gesis
Leibniz Institute for the Social Sciences

Subject Indexing for Author Name Disambiguation

Opportunities and Challenges

Cornelia Hedeler, Andreas Oskar Kempf, Jan Steinberg

Member of the Leibniz
Leibniz Association

Outline

- Introduction
- Methods
- Subject indexing for author disambiguation
- Concrete use case at GESIS
- Experimental pre-study
- Multi-level approach:
  - Macro-level
  - Meso (intermediary/group)-level
  - Micro (individual)-level
- Conclusion
- Outlook

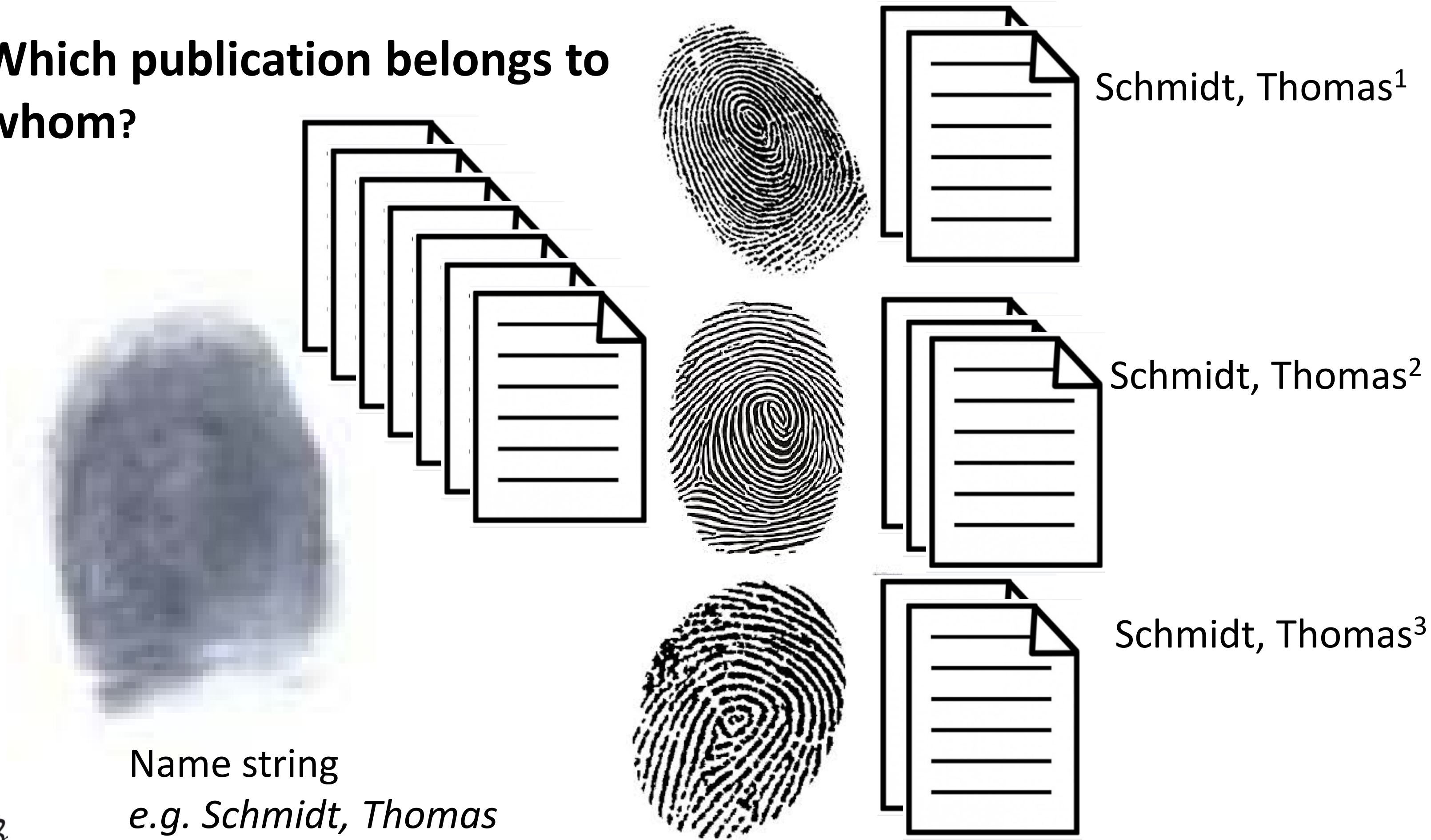
Introduction (1/2)

- Increased demand for research monitoring
- Increased requirements on reporting systems of research activities
- Increased assessment of researchers/research institutions in terms of their impact
- Research information systems have become increasingly widespread – needed for institutional assessment procedures, accreditations, university rankings etc.

> Increased demand for quality-assured data on research activities

Introduction (2/2)

Which publication belongs to whom?



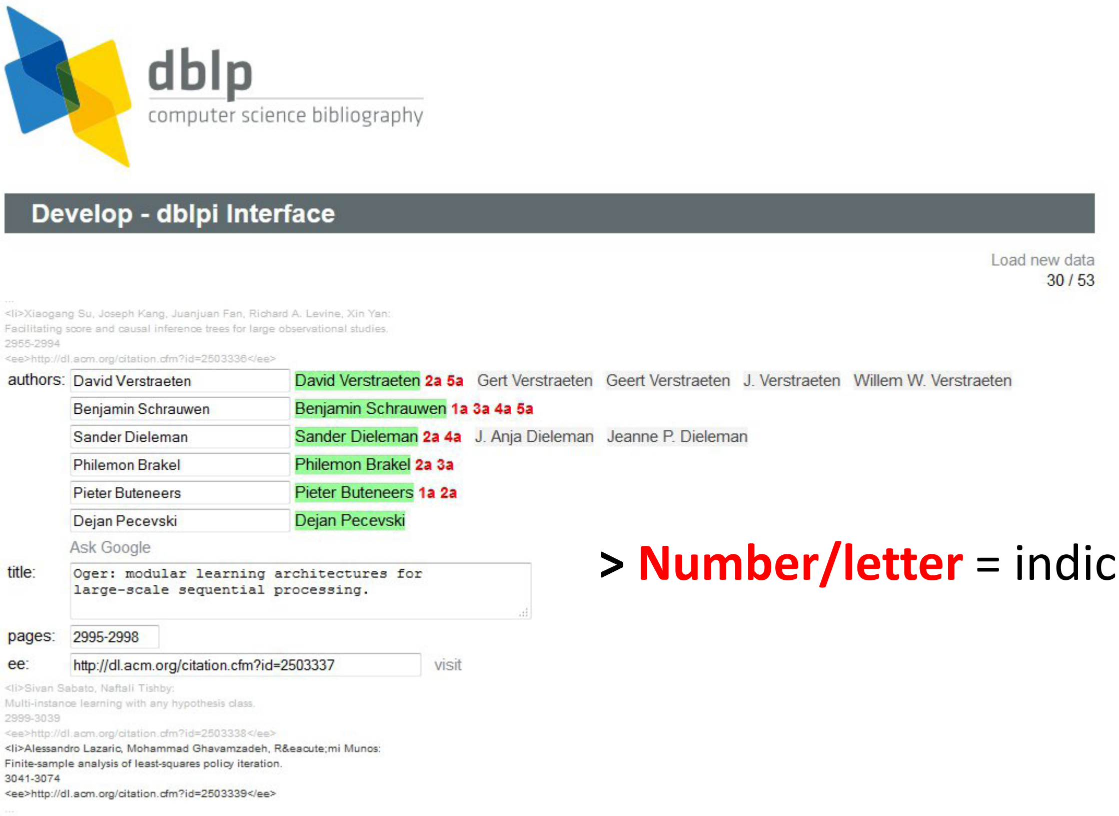
Methods for author name disambiguation

- **Author grouping methods**
  - Using clustering techniques on a given data set on the basis of publication properties (co-authors, publication year etc.) to decide whether to group/subdivide publication records assigned to a certain author name.
- **Author assignment methods**
  - Directly assigning each publication to a given author by constructing a model that represents the author (e.g. the probabilities of an author publishing an article with other co-authors, in a given venue or using specific terms in the title of a publication).

