Investigating interdisciplinary knowledge flow from the content perspective of citances

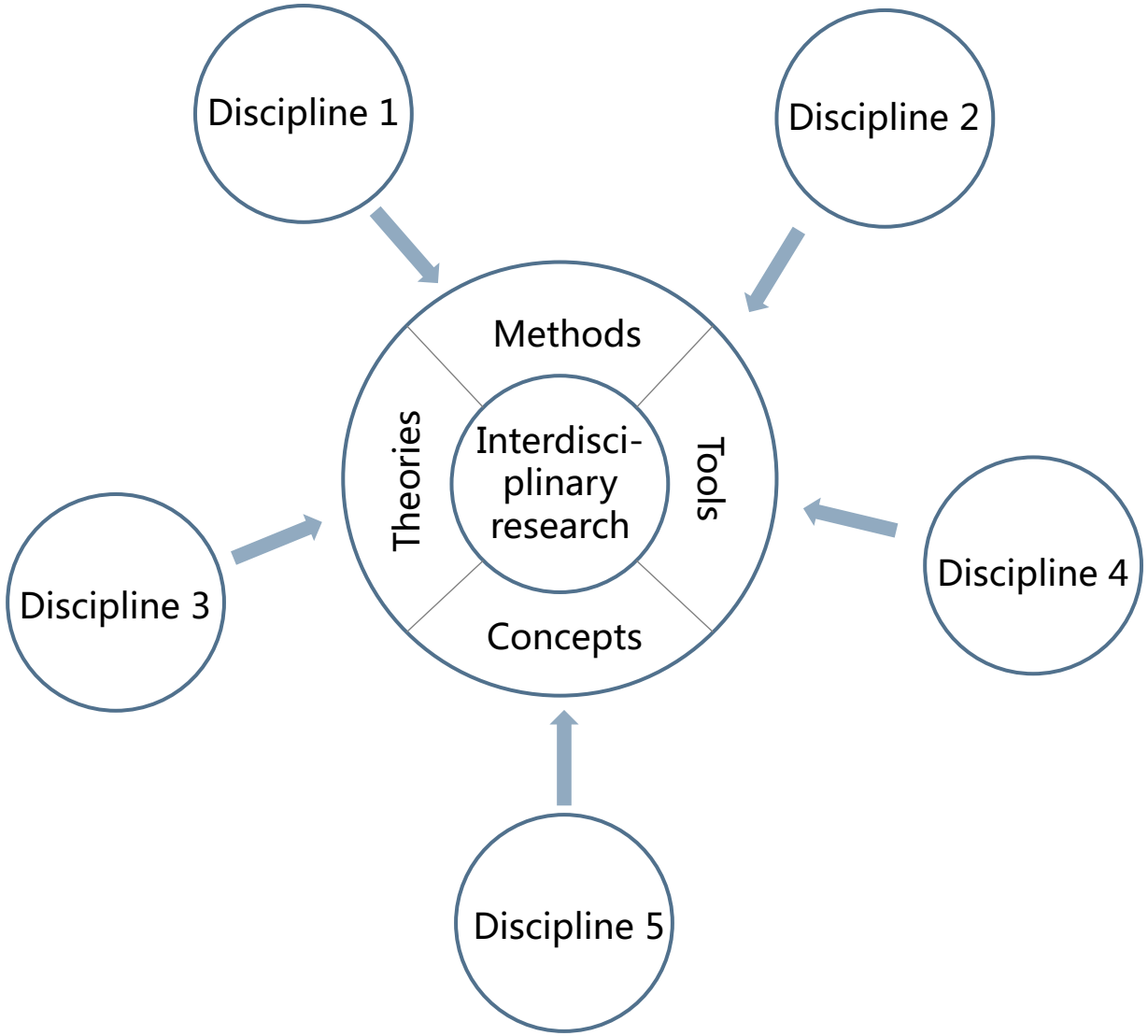
Wuhan University

Jin Mao, Shiyun Wang (Presenter) and Xianli Shang

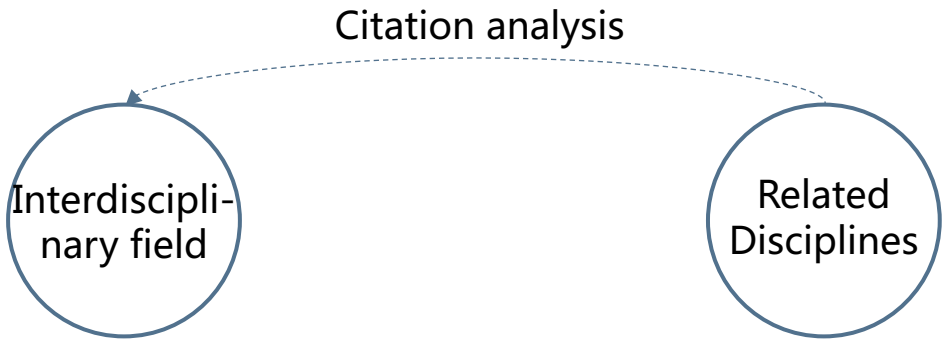
August 1, 2020

Introduction

Interdisciplinary knowledge integration



