



LIBRARY AND INFORMATION CENTER OF  
THE HUNGARIAN ACADEMY OF SCIENCES  
DEPARTMENT OF SCIENCE POLICY AND SCIENTOMETRICS

## Classifying SSH content within the Hungarian Scientific Bibliography (MTMT)

– theory and practice –

Sándor Soós



LIBRARY AND INFORMATION CENTER OF  
THE HUNGARIAN ACADEMY OF SCIENCES  
DEPARTMENT OF SCIENCE POLICY AND SCIENTOMETRICS

1.

Research classification for National publication databases:  
Why?



- Information retrieval
- Representations of scholarly activity and performance
  - Reporting etc
  - Assessment on productivity
- Normalization of performance indicators:
  - Field-normalized citation impact measures

Are these purposes interchangeable? Results in a uniform classification?



2.

Research classification for national publication databases:  
How?



- **Theoretical/Conceptual basis:**
  - Organizational
  - Cognitive
  - Qualificaton-based
- **Methodological aspect:**
  - Mapping via ontologies, thesauri, classifications (external sources)
  - Empirical (data-driven): citations, (co-)authors, referencees, text descriptors
  - Indexer-based (author self-assignment)
- **Technical:**
  - Algorithmic
  - Manual
- **Agency:**
  - Indexer-based
  - Author-based
- **Aggregation level:**
  - Organization
  - Author
  - Paper



LIBRARY AND INFORMATION CENTER OF  
THE HUNGARIAN ACADEMY OF SCIENCES  
DEPARTMENT OF SCIENCE POLICY AND SCIENTOMETRICS

3.

Hungarian Scientific Bibliography (MTMT)



## 4.4.6. *MTMT in Hungary*

The main purpose for MTMT is bibliometrics-based research evaluation; in addition MTMT is intended as a general source of information on research in Hungary.

For MTMT, inclusion criteria are specified in the law on the Hungarian Academy of Sciences. In general, all research output is included in MTMT using a detailed classification of output types. However, further criteria apply for subsets that are used in research evaluation and/or funding allocation. For those subsets, all output from research funded by public funds should be included in MTMT. Nevertheless, it is possible to include unaffiliated publications (e.g. researchers in retirement or without an affiliation) and publications resulting from research funded by other sources.

The main focus is on journal articles, books, book chapters, and conference proceedings. However, MTMT stores data also on research data, engineering and artistic products, and other types of research output (see [Holl et al. 2014](#)).