Publication properties

- Publication information:
  - Stream of the name
  - Co-author names (rate of co-authorship depends on the discipline)
  - Affiliations
  - Publication venue title
  - Publication year
  - Keywords from the title
- Additonal information:
  - Email addresses
  - Postal addresses
  - Data retrieved from the web

Example of application (1/2)



dblp
computer science bibliography

Develop - dblpi Interface

Load new data
30 / 53

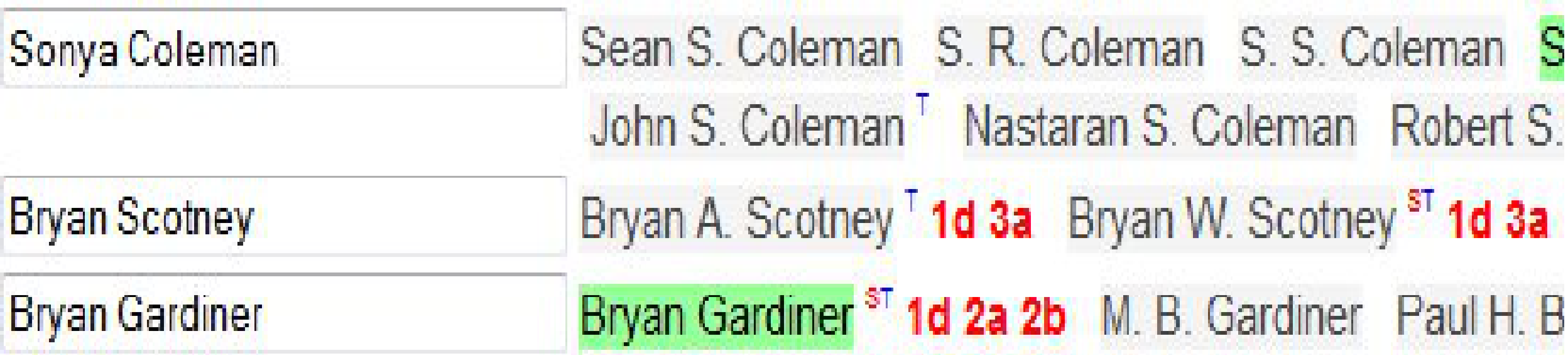
...
<li>Xiaogang Su, Joseph Kang, Juanjuan Fan, Richard A. Levine, Xin Yan:
Facilitating score and causal inference trees for large observational studies.
2955-2994
<ee>http://dl.acm.org/citation.cfm?id=2503336</ee>

authors:	David Verstraeten	David Verstraeten 2a 5a Gert Verstraeten Geert Verstraeten J. Verstraeten Willem W. Verstraeten
	Benjamin Schrauwen	Benjamin Schrauwen 1a 3a 4a 5a
	Sander Dieleman	Sander Dieleman 2a 4a J. Anja Dieleman Jeanne P. Dieleman
	Philemon Brakel	Philemon Brakel 2a 3a
	Pieter Buteneers	Pieter Buteneers 1a 2a
	Dejan Pecevski	Dejan Pecevski
	Ask Google	
title:	Oger: modular learning architectures for large-scale sequential processing.	
pages:	2995-2998	
ee:	http://dl.acm.org/citation.cfm?id=2503337	visit

<li>Sivan Sabato, Naftali Tishby:
Multi-instance learning with any hypothesis class.
2999-3039
<ee>http://dl.acm.org/citation.cfm?id=2503338</ee>
<li>Alessandro Lazaric, Mohammad Ghavamzadeh, R&eacute;mi Munos:
Finite-sample analysis of least-squares policy iteration.
3041-3074
<ee>http://dl.acm.org/citation.cfm?id=2503339</ee>
...

> **Number/letter** = indication of co-authorship

Example of application (2/2)

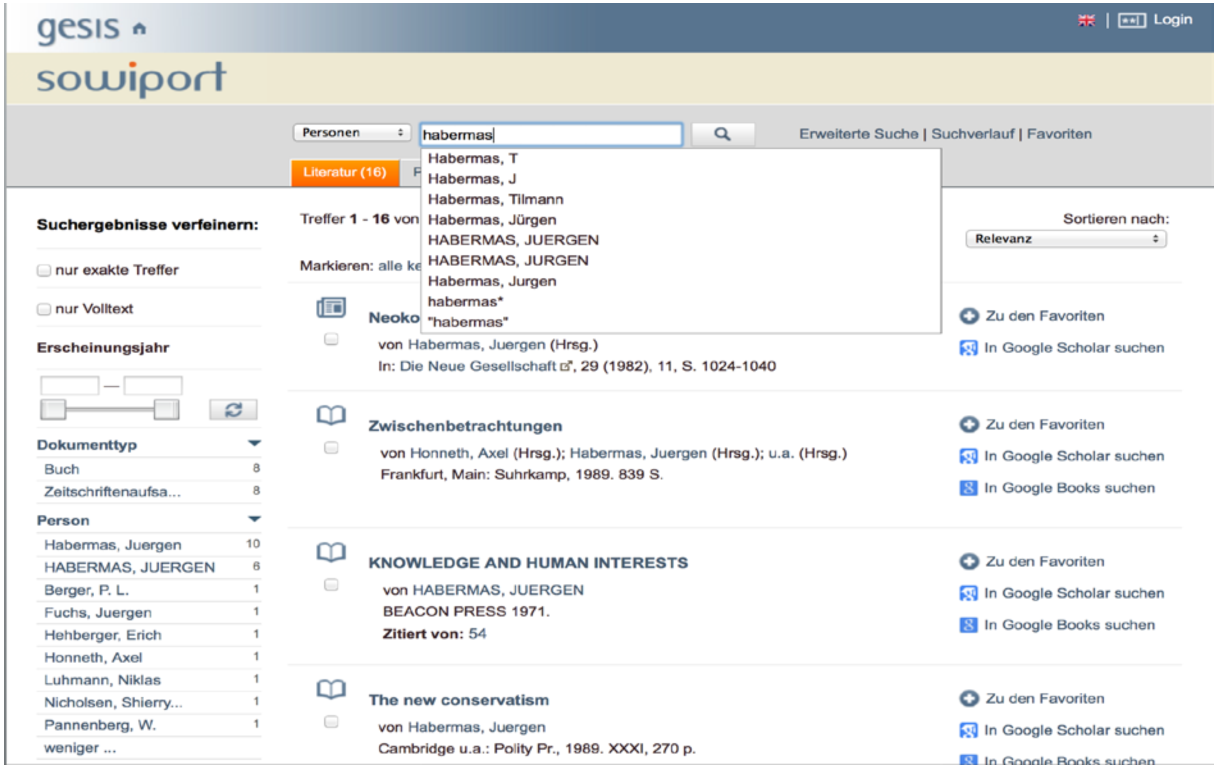


Sonya Coleman	Sean S. Coleman S. R. Coleman S. S. Coleman S John S. Coleman T Nastaran S. Coleman Robert S.
Bryan Scotney	Bryan A. Scotney T 1d 3a Bryan W. Scotney ST 1d 3a
Bryan Gardiner	Bryan Gardiner ST 1d 2a 2b M. B. Gardiner Paul H. B

> **T** = indication of time proximity

> **S** = indication of topical closeness

Concrete use case at GESIS



gesis sowiport

Login

Personen | habermas

Habermas, T
Habermas, J
Habermas, Tilmann
Habermas, Jürgen
HABERMAS, JUERGEN
HABERMAS, JURGEN
Habermas, Jurgen
habermas*
"habermas"

Erweiterte Suche | Suchverlauf | Favoriten

Literatur (16)

Suchergebnisse verfeinern:

□ nur exakte Treffer

□ nur Volltext

Erscheinungsjahr

Dokumenttyp

Buch	8
Zeitschriftenaufsa...	8

Person

Habermas, Juergen	10
HABERMAS, JUERGEN	6
Berger, P. L.	1
Fuchs, Juergen	1
Hehberger, Erich	1
Honneth, Axel	1
Luhmann, Niklas	1
Nicholsen, Shierry...	1
Pannenberg, W.	1

weniger ...

Treffer 1 - 16 von

Markieren: alle ke

Sortieren nach: Relevanz

□ Neoko
von Habermas, Juergen (Hrsg.)
In: Die Neue Gesellschaft, 29 (1982), 11, S. 1024-1040

Zu den Favoriten
In Google Scholar suchen

□ Zwischenbetrachtungen
von Honneth, Axel (Hrsg.); Habermas, Juergen (Hrsg.); u.a. (Hrsg.)
Frankfurt, Main: Suhrkamp, 1989. 839 S.

Zu den Favoriten
In Google Scholar suchen
In Google Books suchen

□ KNOWLEDGE AND HUMAN INTERESTS
von HABERMAS, JUERGEN
BEACON PRESS 1971.
Zitiert von: 54

Zu den Favoriten
In Google Scholar suchen
In Google Books suchen

□ The new conservatism
von Habermas, Juergen
Cambridge u.a.: Polity Pr., 1989. XXXI, 270 p.