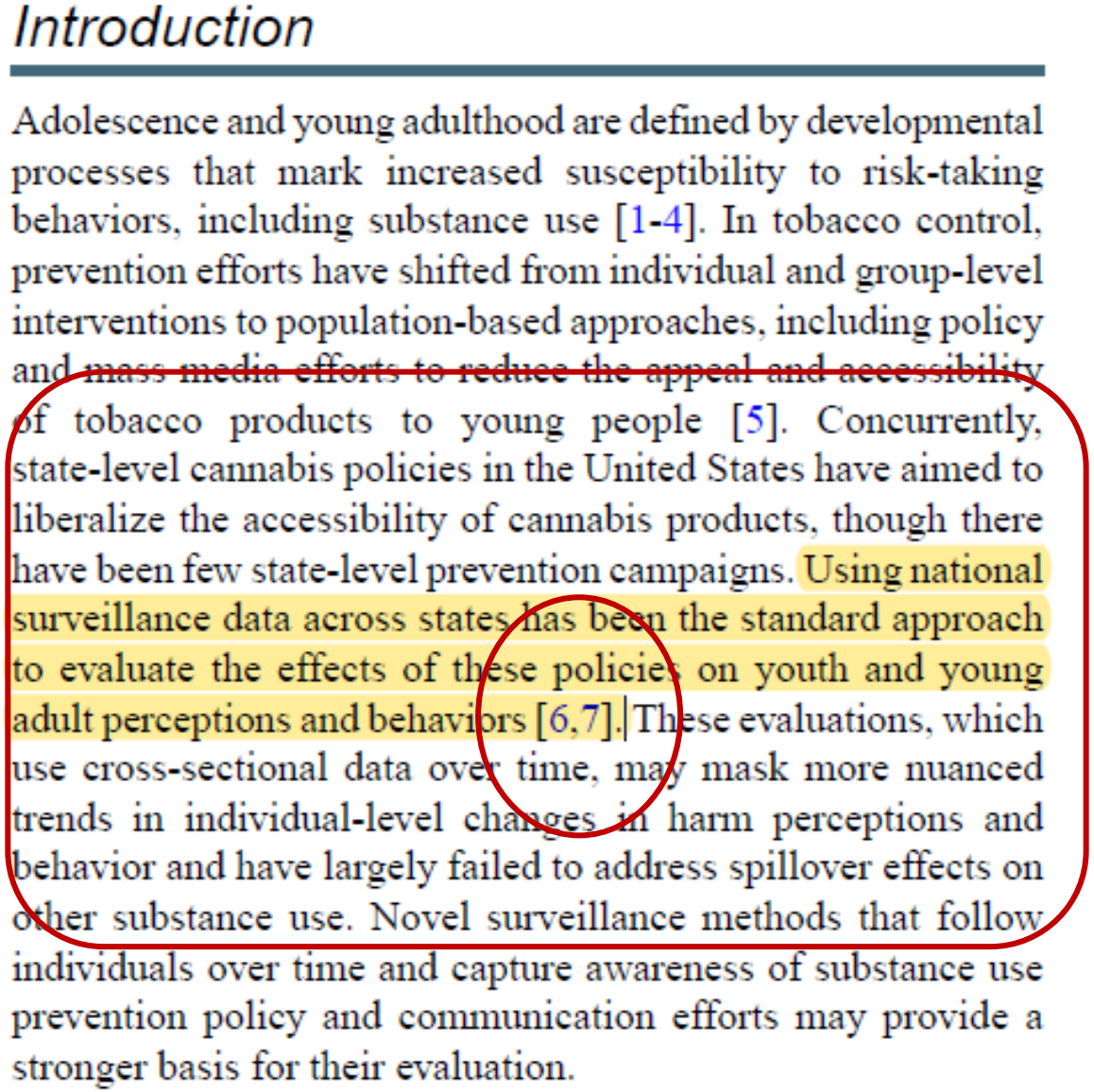
Interdisciplinary knowledge flow



- Conventionally, the knowledge flow to a field is **simply measured** by the number of references cited by the papers in the field.
- Different **importance, functions** and other aspects of citations in a paper are ignored