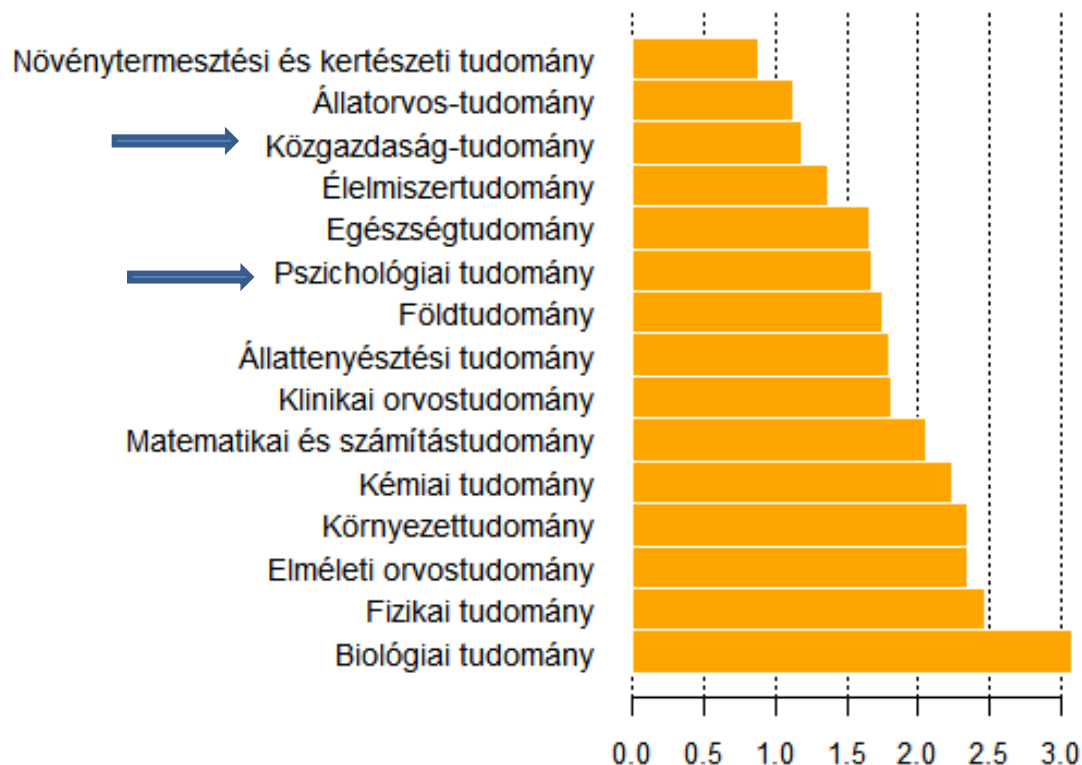


4.

Attempt 1: classification by external sources

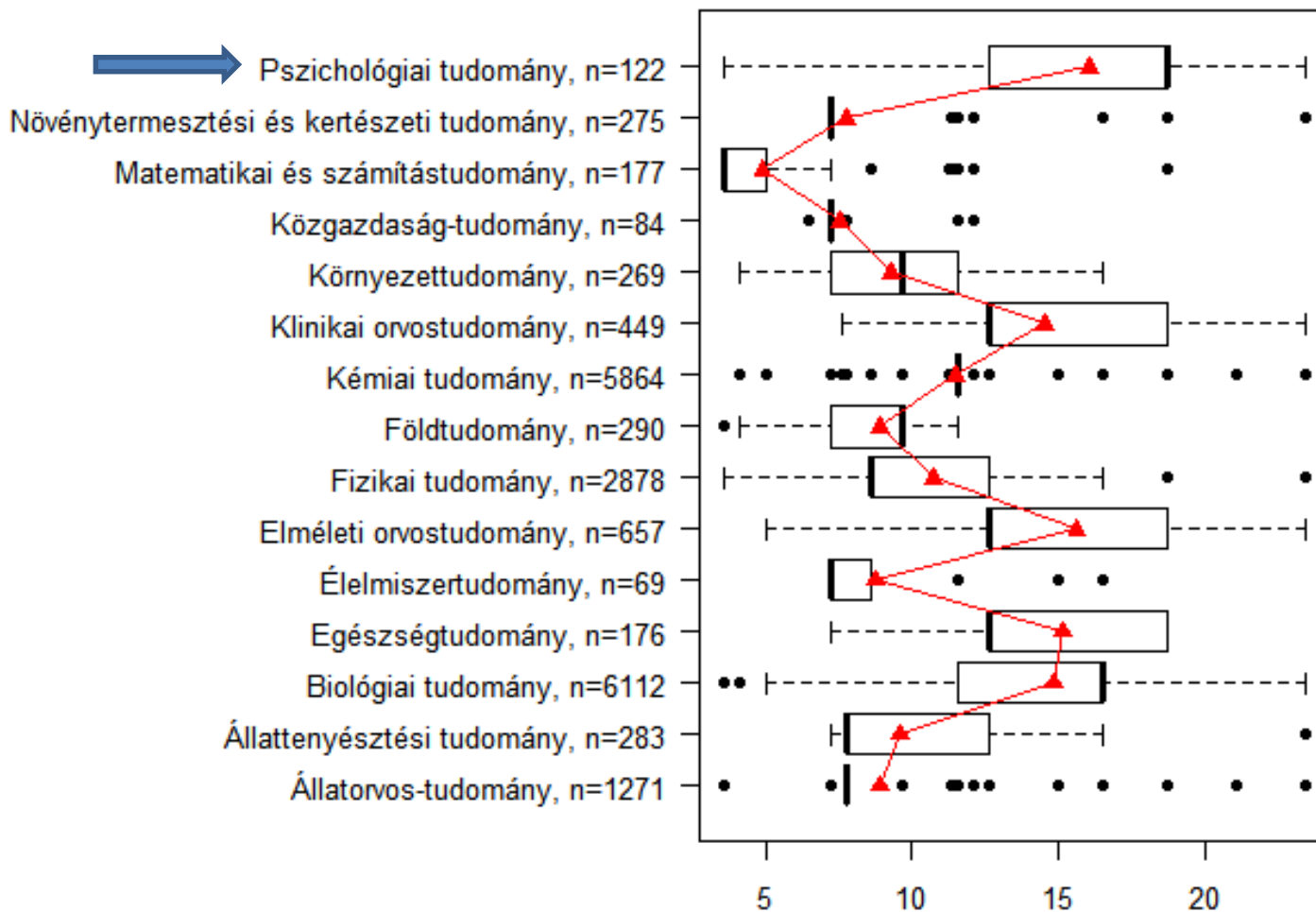


- Comparing a cognitive and a qualification-based classification imposed on the MTMT-WoS intersection (paper vs. Author level)
- ESI (WoS) vs. MAB (Hungarian Accreditation Committee): entropy of MAB categories