Zu den Favoriten
In Google Scholar suchen
In Google Books suchen

GESIS Portal sowiport

Literature

- 10 different data sources

- about 7,5 million publication records

Future work:

> Disambiguation of streams of author names

> Linking of individualized author name records to persistent identifiers

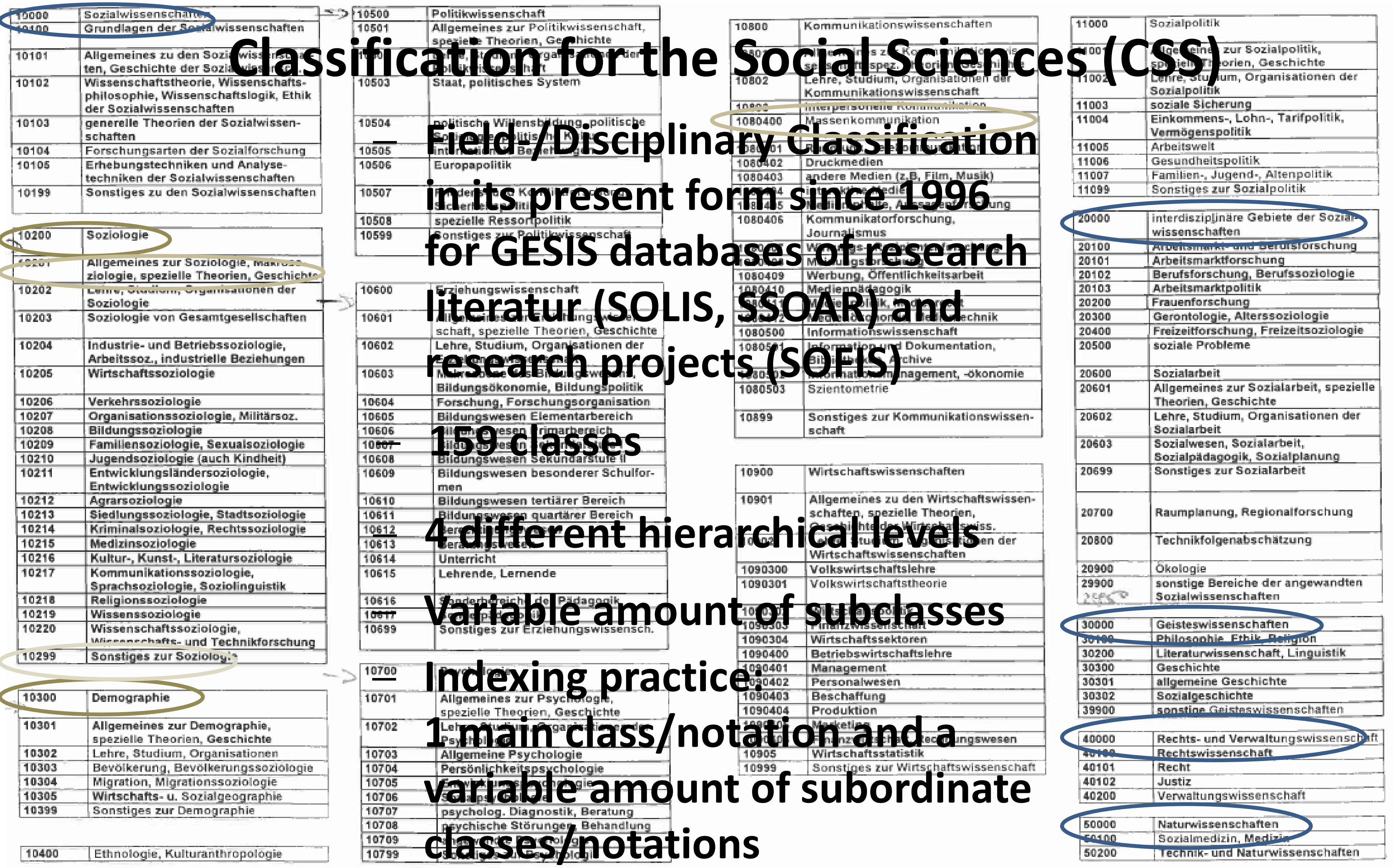
Instruments for content cataloguing at GESIS

- Classification for the Social Sciences (CSS)
  > Disciplinary assignment

- Thesaurus for the Social Sciences (TSS)
  > Content information

Classification for the Social Sciences (CSS)

- Field-/Disciplinary Classification in its present form since 1996 for GESIS databases of research literatur (SOLIS, SSOAR) and research projects (SOFIS)
- 159 classes
- 4 different hierarchical levels
- Variable amount of subclasses
- Indexing practice: 1 main class/notation and a variable amount of subordinate classes/notations



10000	Sozialwissenschaften
10100	Grundlagen der Sozialwissenschaften
10101	Allgemeines zu den Sozialwissenschaften, Geschichte der Sozialwissenschaften
10102	Wissenschaftstheorie, Wissenschaftsphilosophie, Wissenschaftslogik, Ethik der Sozialwissenschaften
10103	generelle Theorien der Sozialwissenschaften
10104	Forschungsarten der Sozialforschung
10105	Erhebungstechniken und Analysetechniken der Sozialwissenschaften
10199	Sonstiges zu den Sozialwissenschaften
10200	Soziologie
10201	Allgemeines zur Soziologie, Makrosoziologie, spezielle Theorien, Geschichte
10202	Lehre, Studium, Organisationen der Soziologie
10203	Soziologie von Gesamtgesellschaften
10204	Industrie- und Betriebssoziologie, Arbeitssoz., industrielle Beziehungen
10205	Wirtschaftssoziologie
10206	Verkehrssoziologie
10207	Organisationssoziologie, Militärsoz.
10208	Bildungssoziologie
10209	Familiensoziologie, Sexualsoziologie
10210	Jugendsoziologie (auch Kindheit)
10211	Entwicklungsländersoziologie, Entwicklungssoziologie
10212	Agrarsoziologie
10213	Siedlungssoziologie, Stadtsoziologie
10214	Kriminalsoziologie, Rechtssoziologie
10215	Medizinsoziologie
10216	Kultur-, Kunst-, Literatursoziologie
10217	Kommunikationssoziologie, Sprachsoziologie, Soziolinguistik
10218	Religionssoziologie
10219	Wissenssoziologie
10220	Wissenschaftssoziologie, Wissenschafts- und Technikforschung
10299	Sonstiges zur Soziologie
10300	Demographie
10301	Allgemeines zur Demographie, spezielle Theorien, Geschichte
10302	Lehre, Studium, Organisationen
10303	Bevölkerung, Bevölkerungssoziologie
10304	Migration, Migrationssoziologie
10305	Wirtschafts- u. Sozialgeographie
10399	Sonstiges zur Demographie
10400	Ethnologie, Kulturanthropologie

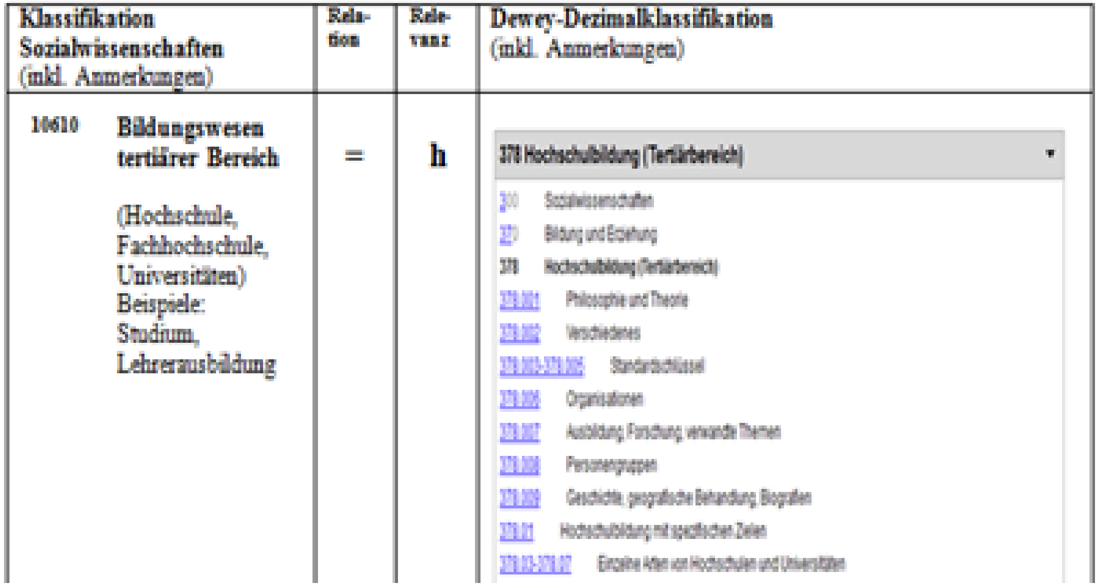
10500	Politikwissenschaft
10501	Allgemeines zur Politikwissenschaft, spezielle Theorien, Geschichte
10502	Lehre, Studium, Organisationen der Politikwissenschaft
10503	Staat, politisches System
10504	politische Willensbildung, politische Soziologie, politische Kultur
10505	internationale Beziehungen
10506	Europapolitik
10507	Friedens- und Konfliktforschung, Sicherheitspolitik
10508	spezielle Ressortpolitik
10599	Sonstiges zur Politikwissenschaft
10600	Erziehungswissenschaft
10601	Allgemeines zur Erziehungswissenschaft, spezielle Theorien, Geschichte
10602	Lehre, Studium, Organisationen der Erziehungswissenschaft
10603	Makroebene des Bildungswesens, Bildungsökonomie, Bildungspolitik
10604	Forschung, Forschungsorganisation
10605	Bildungswesen Elementarbereich
10606	Bildungswesen Primarbereich
10607	Bildungswesen Sekundarstufe I
10608	Bildungswesen Sekundarstufe II
10609	Bildungswesen besonderer Schulformen
10610	Bildungswesen tertiärer Bereich
10611	Bildungswesen quartärer Bereich
10612	Berufsbildung
10613	Beratungswesen
10614	Unterricht
10615	Lehrende, Lernende
10616	Sonderbereiche der Pädagogik
10617	[illegible]
10699	Sonstiges zur Erziehungswissensch.
10700	Psychologie
10701	Allgemeines zur Psychologie, spezielle Theorien, Geschichte
10702	Lehre, Studium, Organisationen der Psychologie
10703	Allgemeine Psychologie
10704	Persönlichkeitspsychologie
10705	Entwicklungspsychologie
10706	Sozialpsychologie
10707	psycholog. Diagnostik, Beratung
10708	psychische Störungen, Behandlung
10709	[illegible]
10799	Sonstiges zur Psychologie