Introduction

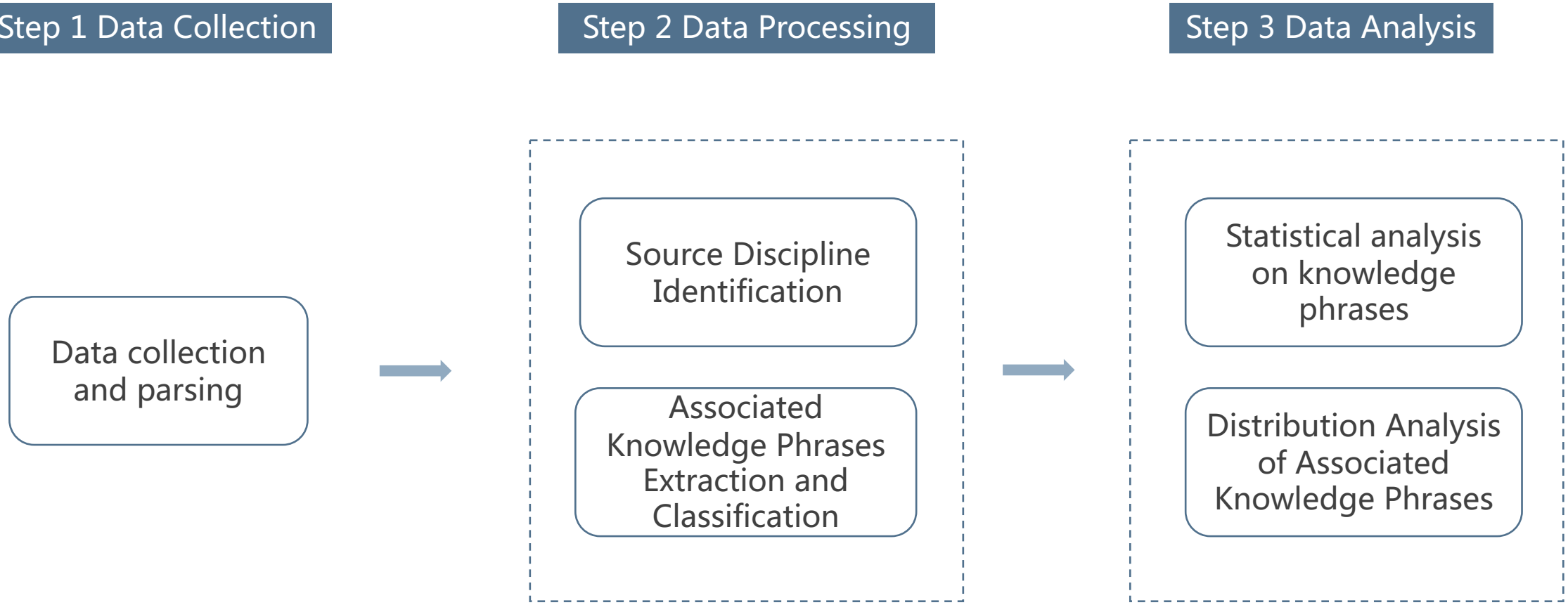
- **Citation contexts** embed the syntactic (e.g., the location of section) and semantic (e.g., the meaning of citation content) information of citations
- In this study, we attempt to explore **what knowledge** is integrated into an interdisciplinary field, **eHealth**, by analyzing the **citances** (i.e. the sentence that contains in-text citations)