- Distribution of citation benchmark values in quali-based categories





LIBRARY AND INFORMATION CENTER OF  
THE HUNGARIAN ACADEMY OF SCIENCES  
DEPARTMENT OF SCIENCE POLICY AND SCIENTOMETRICS

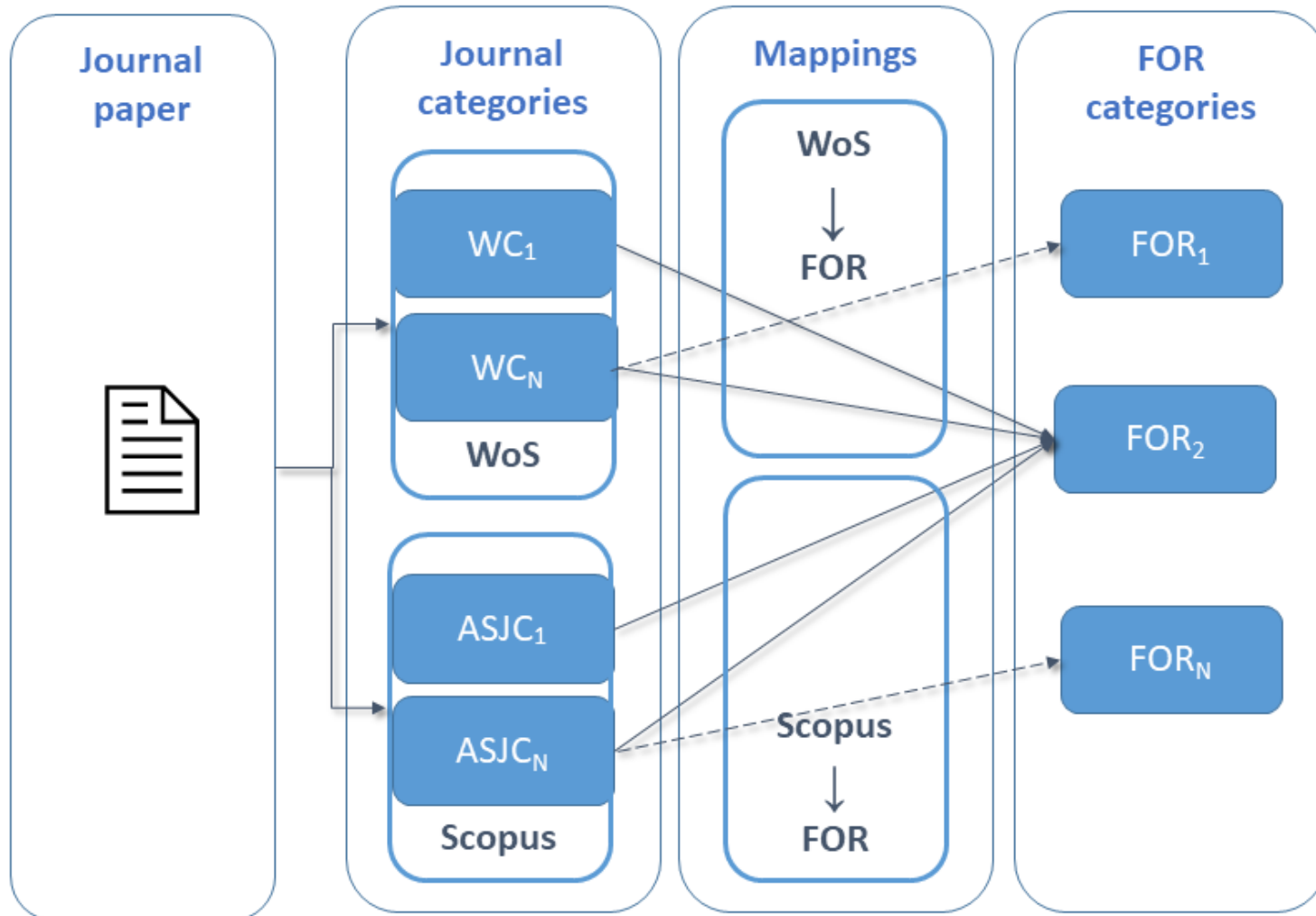
4.