10800	Kommunikationswissenschaften
10801	Allgemeines zu Kommunikationswissenschaften, spez. Theorien, Geschichte
10802	Lehre, Studium, Organisationen der Kommunikationswissenschaft
10803	interpersonelle Kommunikation
1080400	Massenkommunikation
1080401	Rundfunk, Telekommunikation
1080402	Druckmedien
1080403	andere Medien (z.B. Film, Musik)
1080404	interaktive, elektronische Medien
1080405	Medieninhalte, Aussagenforschung
1080406	Kommunikatorforschung, Journalismus
1080407	Wirkungsforschung, Rezipientenforschung
1080408	Nutzungsforschung
1080409	Werbung, Öffentlichkeitsarbeit
1080410	Medienpädagogik
1080411	Medienpolitik, Informationspolitik, Medienrecht
1080412	Medienökonomie, Medientechnik
1080500	Informationswissenschaft
1080501	Information und Dokumentation, Bibliothekswesen, Archive
1080502	Informationsmanagement, -ökonomie
1080503	Szientometrie
10899	Sonstiges zur Kommunikationswissenschaft
10900	Wirtschaftswissenschaften
10901	Allgemeines zu den Wirtschaftswissenschaften, spezielle Theorien, Geschichte der Wirtschaftswiss.
10902	Lehre, Studium, Organisationen der Wirtschaftswissenschaften
1090300	Volkswirtschaftslehre
1090301	Volkswirtschaftstheorie
1090302	Wirtschaftspolitik
1090303	Finanzwissenschaft
1090304	Wirtschaftssektoren
1090400	Betriebswirtschaftslehre
1090401	Management
1090402	Personalwesen
1090403	Beschaffung
1090404	Produktion
1090405	Marketing
1090406	Finanzwirtschaft, Rechnungswesen
10905	Wirtschaftsstatistik
10999	Sonstiges zur Wirtschaftswissenschaft

11000	Sozialpolitik
11001	Allgemeines zur Sozialpolitik, spezielle Theorien, Geschichte
11002	Lehre, Studium, Organisationen der Sozialpolitik
11003	soziale Sicherung
11004	Einkommens-, Lohn-, Tarifpolitik, Vermögenspolitik
11005	Arbeitswelt
11006	Gesundheitspolitik
11007	Familien-, Jugend-, Altenpolitik
11099	Sonstiges zur Sozialpolitik
20000	interdisziplinäre Gebiete der Sozialwissenschaften
20100	Arbeitsmarkt- und Berufsforschung
20101	Arbeitsmarktforschung
20102	Berufsforschung, Berufssoziologie
20103	Arbeitsmarktpolitik
20200	Frauenforschung
20300	Gerontologie, Alterssoziologie
20400	Freizeitforschung, Freizeitsoziologie
20500	soziale Probleme
20600	Sozialarbeit
20601	Allgemeines zur Sozialarbeit, spezielle Theorien, Geschichte
20602	Lehre, Studium, Organisationen der Sozialarbeit
20603	Sozialwesen, Sozialarbeit, Sozialpädagogik, Sozialplanung
20699	Sonstiges zur Sozialarbeit
20700	Raumplanung, Regionalforschung
20800	Technikfolgenabschätzung
20900	Ökologie
29900	sonstige Bereiche der angewandten Sozialwissenschaften
30000	Geisteswissenschaften
30100	Philosophie, Ethik, Religion
30200	Literaturwissenschaft, Linguistik
30300	Geschichte
30301	allgemeine Geschichte
30302	Sozialgeschichte
39900	sonstige Geisteswissenschaften
40000	Rechts- und Verwaltungswissenschaft
40100	Rechtswissenschaft
40101	Recht
40102	Justiz
40200	Verwaltungswissenschaft
50000	Naturwissenschaften
50100	Sozialmedizin, Medizin
50200	Technik- und Naturwissenschaften

Classification cross-concordances

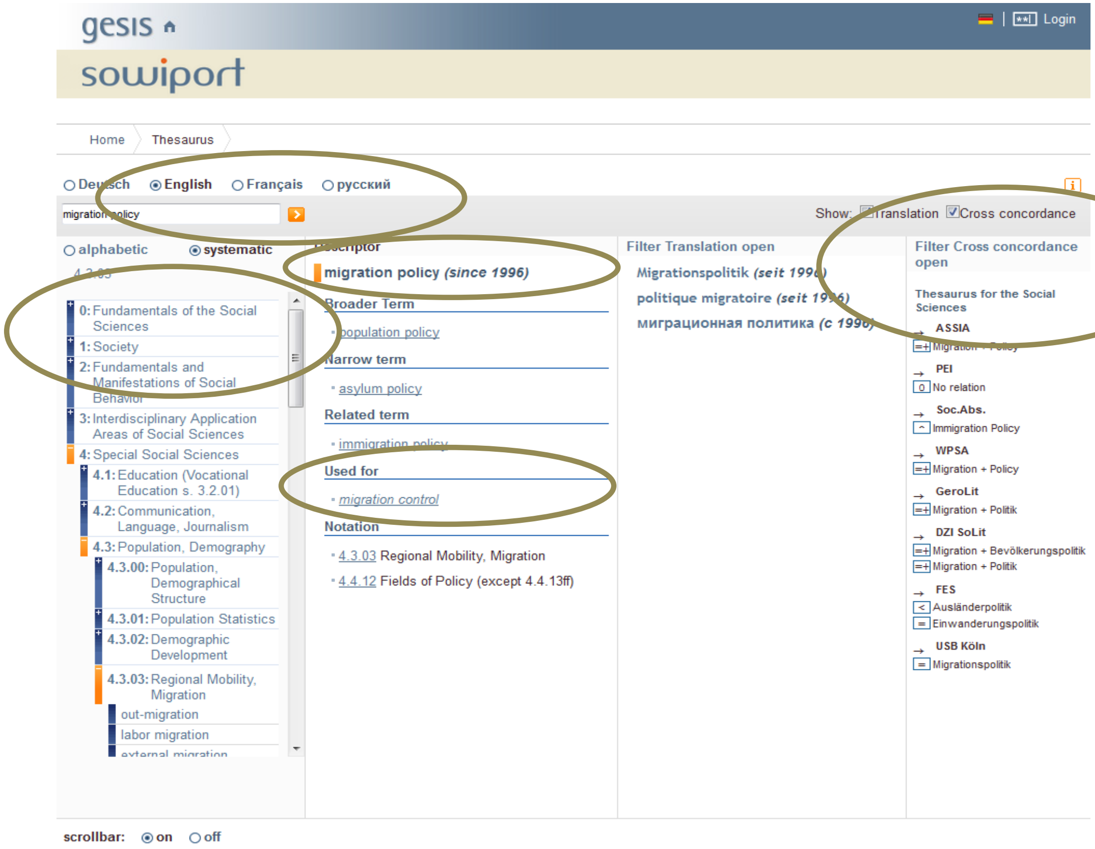
- Bilateral mapping between CCS and DDC (CSS > DDC / DDC > CSS) (2012/2013)



Klassifikation Sozialwissenschaften (inkl. Anmerkungen)	Relation	Relevanz	Dewey-Dezimalklassifikation (inkl. Anmerkungen)
10610 Bildungswesen tertiärer Bereich (Hochschule, Fachhochschule, Universitäten) Beispiele: Studium, Lehrerausbildung	=	h	378 Hochschulbildung (Tertiärbereich) ▾ 300 Sozialwissenschaften 370 Bildung und Erziehung 378 Hochschulbildung (Tertiärbereich) 378.001 Philosophie und Theorie 378.002 Verschiedenes 378.003-378.005 Standardschlüssel 378.006 Organisationen 378.007 Ausbildung, Forschung, verwandte Themen 378.008 Personengruppen 378.009 Geschichte, geografische Behandlung, Biografien 378.01 Hochschulbildung mit spezifischen Zielen 378.03-378.07 Einzelne Arten von Hochschulen und Universitäten

Start voc.	End voc. (main panels)	Rel. in total	Exact Match	Broader Match	Narrower Match	Null Rel.	Simple Matches	Multiple Matches (1:n)	Multiple Matches (n:1)
159	>27,000	169	45	15	89	1	20	41	16

Thesaurus for the Social Sciences (since 1979)



○ Deutsch ● English ○ Français ○ русский

migration policy

Show: ☑ Translation ☑ Cross concordance

○ alphabetic ● systematic

4.3.03

- 0: Fundamentals of the Social Sciences
- 1: Society
- 2: Fundamentals and Manifestations of Social Behavior
- 3: Interdisciplinary Application Areas of Social Sciences
- 4: Special Social Sciences
  - 4.1: Education (Vocational Education s. 3.2.01)
  - 4.2: Communication, Language, Journalism
  - 4.3: Population, Demography
    - 4.3.00: Population, Demographical Structure
    - 4.3.01: Population Statistics
    - 4.3.02: Demographic Development
    - 4.3.03: Regional Mobility, Migration
      - out-migration
      - labor migration
      - external migration

