

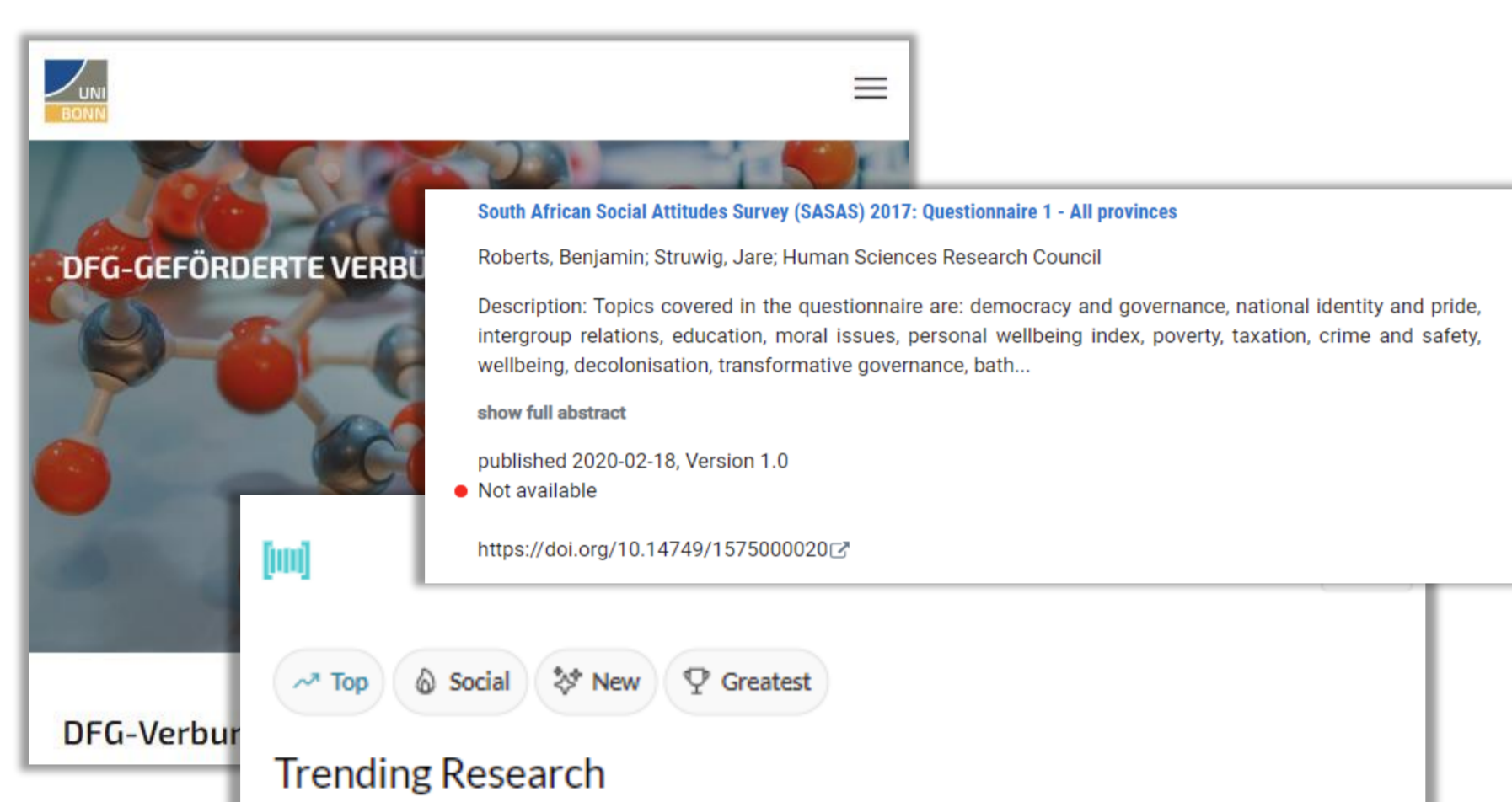
Scholarly Information Extraction for Research Data

Motivation

- Trace data usage
- Interlink scholarly artifacts
- Contextualize data (e.g. for better data search)