Introduction

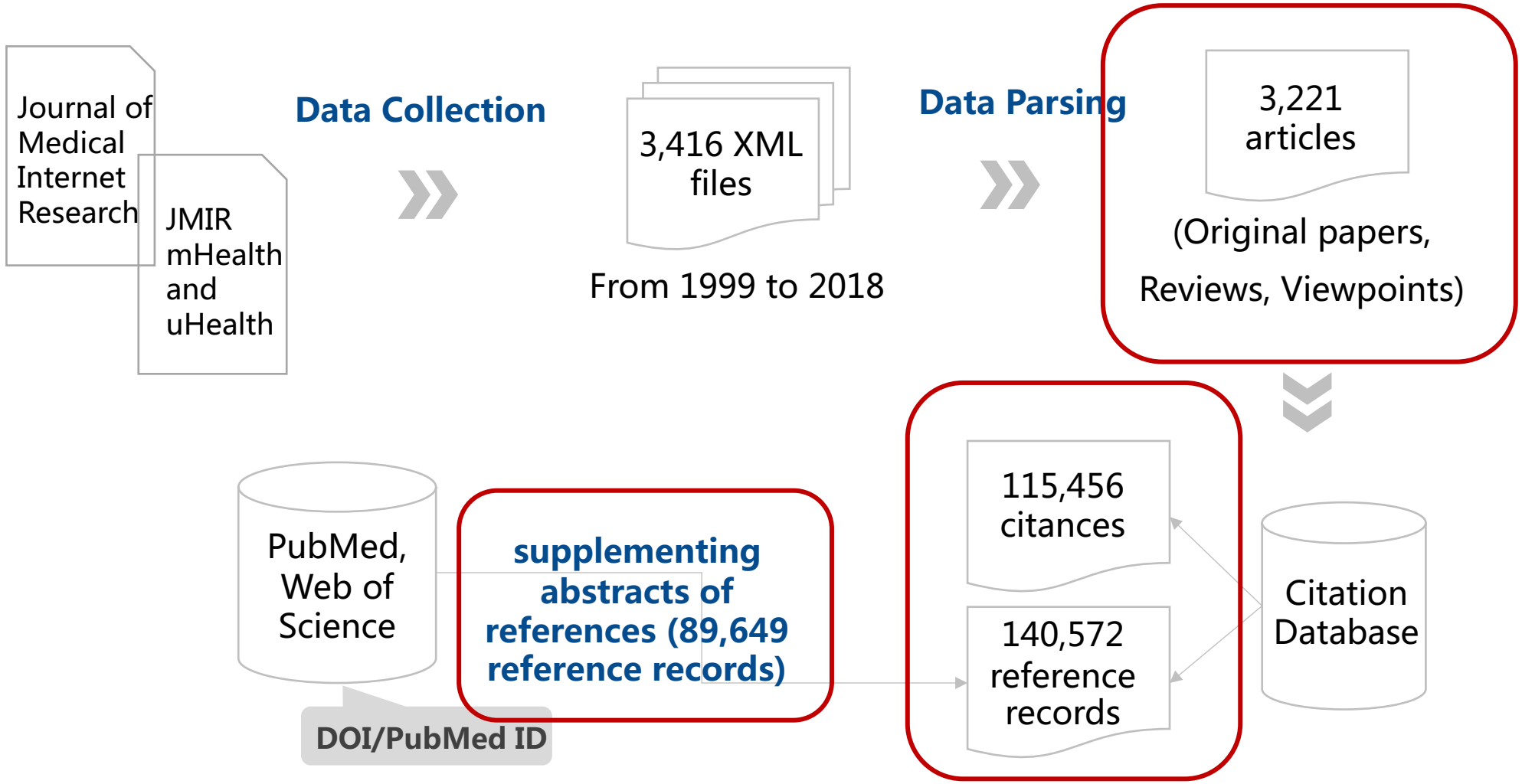
Adolescence and young adulthood are defined by developmental processes that mark increased susceptibility to risk-taking behaviors, including substance use [1-4]. In tobacco control, prevention efforts have shifted from individual and group-level interventions to population-based approaches, including policy and mass media efforts to reduce the appeal and accessibility of tobacco products to young people [5]. Concurrently, state-level cannabis policies in the United States have aimed to liberalize the accessibility of cannabis products, though there have been few state-level prevention campaigns. Using national surveillance data across states has been the standard approach to evaluate the effects of these policies on youth and young adult perceptions and behaviors [6,7]. These evaluations, which use cross-sectional data over time, may mask more nuanced trends in individual-level changes in harm perceptions and behavior and have largely failed to address spillover effects on other substance use. Novel surveillance methods that follow individuals over time and capture awareness of substance use prevention policy and communication efforts may provide a stronger basis for their evaluation.

Methodology



Data Collection

- Two high impact journals, Journal of Medical Internet Research (JMIR) and JMIR mHealth and uHealth, in the eHealth fields, were selected as our data sources.



Source Discipline Identification

- 2018 version of Essential Science Indicators **(ESI) journal list** were used to identify the disciplines of our reference journals.

