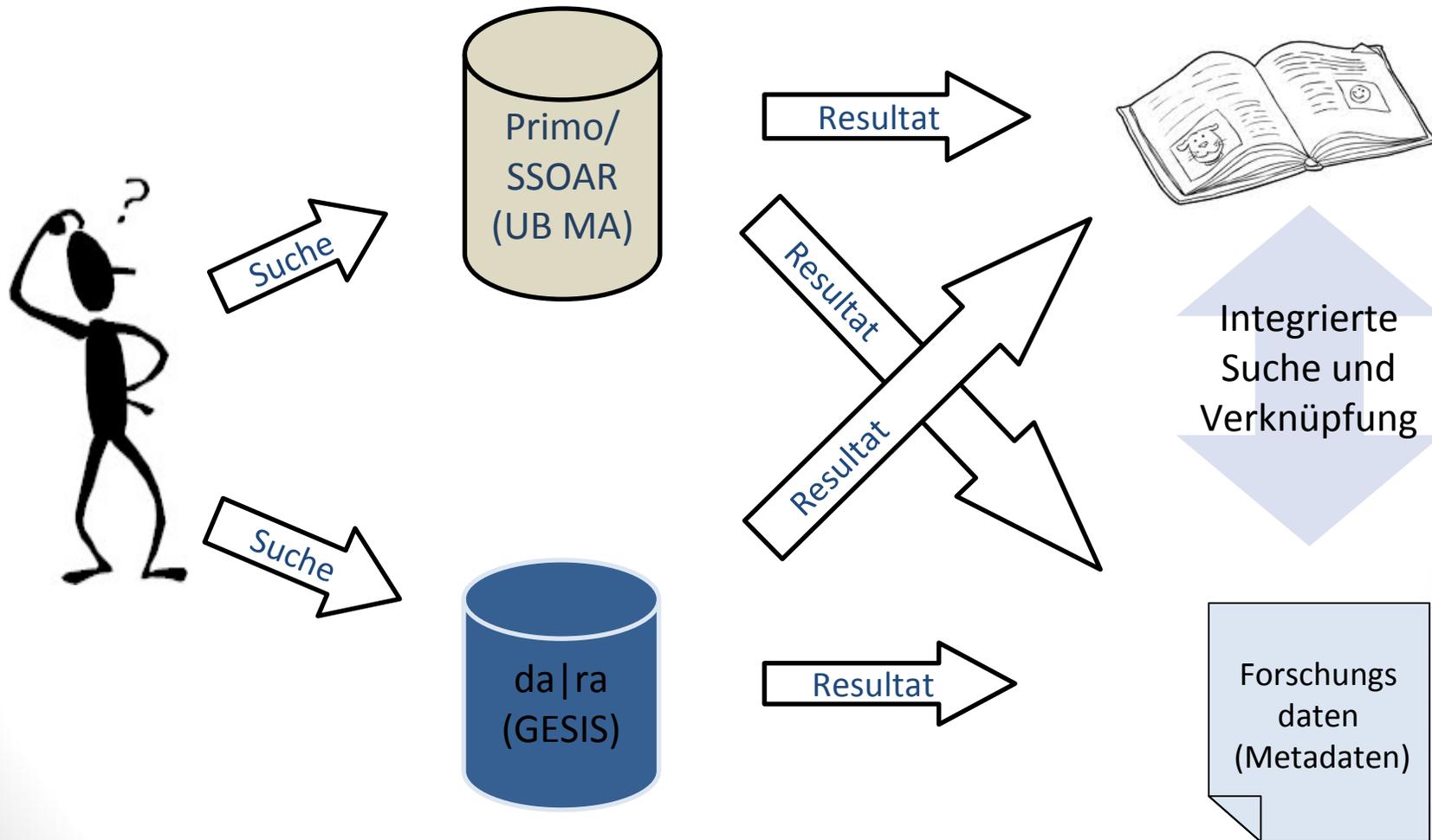
# Gliederung

- Motivation
- InFoLiS-Projekt
- Verfahren zum Finden von Referenzen
- Integration der Daten und Referenzen
- Fazit

# InFoLiS

- InFoLiS: Integration von Forschungsdaten und Literatur in den Sozialwissenschaften
- Projektzeitraum: August 2011- August 2013
- Gefördert durch die DFG  Deutsche Forschungsgemeinschaft

# Ziel



# Vorgehensweise

- Integration
  - Metadaten Schemata analysieren
  - In ein Format entsprechend dem Resource Discovery System (RDS) überführen
  - Laden der Daten
- Verlinkung
  - Daten analysieren
  - Verfahren entwickeln um Referenzen zu erkennen
  - Anwendung auf die eigenen Daten
  - Integration der Referenzen in das RDS

# Referenzierungsarten

erfolgt die Darstellung und Diskussion der empirischen Ergebnisse. Hierfür werden die Daten des **Sozio-oekonomischen Panels (SOEP)** aus den Jahren 1990 und 2003 verwendet und für beide Zeitpunkte werden die Einflussfaktoren mittels linearer Regressionsmodelle geschätzt.