Extraction from Web Data

Web pages as data sources

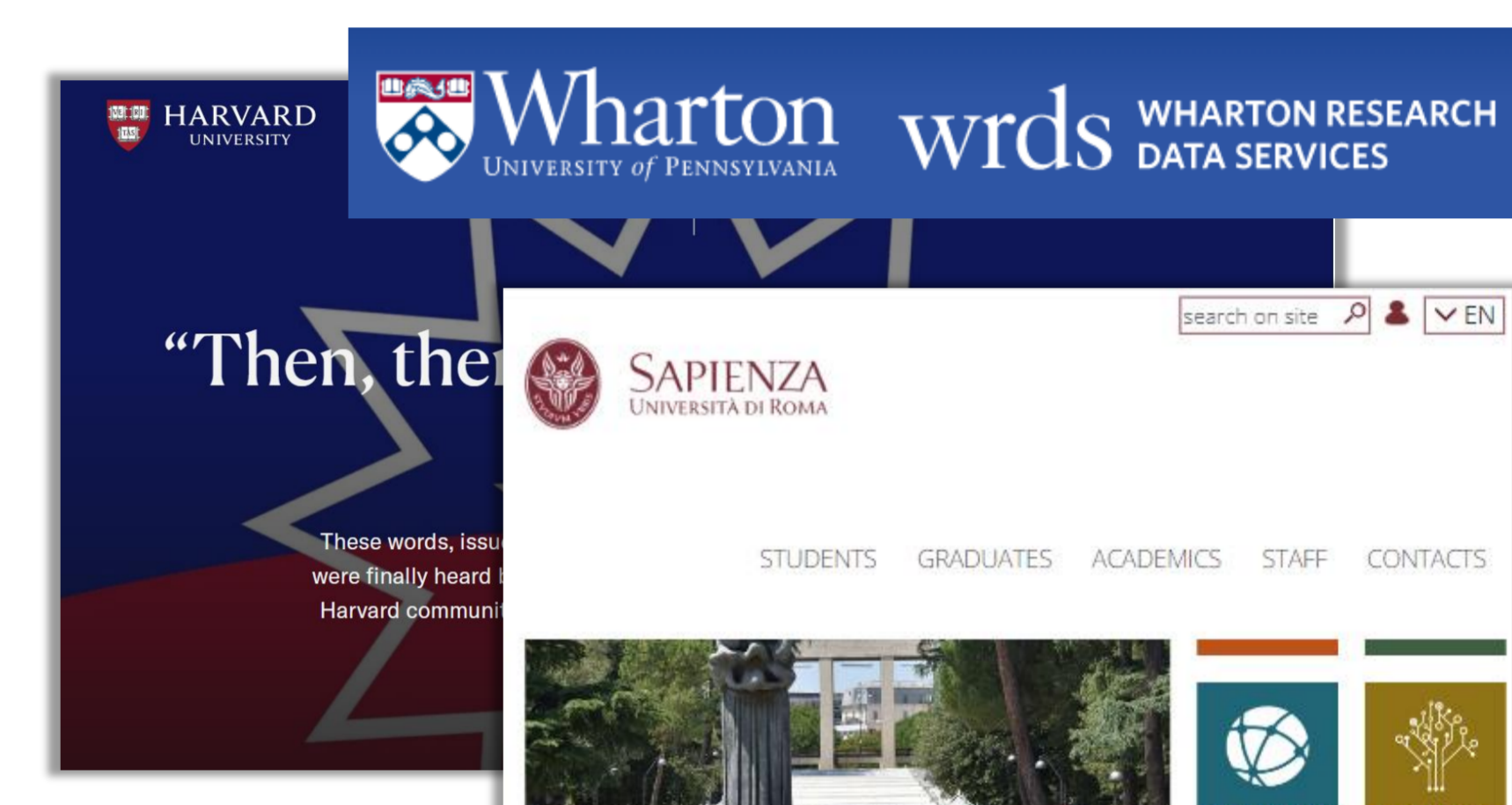


Web APIs as data sources

```
requests.get(
    "https://data.org/api/v2/r")

{"type": "dataset",
 "name": "Survey of ..."
```

Websites as data provider



Dataset level tasks

- Quality filtering
- Data fusion
- Enrichment
- Research field
- Information from Web pages