Descriptor

migration policy *(since 1996)*

Broader Term

- population policy

Narrow term

- asylum policy

Related term

- immigration policy

Used for

- *migration control*

Notation

- 4.3.03 Regional Mobility, Migration
- 4.4.12 Fields of Policy (except 4.4.13ff)

Filter Translation open

Migrationspolitik *(seit 1996)*

politique migratoire *(seit 1996)*

миграционная политика *(с 1996)*

Filter Cross concordance open

Thesaurus for the Social Sciences

→ **ASSIA**
=+ Migration + Policy

→ **PEI**
0 No relation

→ **Soc.Abs.**
^ Immigration Policy

→ **WPSA**
=+ Migration + Policy

→ **GeroLit**
=+ Migration + Politik

→ **DZI SoLit**
=+ Migration + Bevölkerungspolitik
=+ Migration + Politik

→ **FES**
< Ausländerpolitik
= Einwanderungspolitik

→ **USB Köln**
= Migrationspolitik

scrollbar: ● on ○ off

- Translation into English, French, (Russian)

- About 8,000 subject headings

- About 4,000 non-descriptors/synonyms

- Classification scheme

- Cross-concordances to other vocabularies

- Indexing practice: 10-15 subject headings/document

Thesaurus cross-concordances

- Major terminology mapping initiative (KoMoHe: 2004 – 2007) funded by the German Federal Ministry of Education and Research
- Mapped vocabularies i.a. AGROVOC, Medical Subject Headings (MeSH), Thesaurus for Economics (STW)

Start voc.	**End voc.**	**Rel. in total**	**Equiv. Rel.**	**BT-Rel.**	**NT-Rel.**	**Assoc. Rel.**	**Null-Rel.**	**Start terms**	**End terms**	**Term combi-nations**
TSS	IAF (*SWD*)	8,208	7,098	295	292	356	160	7,662	6,838	551
IAF (*SWD*)	TSS	9,432	6,276	1,831	134	640	594	8,890	5,556	182

TSS > IAF/IAF > TSS: Bilateral cross-concordance is mutually continuously developed

Experimental pre-study

- **Research question:**
  Can topic information (subject headings, class notations) help to distinguish between different authors with the same name?
- **Data set:** Social Science Research Literature Information System (SOLIS) (since 1978):
  - Bibliographic records including content information on German social science literature (monographs, compilations, journal articles (300 journals) and grey literature)
  - 450,000 social science publications (Jan. 2014) from 1945-

> At this point of our study we perceive name records in SOLIS as being fully disambiguated.

Experimental pre-study – set up

- Approach: Longitudinal analysis of content information
  - Publication years: 1954 - 2013
  - „Profiling“ of authors (more than two publications): 63,683 author names (81.14% of publications in single authorship)
  - Multi-level approach:
    - Macro-level:
      How discriminative/expressive are subject headings/class notations in general?
    - Meso- (group) level:
      How do topic distributions of research interests along a career look like?
      - Group 1 (5-10 years of publication activity): 16,108 author names
      - Group 2 (20-30 years of publication activity): 7,953 author names
      - Group 3 (40-50 years of publication activity): 482 author names
    - Micro-level:
      How do individual topic distributions of research interests along a career look like?
      - Group 1 (5-10 years of publication activity): one example
      - Group 2 (20-30 years of publication activity): two examples
      - Group 3 (40-50 years of publication activity): one example

Preliminary results: Macro-level

Subject headings

- Mean value of subject headings (whole thesaurus) for the average author: 48.11
- Mean value of authors per subject heading (whole thesaurus): 375.46
- for the selection of publications we have included in our pre-study frequencies of use (subject heading/author) range from 1 author (e.g. *official title* – ger. Amtsbezeichnung) to 29,599 authors (*historical development*)

Preliminary results: Macro-level

Classification (1/2)

- Mean value of classes (**whole classification**) for the average author: 6.57
- Mean value of **core classes** (#1….) for the average author: 6.02
- Mean value of **aggregated classes** (first 3 digits of classification ID for whole classification) for the average author: 3.98
- Mean value of **aggregated core classes** (first 3 digits of classification ID for core classes) for the average author: 3.47

Preliminary results: Macro-level

Classification (2/2)

- Average number of author names per class (whole classification): 2,631
- Average number of author names per class (**core classes**): 2,655
- Average number of author names per **aggregated class** (first 3 digits of class-ID – whole classification): 7,908
- Average number of author names per **aggregated class** (first 3 digits of class-ID) **of core classes**: 10,421

> The coverage of the core areas/classes is higher than for the more marginal areas/classes of the classification.

Preliminary results: Meso-level

Classification

- Average number of classes per author (**1st group: 5-10** years of publication):
  - Whole classification: 5.45
  - Aggregated classes: 3.61
- Average number of classes per author (**2nd group: 20-30** years of publication):
  - Whole classification: 10.98
  - Aggregated classes: 5.79
- Average number of classes per author (**3rd group: 40-50** years of publication):
  - Whole classification: 21.50
  - Aggregated classes: 9.31

Preliminary results: Meso-level

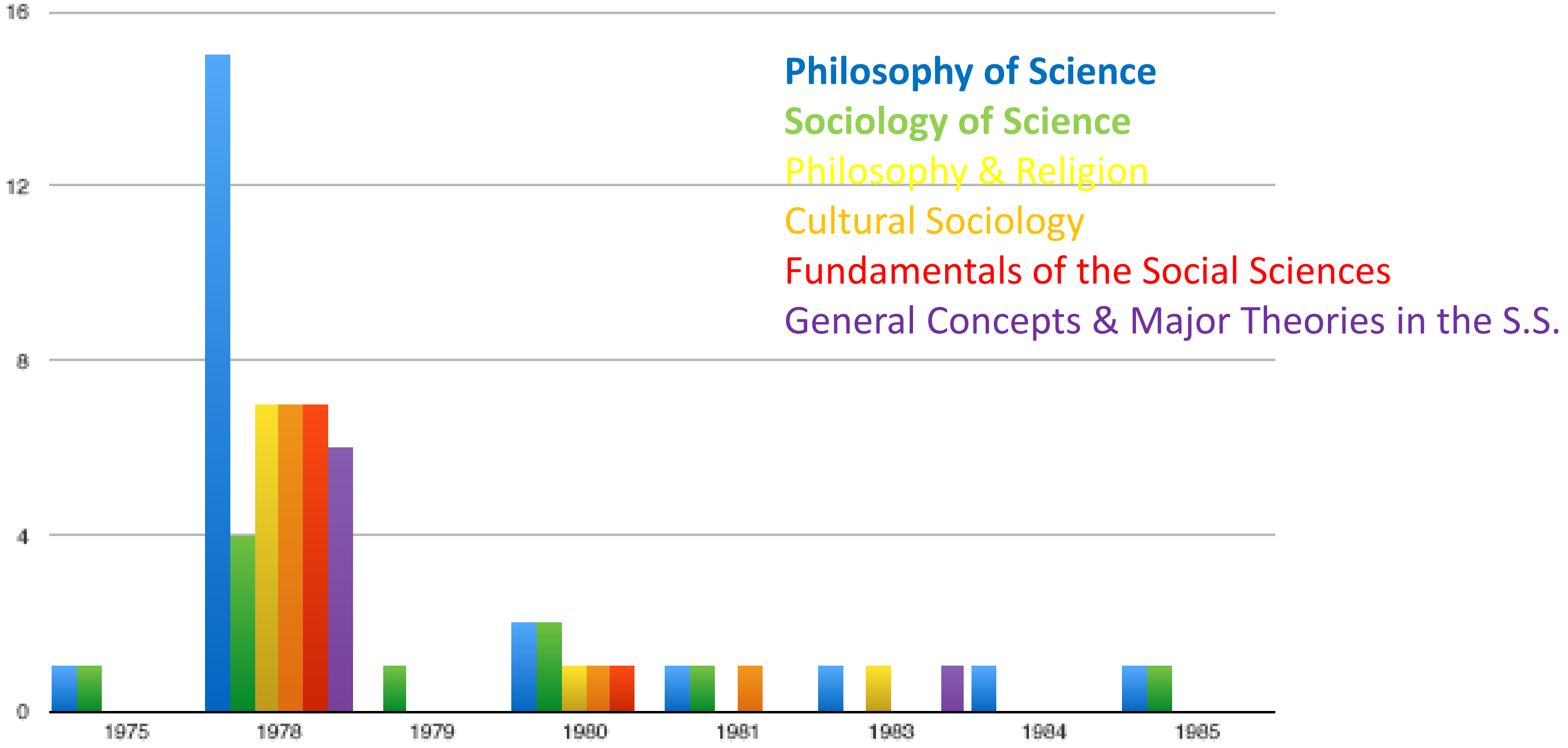
Subject headings

- Average number of different subject headings per author (**1st group: 5-10** years of publication):
  - 37.87
- Average number of different subject headings per author (**2nd group: 20-30** years of publication):
  - 84.48
- Average number of different subject headings per author (**3rd group: 40-50** years of publication):
  - 191.92