---

1 Herangezogen wurden außerdem **Allbus, Allensbacher Erhebungen, Eurobarometer, International Social Survey Program, International Social Justice Project, Sozio-ökonomisches Panel, World Values Survey.**

---

*Tabelle 1: Bevölkerungsvorausberechnung für Deutschland nach Altersgruppen – Anteile in Prozent*

*(Datenbasis: 10. Bevölkerungsvorausberechnung des Statistischen Bundesamtes, Variante 5)*

---

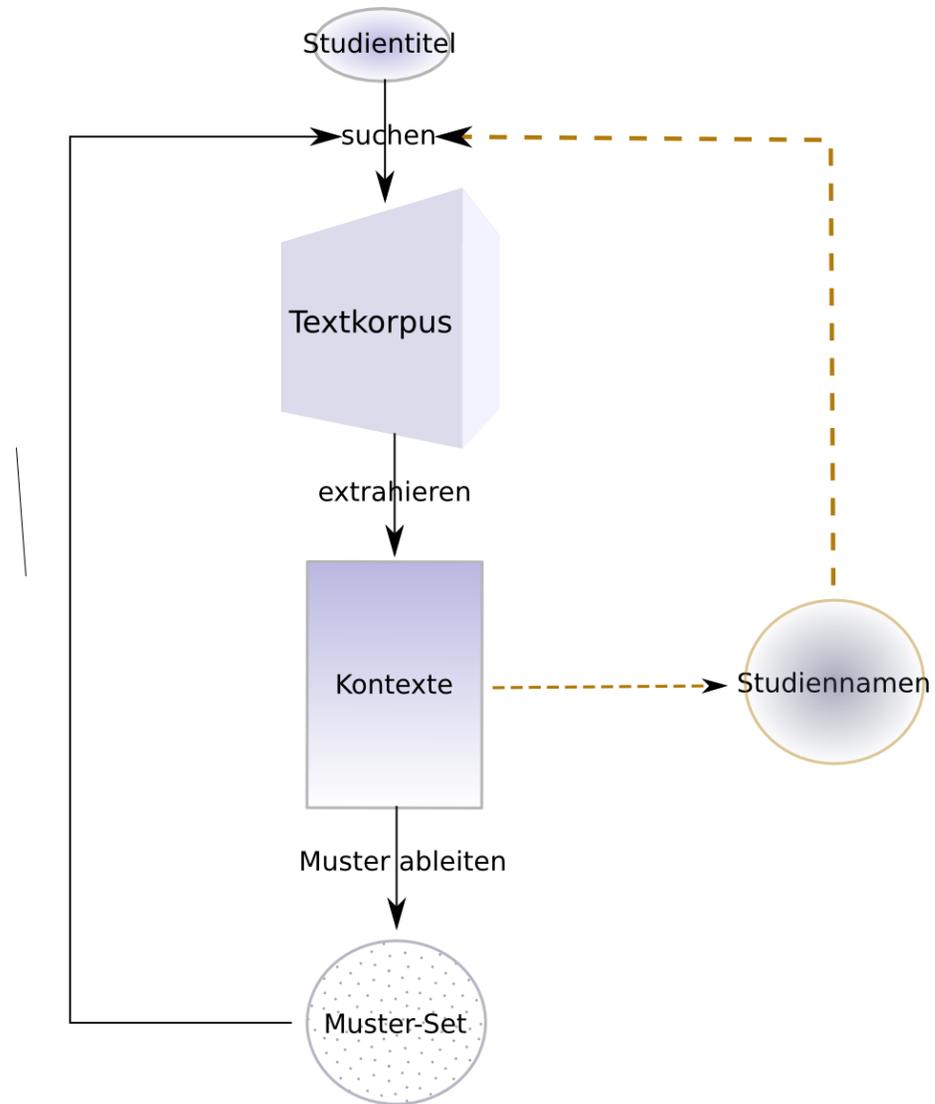
*Tabelle 3: Stichprobe der Untersuchung in den Jahren 2003 und 2004 sowie Größe der Stichprobe, mit gültigen Daten aus beiden Erhebungen*