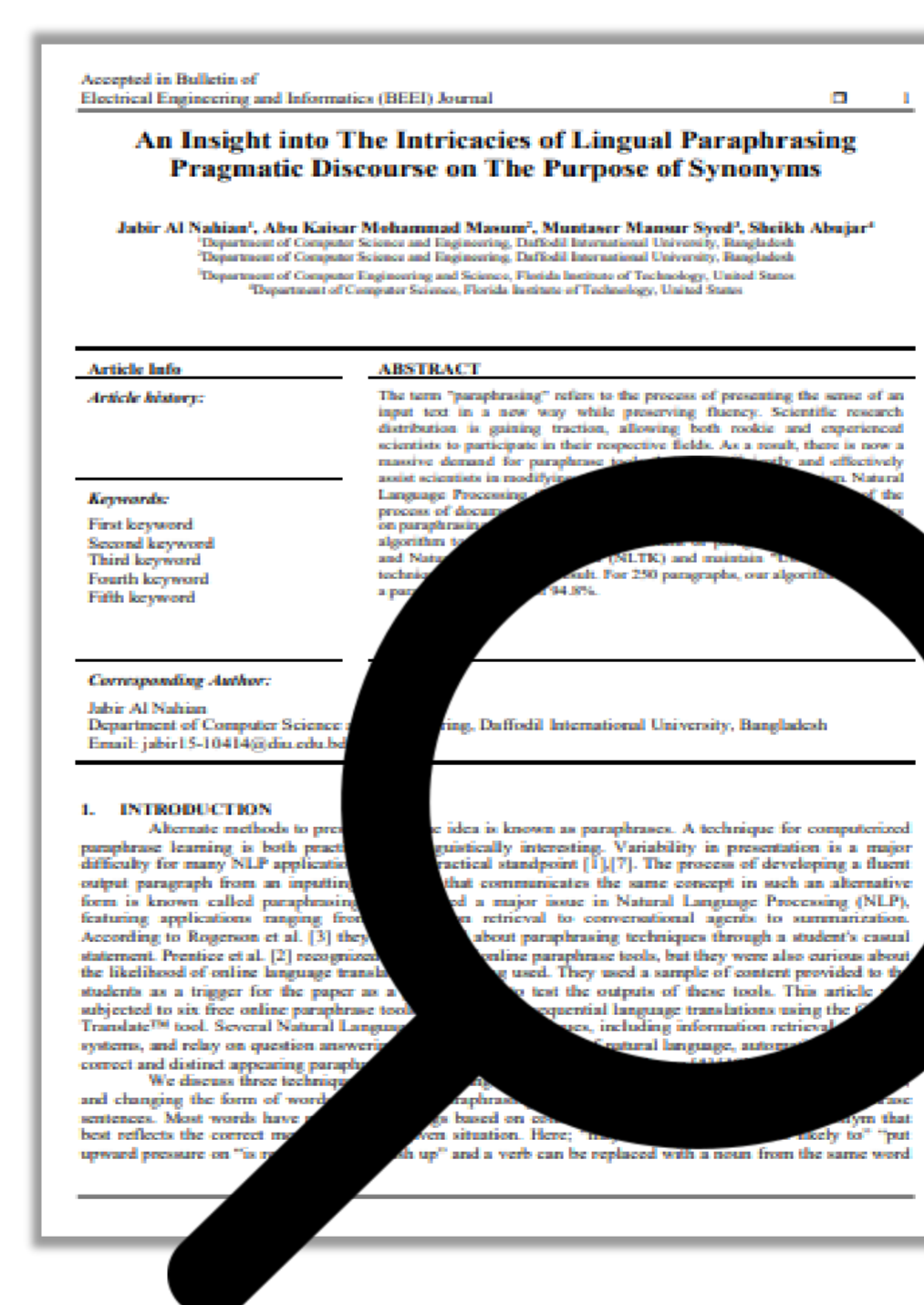
Data provider level tasks

- Quality measure
- Trustworthiness
- Source classification
- Research topic
- Datatypes

Extraction from Publications

Usage of data in research

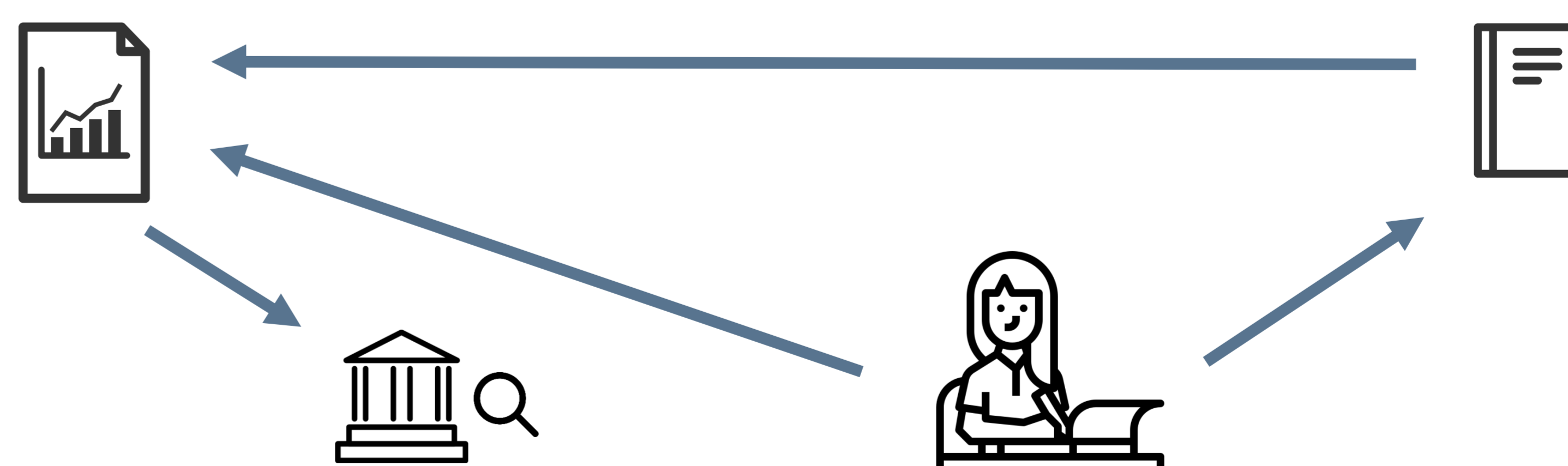
- Contextualize
 - Research field
 - Research questions
 - Methods
- Categorize utilization
- Link to known datasets



Extract research data mentions

- South African Social Attitudes Survey (SASAS)
- Gesis Panel
- ...

Integrated Research Knowledge Graph



Otto, Wolfgang et al. 2020. "Knowledge extraction from scholarly publications: The GESIS contribution to the rich context competition." In Rich search and discovery for research datasets, edited by Julia Lane, et al., 107-127. Los Angeles u.a.: Sage. <https://study.sagepub.com/richcontext>.