Preliminary results: Micro-level

Example 1st group: P. K. F.
10 years of publication activity)/29 publications

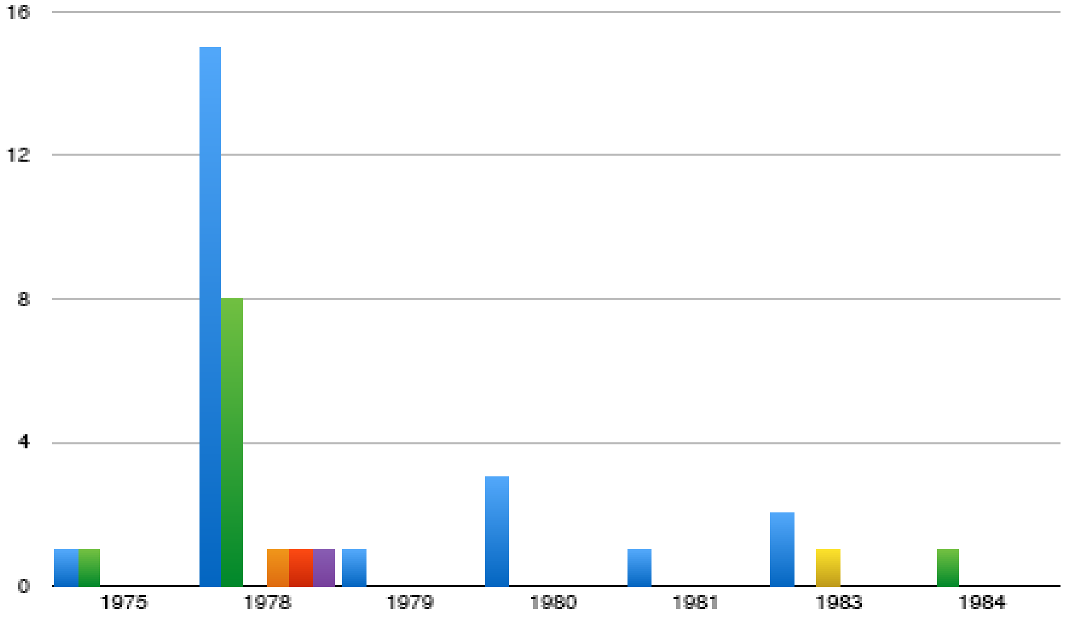
Distribution of most frequently assigned classes



Preliminary results: Micro-level

Example 1st group: P. K. F. 10 years of publication activity)/29 publications

Distribution of most frequently assigned subject headings

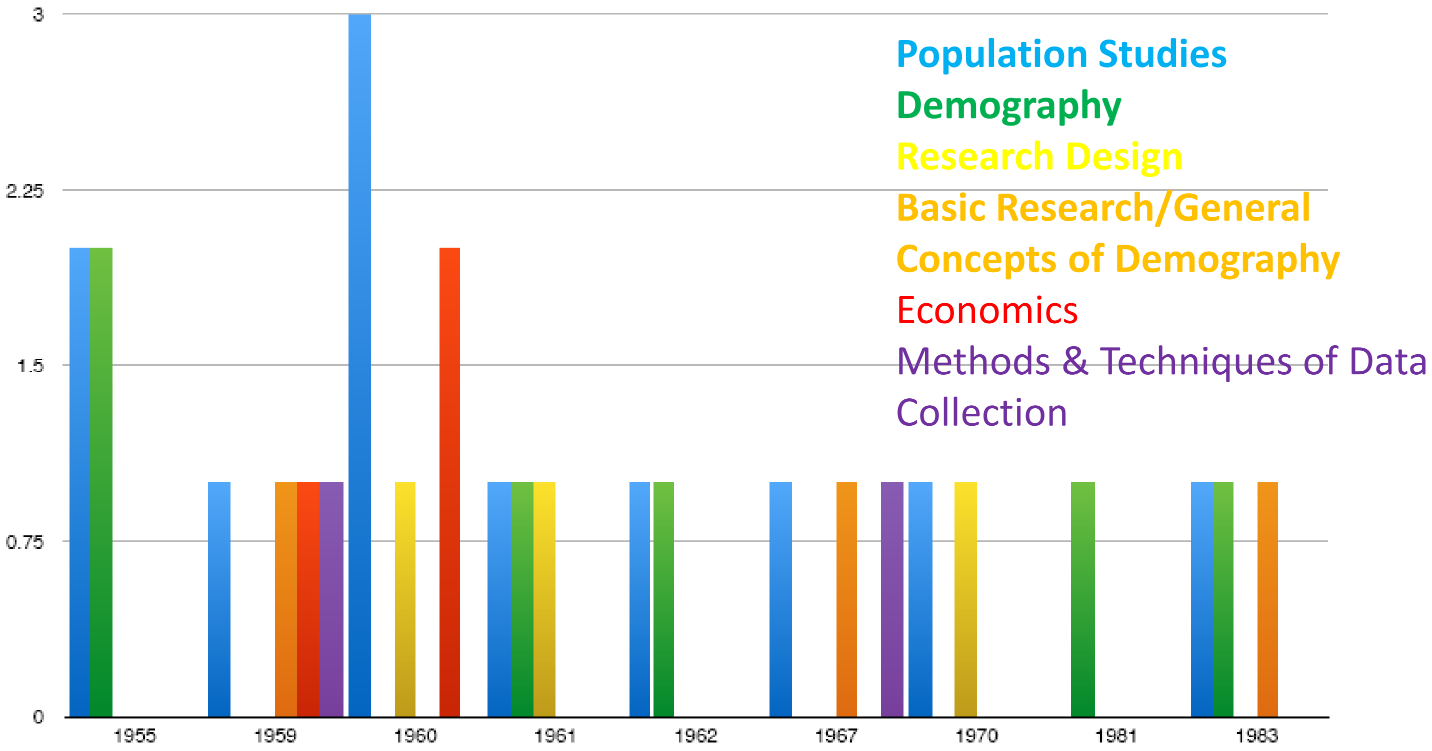


Science
Knowledge
Theory
Logic
Experience
Philosophy of science

Preliminary results: Micro-level

Example (a) 2nd group: S. K.
28 years of publication activity/17 publications

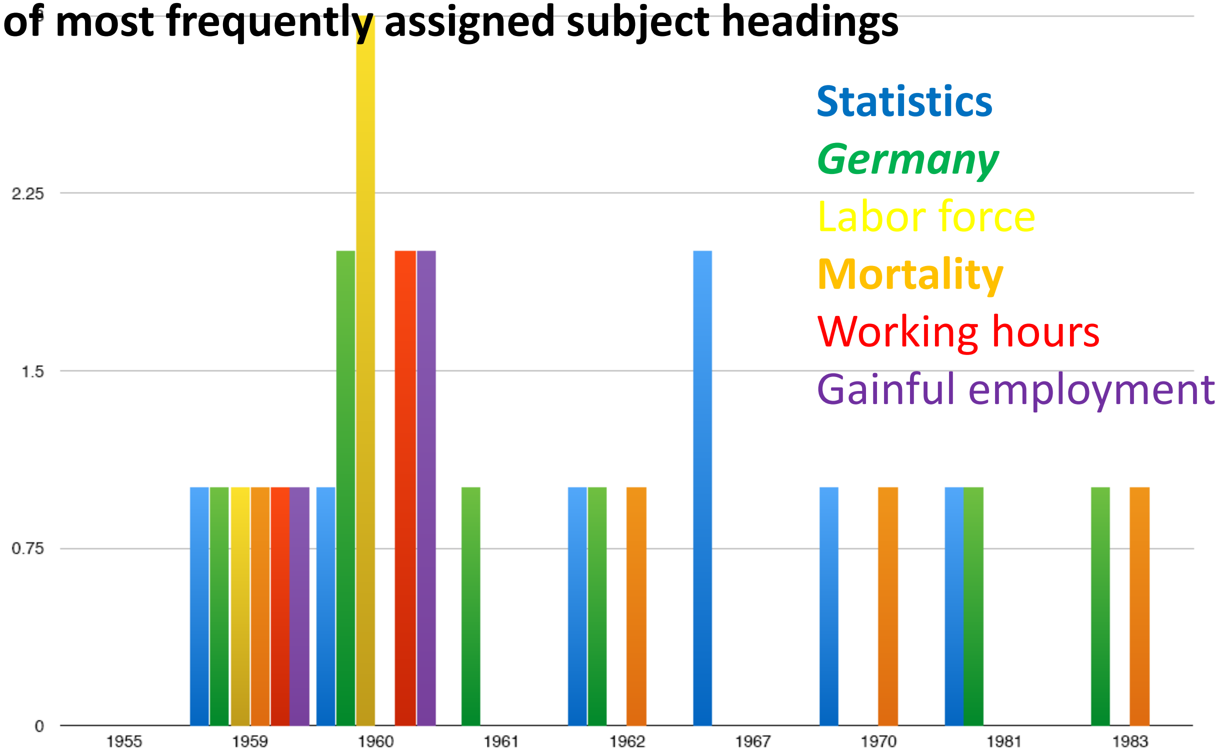
Distribution of most frequently assigned classes