*(Quelle: **Ditton u.a. 2005a**)*

# Wie findet man Referenzen?

- Suchen nach Studiennamen ineffizient
  - Studiennamen wie „Gesundheit“ oder „Wirtschaft und Ethik“
  - Oft werden Synonyme und Abkürzungen benutzt
  - Keine vollständige Liste mit Studiennamen
- Machine Learning Methoden um Muster zu erkennen in denen Referenzen typischerweise auftreten
- Bootstrapping um das “sparse data“ Problem zu lösen

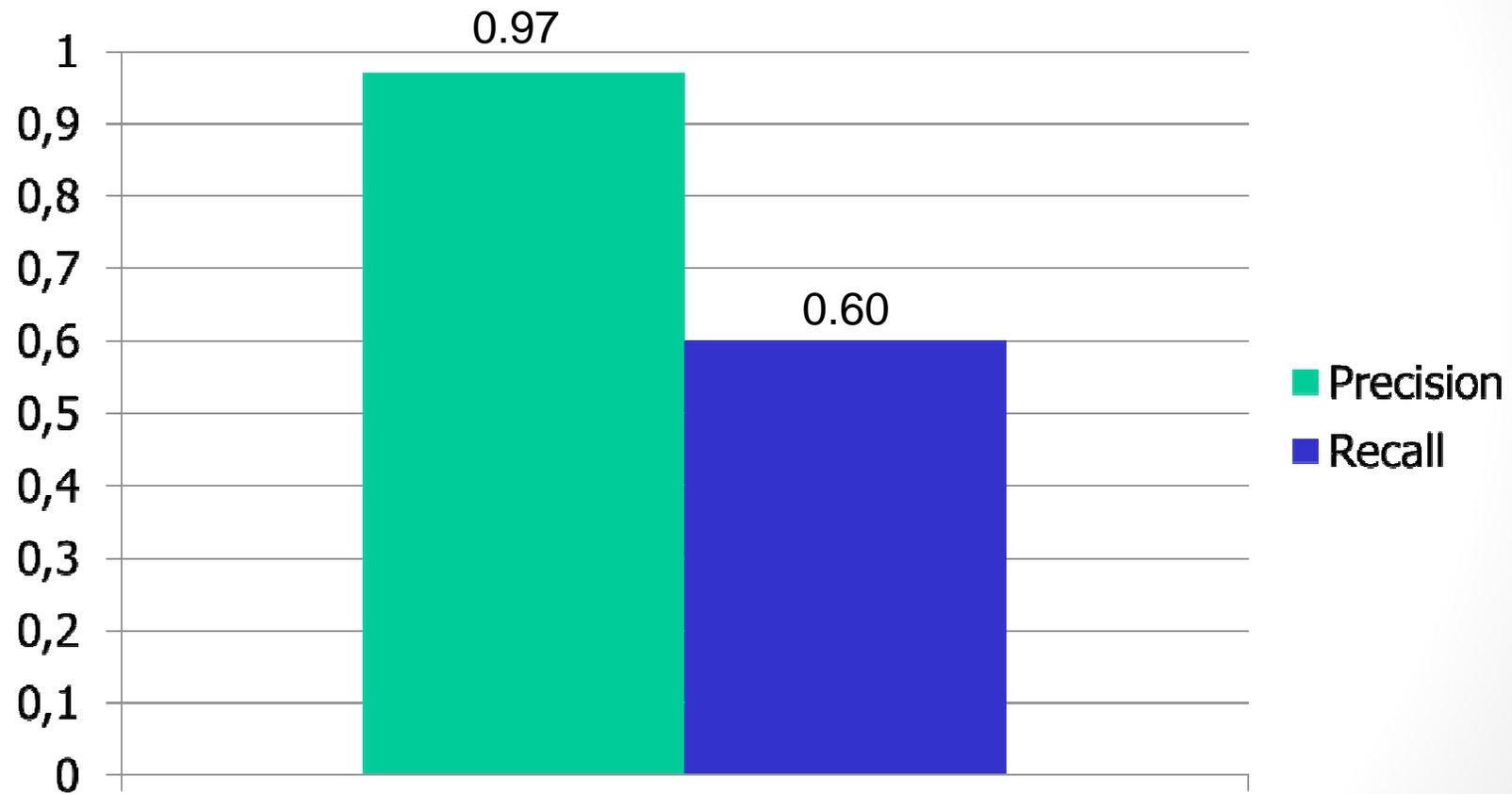
# Algorithmus



# Experiment

- SSOAR Dokumente mit dem Schlagwort „empirisch-quantitativ“
- 259 Dokumente auf deutsch
- Unabhängig vom ersten Studiennamen
  