7,393 distinct journal titles

» **manually compensated** 2,561 journal full titles

» matching reference journal titles with the ESI journal list

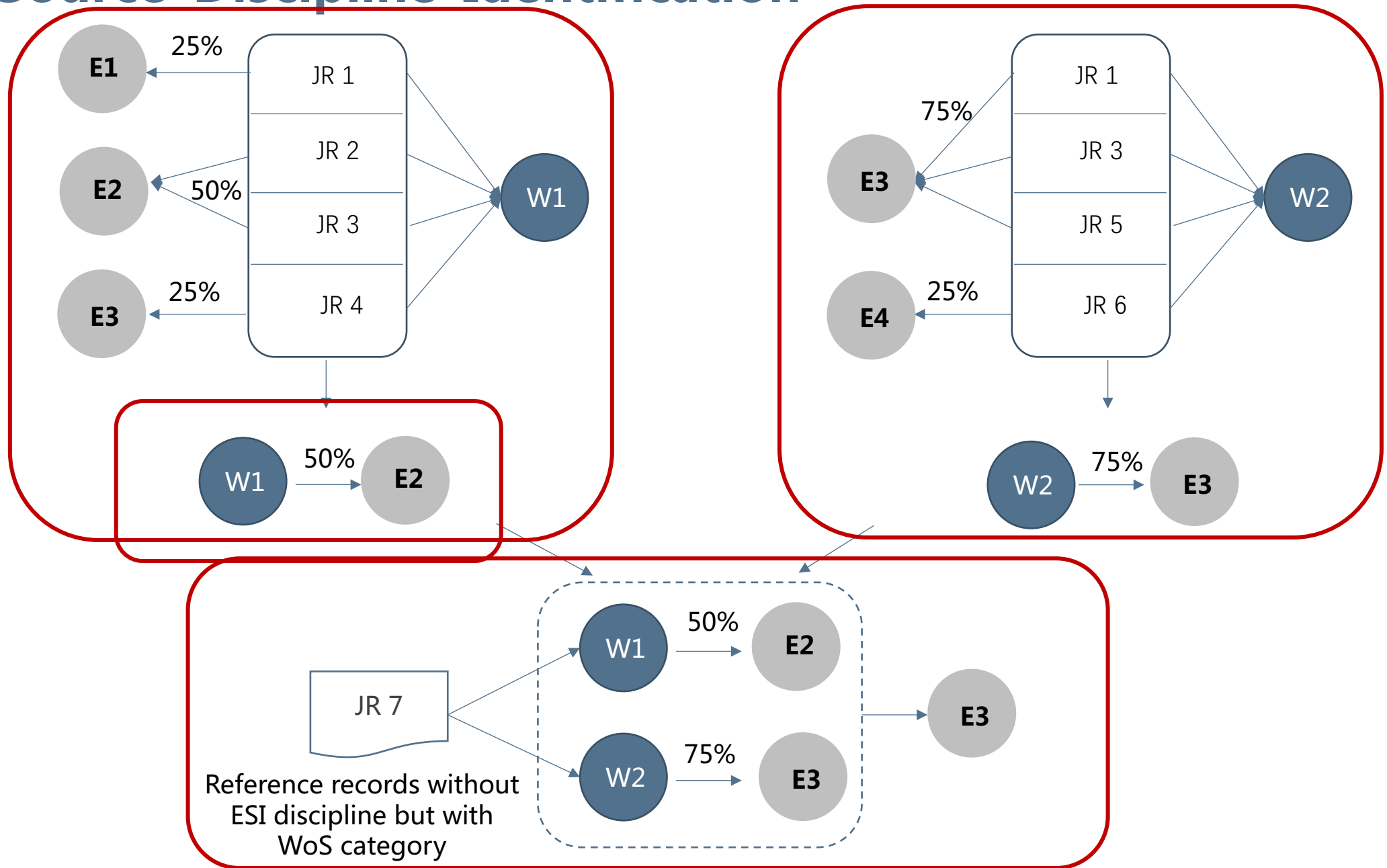
» **still 8,393 reference records without the ESI discipline information**

Web of Science (WoS) subject categories were used to infer the ESI disciplines of the not matched reference records

Probability calculation »