Preliminary results: Micro-level

Example (a) 2nd group: S. K.
28 years of publication activity/17 publications

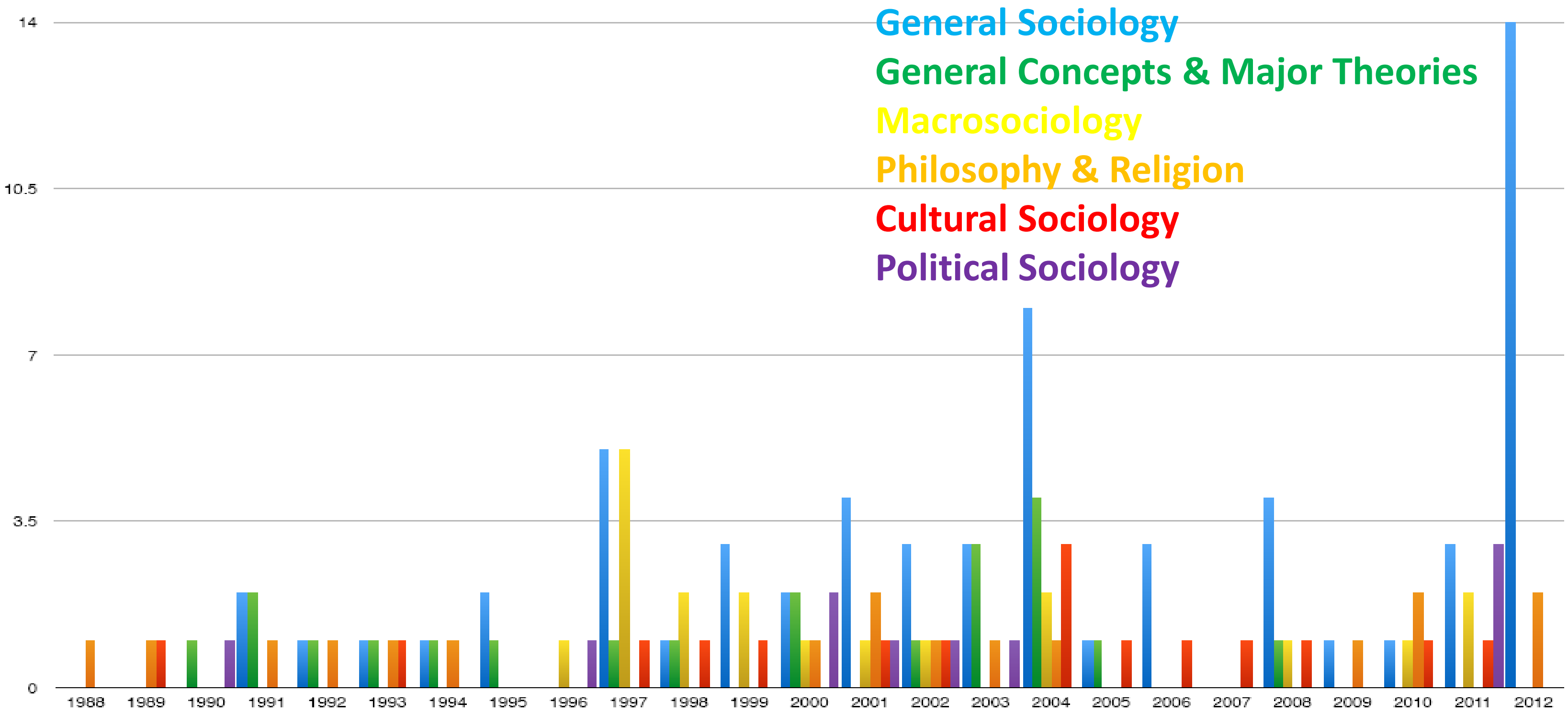
Distribution of most frequently assigned subject headings



Preliminary results: Micro-level

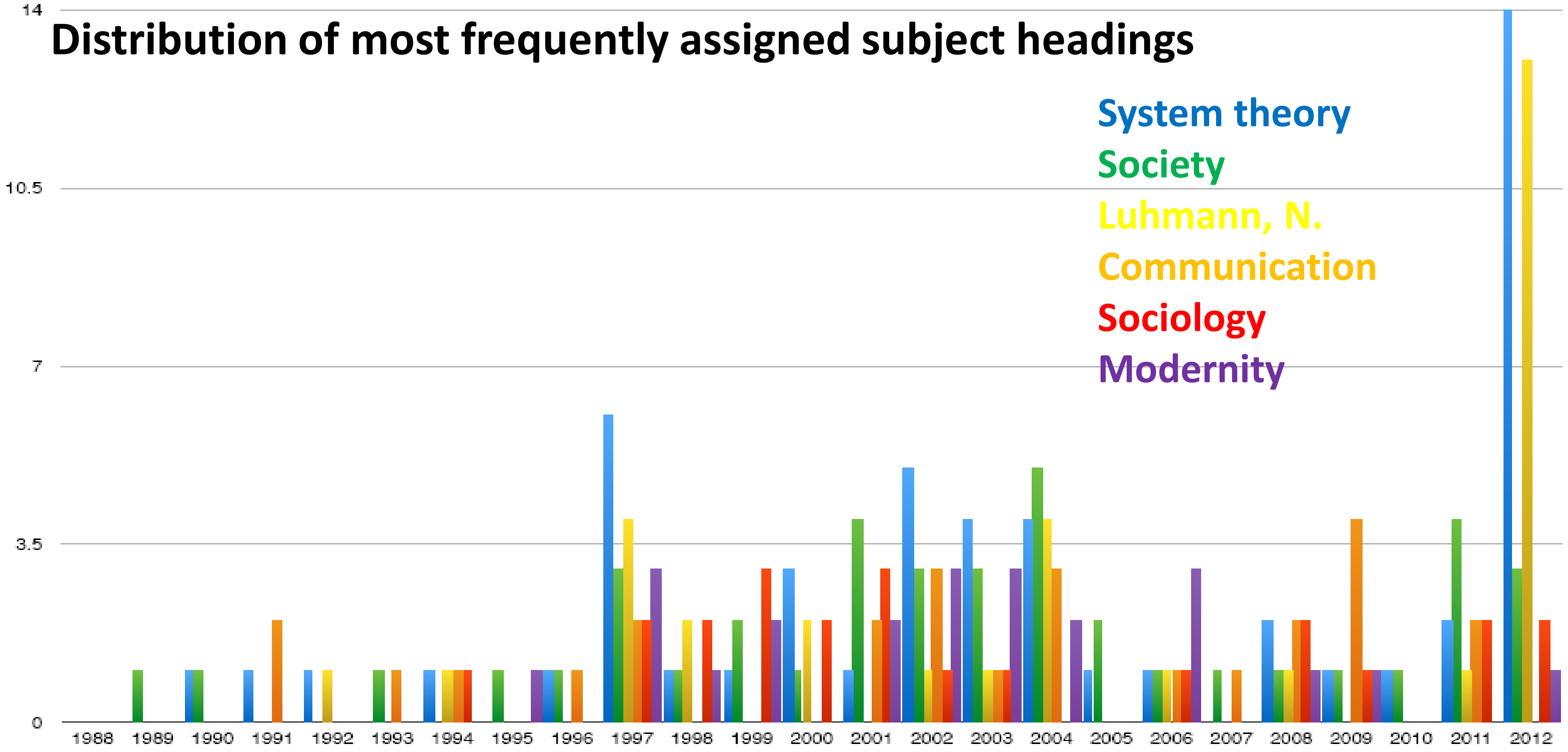
Example (b) 2nd group: A. N.
24 years of publication activity/146 publications