- Vorverarbeitung
  - Konvertierung Pdf → Txt
  - Automatische Verbesserung
  - Entfernen von Bibliographien

# Ergebnis



# Anwendung

- Studien: da|ra
- Publikationen: SSOAR
- Gefundene Studiennamen müssen mit Studien in da|ra gematcht werden
  
- Darstellung der Referenzen in RDF mit der CiTO Ontologie
- Veröffentlichung als LOD mit SPARQL-Endpoint

# Verknüpfungen integrieren



# Integration von Studien

Books & more **GESIS**

employment and wages search

that contain my query words  anywhere in the record

1 Results for **MAN GESIS**

sorted by: Relevance

1



resdata

**Employment and Wages in German Industries between 1888 and 1954**  
**Beschäftigung und Löhne der deutschen Industriegewirtschaft 1888-1954**  
 DIVO, Frankfurt [Data Coll.]  
 GESIS - Leibniz-Institut für Sozialwissenschaften 2010-04-13

[Details](#)

1 Results for **MAN GESIS**

sorted by: Relevance

# Integration von Studien

Books & more **GESIS**

employment and wages  advanced search

that contain my query words  anywhere in the record

**Employment and Wages in German Industries between 1888 and 1954**  
 Beschäftigung und Löhne der deutschen Industrielwirtschaft 1888-1954  
 DIVO, Frankfurt [Data Coll.]  
 GESIS - Leibniz-Institut für Sozialwissenschaften 2010-04-13

**Details**

Title:	Employment and Wages in German Industries between 1888 and 1954 Beschäftigung und Löhne der deutschen Industrielwirtschaft 1888-1954
Author:	DIVO, Frankfurt [Data Coll.] GESIS - Leibniz-Institut für Sozialwissenschaften [Publ. Agent]
Subjects:	Historische Sozialforschung ; Historical Social Research
Description:	Methodik: Mündliche Befragung mit standardisiertem Fragebogen Methodik: Oral survey with standardized questionnaire Beschreibung: In der vorliegenden Studie wird versucht, Beschäftigungs- und Lohnentwicklung in Deutschland auf Grund der Unterlagen der Unfallversicherungsstatistik zu ermitteln, da die oben erwähnten Bedenken diesem Material gegenüber entfallen. Die Angaben über die Zahl der
Publisher:	GESIS - Leibniz-Institut für Sozialwissenschaften
Creation Date:	2010-04-13

**Links**  
 > Research Data Source

# Integration von Referenzen

Standorte **Details** Bestand Weitere Services

Titel:	Altersbild und Altersvorsorge der Arbeiter und Angestellten
Autor:	Ludwig von Friedeburg ; Friedrich Weltz
Schlagwörter:	Arbeiter / Altersversorgung ; Angestellter / Altersversorgung
Verknüpfte Titel:	Frankfurter Beiträge zur Soziologie : Sonderheft ; 1
Verlag:	Frankfurt am Main : Europ. Verlagsanst.
Erscheinungsjahr:	1958

**Links**

> [Find Primary Research Data](#)

**➔ Verknüpfung zu Forschungsdaten**

- 1  **Image of the Old and Provision for Old Age for Workers and Employees**  
**Altersbild und Altersvorsorge der Arbeiter und Angestellten**  
 DIVO, Frankfurt [Data Coll.]  
 GESIS - Leibniz-Institut für Sozialwissenschaften 2010-04-13

# Fazit

- Publikationen und Studien sind meist unverbunden
- Keinerlei Standards zur Referenzierung
- Bootstrapping Algorithmus kann zur Erkennung der Referenzen verwendet werden
- Representation der Referenzen als LOD
- Ziel: Integriertes Informationssystem

Vielen Dank  
für Ihre  
Aufmerksamkeit!

# InFoLiS Projekt

- Prof. Dr. York Sure-Vetter
  - GESIS – Leibniz-Institut für Sozialwissenschaften
- Prof. Dr. Heiner Stuckenschmidt
  - Lehrstuhl für Künstliche Intelligenz, Universität Mannheim
- Christian Benz
  - Universitätsbibliothek Mannheim
- InFoLiS wird gefördert durch die DFG (SU 647/2-1)

[dominique.ritze@bib.uni-mannheim.de](mailto:dominique.ritze@bib.uni-mannheim.de)