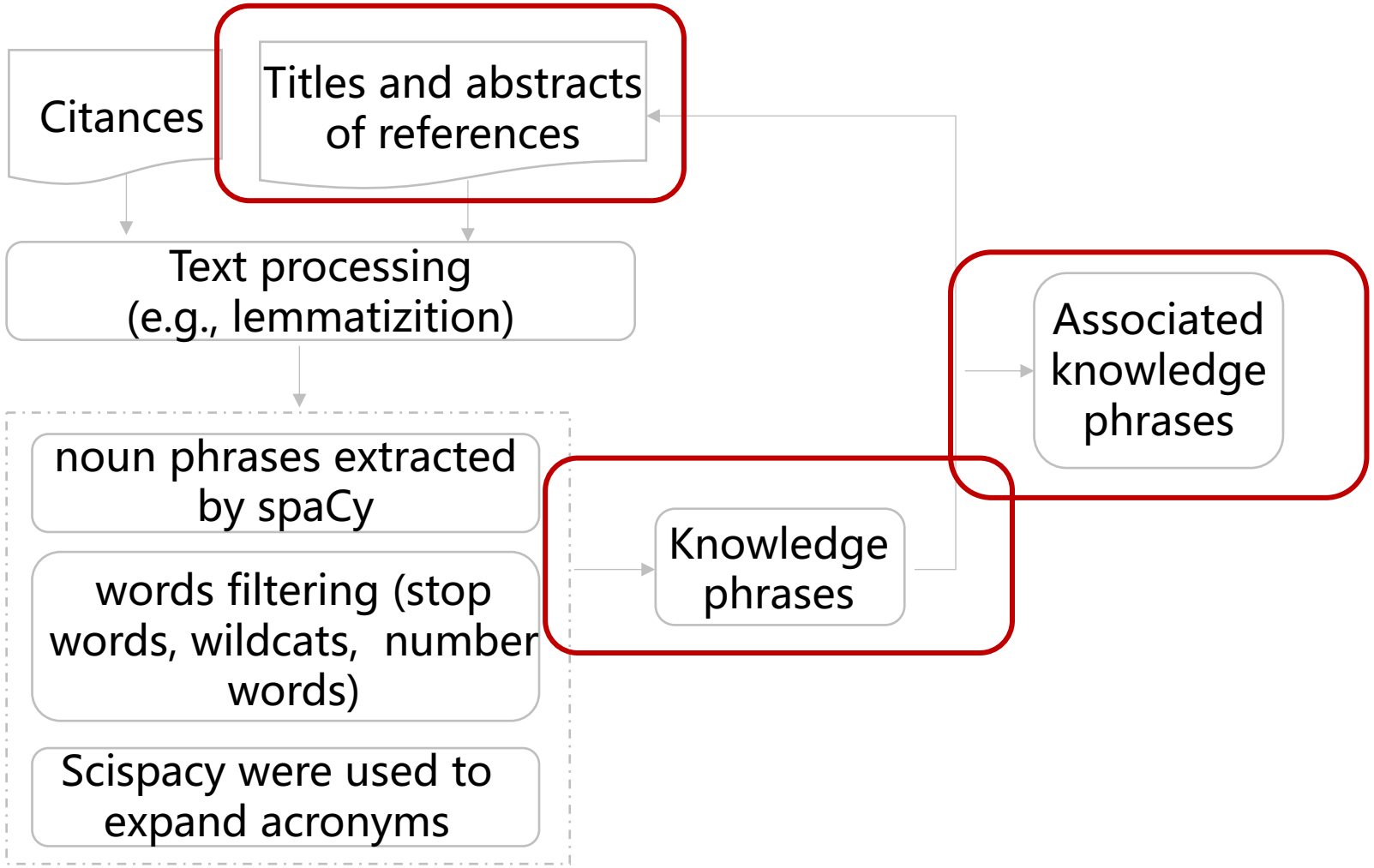
finally, approximately **94%** of journal reference records (98,685) get the discipline information

Source Discipline Identification



Associated Knowledge Phrases Process

- We defined the noun phrases that appeared in both a citance and its reference as **associated knowledge phrases**.



Associated Knowledge Phrases Process

Annotation work

01

Initializing knowledge classification framework.

- a author constructed a **preliminary classification schema** based on literature review
- randomly selected 100 knowledge phrases for trial annotation, and wrote an **annotation specification document**

02

Pre-annotation.

- two coders independently annotated **500 identical knowledge phrases**
- coder **discussed the ambiguous cases** with a professional in eHealth to reach an agreement after the annotation process

03

Formal annotation.

- two coders annotated all **24,132 unique phrases**, respectively
- maintained communication with the professional to reach a consensus during labeling

Associated Knowledge Phrases Classification Framework

Category	Description	Exemplar phrases
Research Subject	subject terms related to research problems, e.g., drugs, diseases, research areas	e.g., *information, depression, diabetes, health information*
Theory	theory related phrases, e.g., specific names of theories, frameworks, laws, etc.	e.g., *TAM, social cognitive theory, transtheoretical model*
Research Methodology	methodology used in research, including research methods, scales, guidelines, evaluation indicators	e.g., *systematic review, analysis, meta analysis, questionnaire, randomize control trial*
Technology	technique, device and system that used in research	e.g., *mobile phone, web, smartphone, app*
Human Entity	people or organizations that are targeted by the experiment	e.g., *patient, woman, child, adolescent*
Data	phrases related to dataset, data source and data material	e.g., *twitter, qualitative datum, clinical datum*
Others	other phrases that cannot be included in the above categories, e.g., geolocations, funding, or some meaningless phrases	e.g., *study, use, result, outcome, number, canada, project, USA*

Main Result 1

- The ranks of disciplines by the frequency of associated knowledge phrases are **in harmony with** the ranks by the frequency of in-text citations
- The scores of **knowledge density are slightly different** between the 10 disciplines.