Attempt2: classifying SSH content

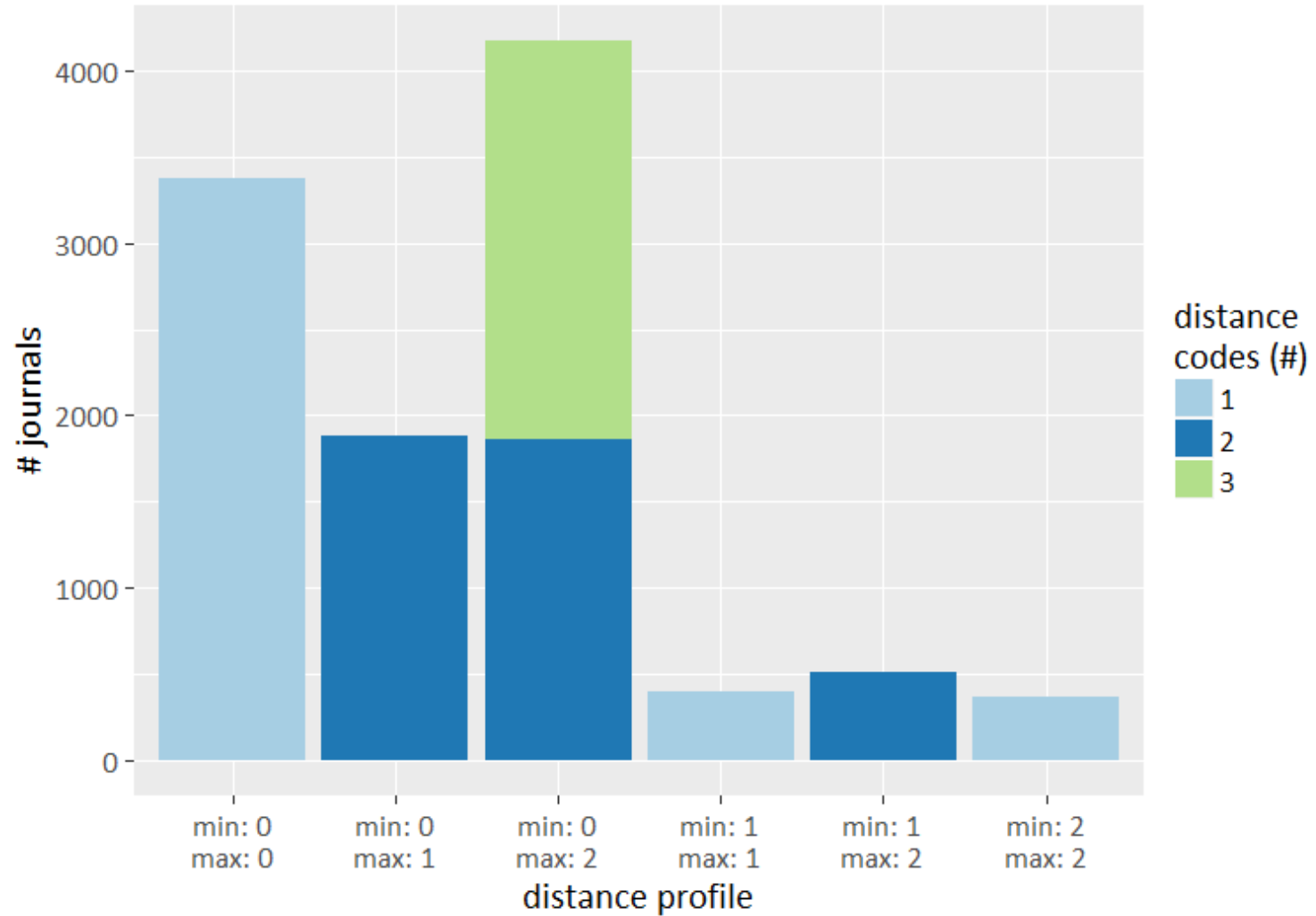


- Objective: classifying SSH content using the OECD FOR (Fields of Science) system (algorithmically)
- Option1: mapping-based
  - Only applicable to MTMT intersections with WoS or Scopus
- Option2: empirical (for non-WoS/Scopus content)
  - References (bibliographic coupling): not available in MTMT
  - Citatios (co-citation analysis): citation numbers are very low and many zeros
  - Co-authorship: rare phenomenon in many SS and H fields...
  - Still need a mapping to FOS