Distribution of most frequently assigned classes



Preliminary results: Micro-level

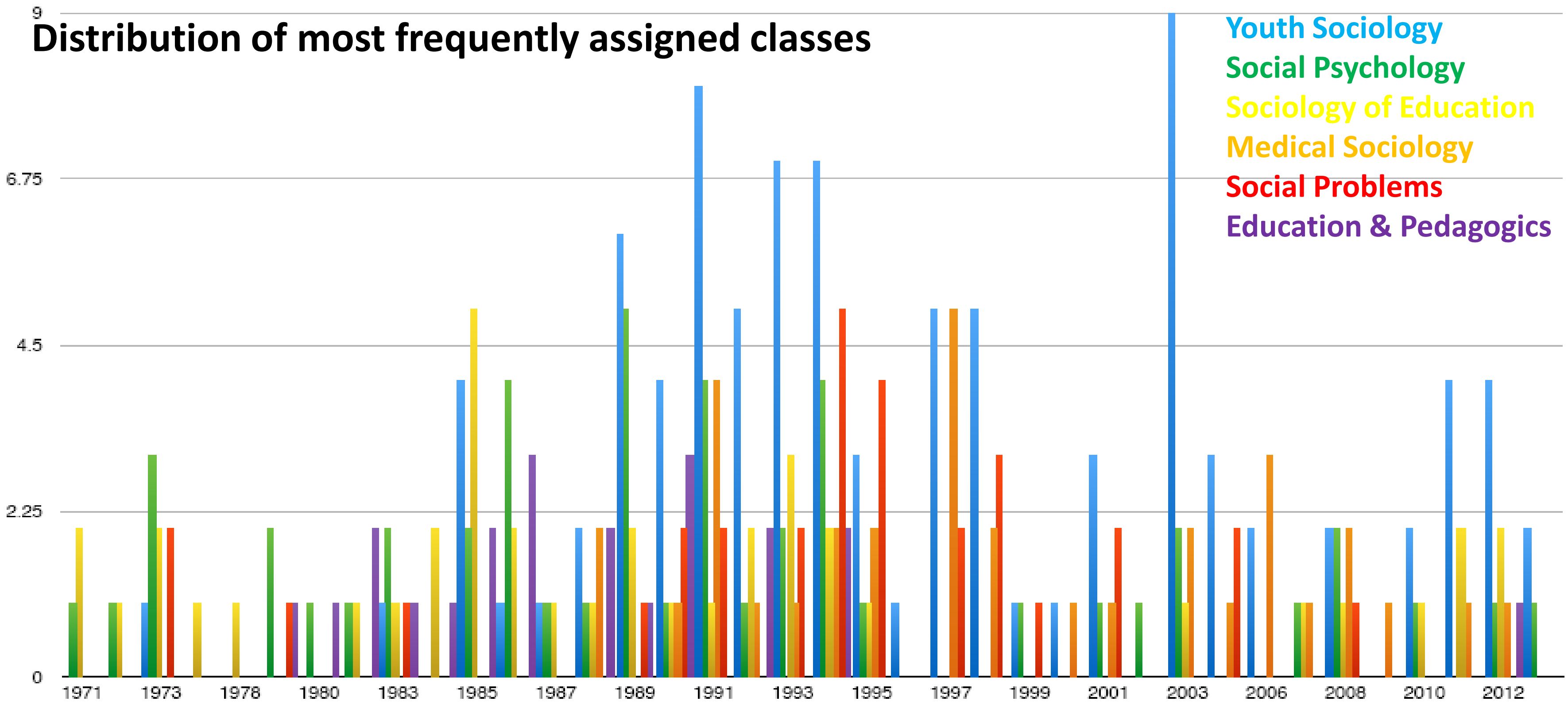
Example (b) 2nd group: A. N.
24 years of publication activity/146 publications



Distribution of most frequently assigned subject headings

Preliminary results: Micro-level

Example 3rd group: K. H.
41 years of publication activity/241 publications

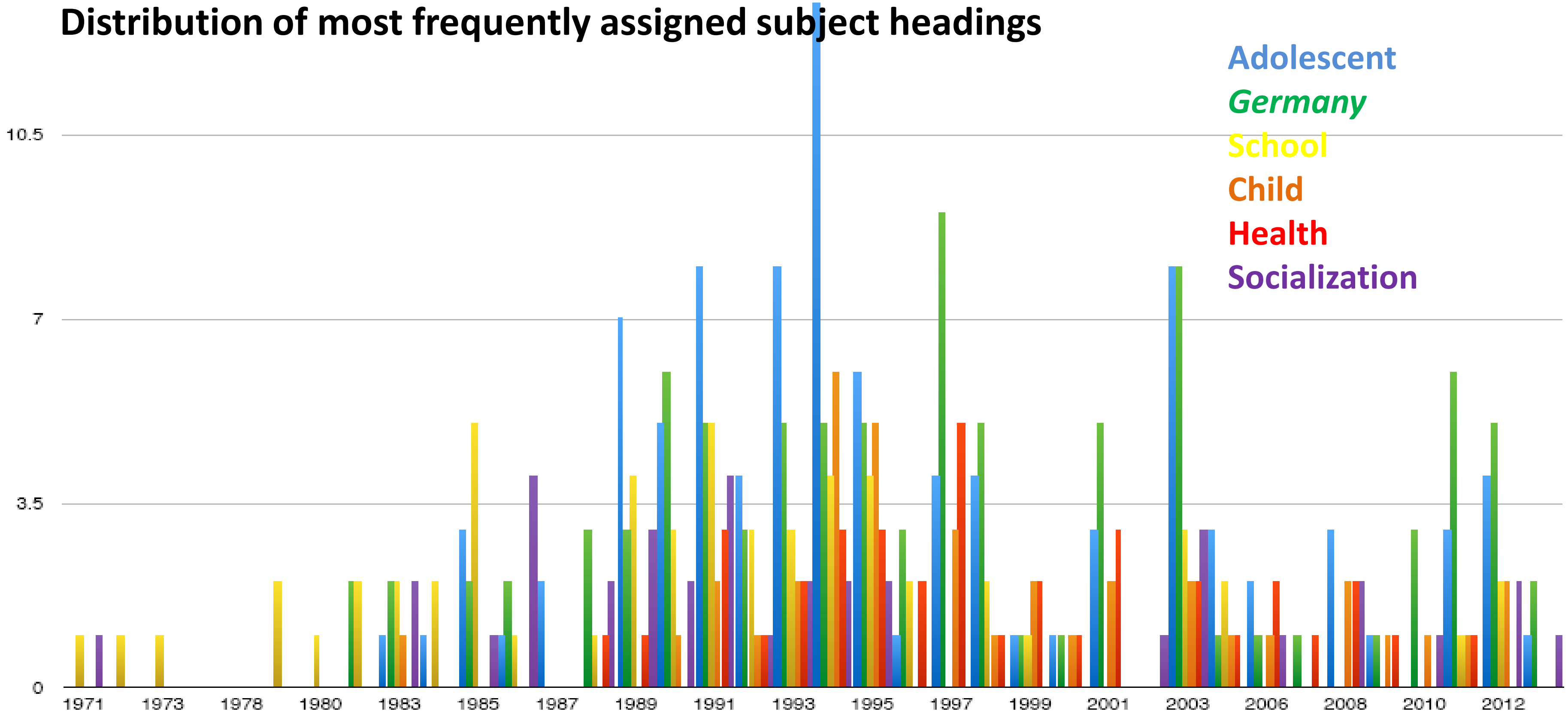


Distribution of most frequently assigned classes

Preliminary results: Micro-level

Example 3rd group: K. H. 41 years of publication activity/241 publications

Distribution of most frequently assigned subject headings



Preliminary results –

for the German-speaking social sciences

- **Research question:**

  Can topic information (subject headings, classification) help to distinguish between different authors with the same name?
  OR: Is there subject continuity along a researcher´s career?

- **Macro-level of study:**
  - Speaking of an „average author“ could only serve as an approximation to answer the research question.
  - The frequency of use differs enormously between different classes/subject headings.
- **Meso-level of study:**
  - The average number of classes/subject headings grows continuously along a researcher´s career. Using subject information for author name disambiguation therefore might appear rather fruitless. However,…
- **Micro-level of study:**
  - Considering the most frequently used classes/subject headings assigned to publications of an author could help to profile/ disambiguate an author name. Subject continuity could be identified. This already applies to authors with a rather short period of publication activity.

Outlook

- **Future workflow**
  - ↓ 1st step of disambiguation: Preprocessing via standard algorithms on the basis of **database-internal** publication/author-centred reference information (publication properties like co-authors, year of publication etc.)
  - ↓ 2nd step of disambiguation: Consideration of classification and subject heading information for cases of doubt
  - ↓Linking person records to authority files: Use of **database-external** person-centred reference information (individualized/differentiated person records of the Integrated Authority File (IAF)) via mapping of bibliographic records and individualized person IDs on the basis of an overlap of subject information using cross-concordances between classifications (DDC/CSS) and thesauri (IAF/TSS).
    - Preliminary results: Differentiated person records in the IAF only cover a small section of author strings included in the GESIS portal sowiport (mainly German-speaking social scientists)
    - Future work: Taking cross-concordances between thesauri (TSS – IAF) and classifications (CSS – DDC) for author name disambiguation into account.

Thank you very much for your attention.

Contact

Dr. Andreas Oskar Kempf (GESIS)
Andreas.Kempf@gesis.org

Dr. Cornelia Hedeler (University of Manchester)
chedeler@cs.manchester.ac.uk

Jan Steinberg (GESIS)
Jan.Steinberg@gesis.org

References

- Ferreira, Anderson A.; Goncalves, Marcos André; Laender, Albert H. F. (2012) A Brief Survey of Automatic Methods for Author Name Disambiguation, *SIGMOD Record*, vol. 41, No. 2.
- Liu, Wanli; Dogan, Rezarta Islamaj; Kim, Sun; Corneau, Donald C.; Kim, Won; Yeganova, Lena; Lu, Zhiyong; Wilbur, John W. (2013) Author Name Disambiguation for PubMed, *JASIST*.
- Mayr, Philipp; Petras, Vivien (2008b): *Cross-concordances: terminology mapping and its effectiveness for information retrieval.* In: 74th IFLA World Library and Information Congress. Québec, Canada URL: www.ifla.org/IV/ifla74/papers/129-Mayr_Petras-en.pdf.
- Riege, Udo (1998) *Elektronische Version des Thesaurus und der Klassifikation Sozialwissenschaften*, IZ-Arbeitsbericht Nr. 14, GESIS IZ InformationsZentrum Sozialwissenschaften.
- http://www.gesis.org/unser-angebot/recherchieren/thesauri-und-klassifikationen/thesaurus-sozialwissenschaften/#c25531 (last accessed July 1st, 2014).
- http://www.gesis.org/fileadmin/upload/dienstleistung/tools_standards/KassifikationSozialwissenschaften_Stand_Juli_2013_dt_en__2_.pdf (last accessed July 1st, 2014).