TABLE 4. The frequency of associated knowledge phrases

Disciplines	Knowledge phrases	In-text citations	Knowledge density
Clinical Medicine	113,424	61,385	1.848
Social Sciences, General	46,532	28,008	1.661
Psychiatry / Psychology	31,765	19,446	1.633
Neuroscience & Behavior	5,365	3,014	1.780
Multidisciplinary	4,470	2,561	1.745
Computer Science	2,750	1,979	1.390
Immunology	2,434	1,352	1.800
Biology & Biochemistry	1,905	1,301	1.464
Pharmacology & Toxicology	1,620	876	1.849
Economics & Business	1,189	855	1.391

Main Result 2

- the frequency distribution of knowledge phrases over the categories is **heavily skewed**
- except **others**, the associated phrases of **research subject** are the most, followed by **entity** and **technology**

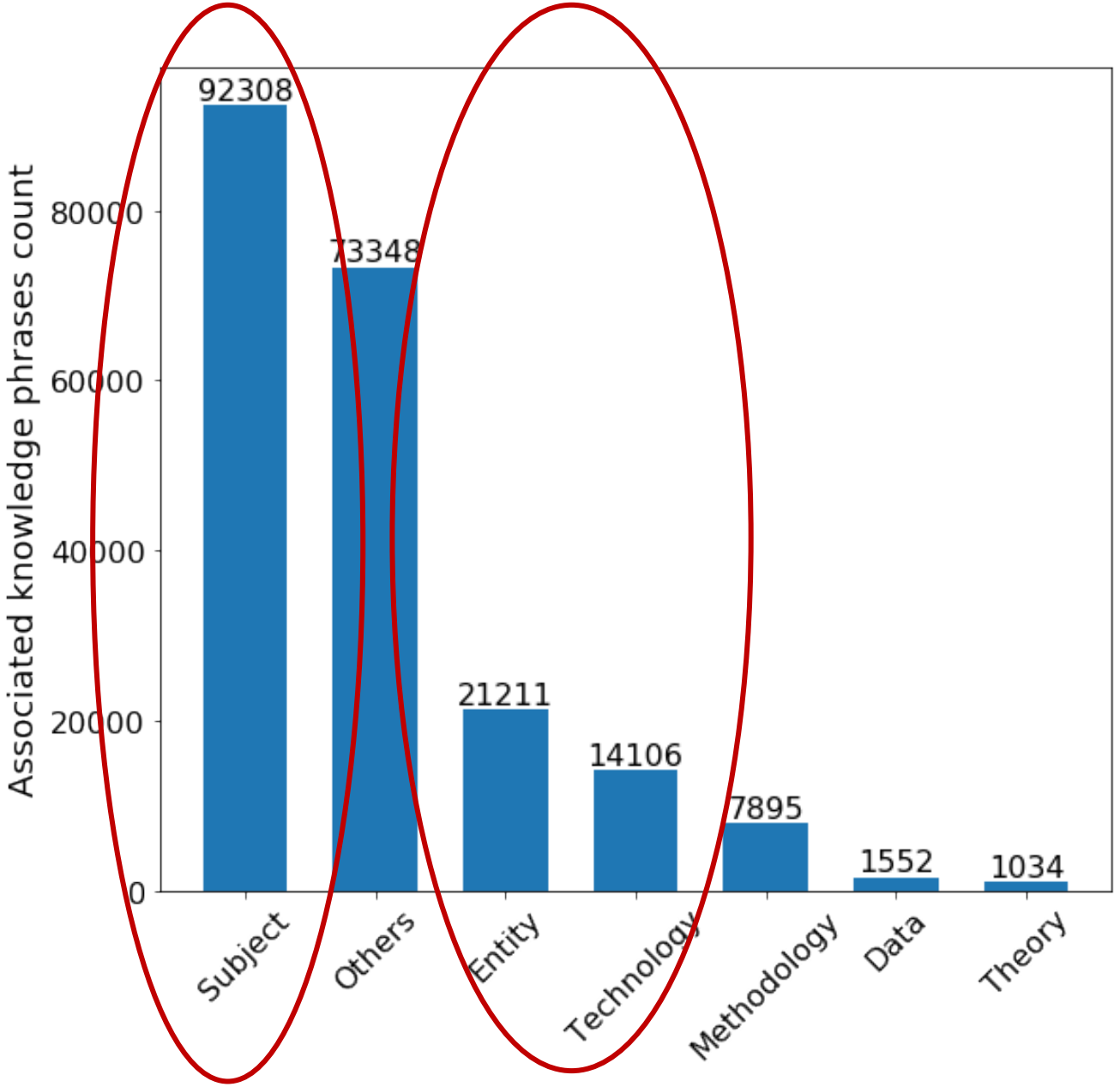


Figure 1: Frequency distribution of knowledge categories

Main Result 3

- The knowledge category distribution over different disciplines is **significantly different** (Pearson Chi Square test, p-value < 0.001)
- the proportion of theory phrases in **Economics & Business** is much higher than that in other disciplines

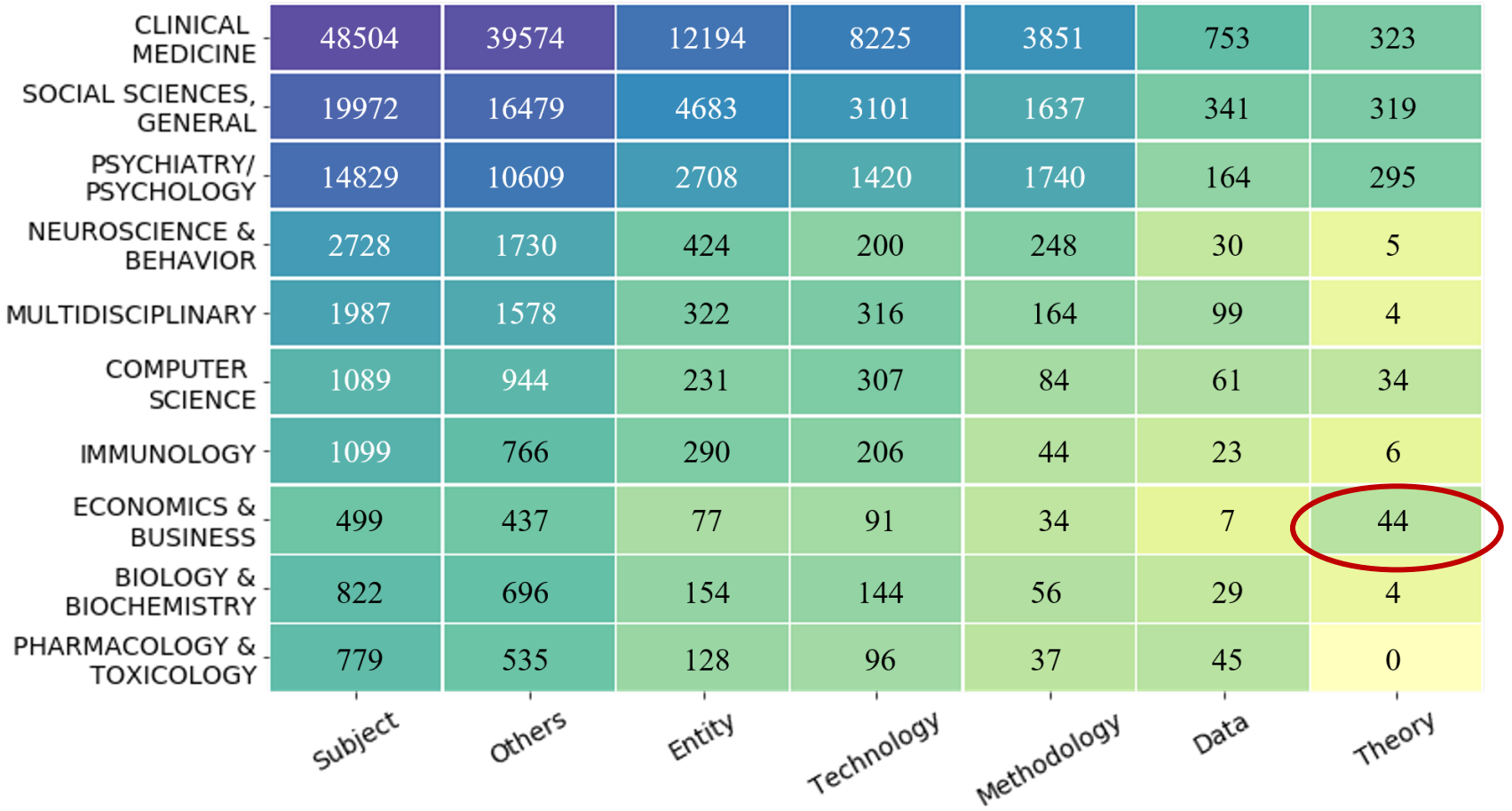


Figure 2: Frequency distribution of knowledge categories over disciplines

Discussion & Conclusion

◇ Implications

1

➢ Associated knowledge phrases can indicate the spread knowledge content, which may be useful to **generate a knowledge map** of interdisciplinary knowledge integration

2

➢ knowledge categories will be helpful to understand **the roles of different disciplines** in the knowledge integration of an interdisciplinary field

Thanks