# FOR mapping problems



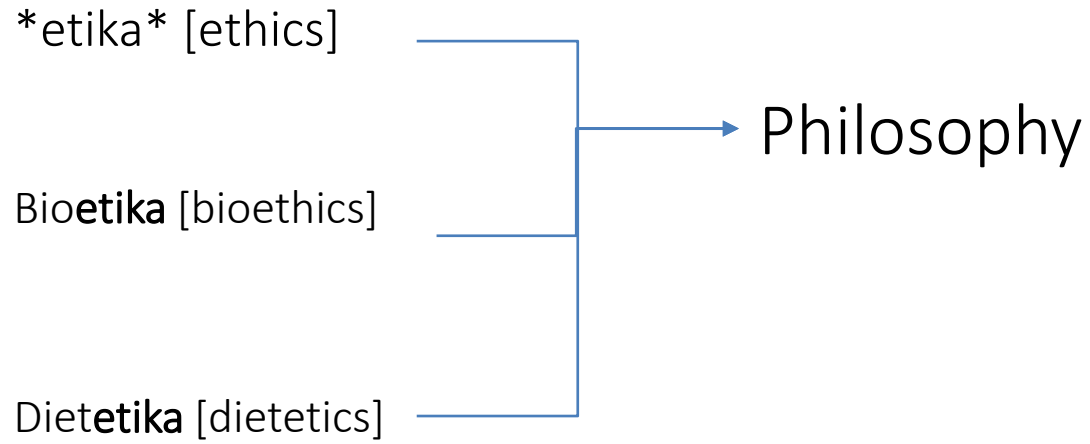
# FOR mapping problems



A hybrid approach:

- **Mapping-based part**
  - Journal content
  - External classifications: ERIH, national library
  - Manual mapping of external classifications to FOS categories
  - Mapping journal content to FOS categories
- **Text mining part**
  - Non-Journal content
  - Using a detailed „ontology” with Hungarian subject terms (committee of HAS)
  - Mapping the upper level of this ontology to FOS
  - Classifying items at **author level** via occurrence of subject terms in author-provided field descriptions (MTMT field, no controlled vocabulary)
  - Mapping items to FOS categories





| FOS_field                            | Total number of publications (#) | Number of publications in the main national language(-s) (#) | Number of English language publications (#) | Number of articles in books (#) | Number of monographs (#) | Number of articles in conference proceedings |
|--------------------------------------|----------------------------------|--------------------------------------------------------------|---------------------------------------------|---------------------------------|--------------------------|----------------------------------------------|
| <b>5. Social Sciences</b>            | 39,414                           | 26,814                                                       | 10,159                                      | 14,773                          | 4,915                    | 1,799                                        |
| 5.01 Psychology                      | 4,354                            | 2,487                                                        | 1,760                                       | 1,388                           | 349                      | 523                                          |
| 5.02 Economics and business          | 9,761                            | 6,141                                                        | 3,032                                       | 3,025                           | 1,126                    | 371                                          |
| 5.03 Educational sciences            | 8,725                            | 6,581                                                        | 1,642                                       | 3,662                           | 1,155                    | 498                                          |
| 5.04 Sociology                       | 3,897                            | 2,752                                                        | 954                                         | 1,096                           | 443                      | 138                                          |
| 5.05 Law                             | 9,211                            | 6,944                                                        | 1,545                                       | 3,989                           | 1,290                    | 144                                          |
| 5.06 Political science               | 3,567                            | 2,695                                                        | 714                                         | 1,013                           | 446                      | 43                                           |
| 5.07 Social and economic geography   | 2,199                            | 1,300                                                        | 766                                         | 1,026                           | 273                      | 74                                           |
| 5.08 Media and communication         | 1,965                            | 1,381                                                        | 419                                         | 621                             | 253                      | 90                                           |
| 5.09 Other social sciences           | 1,126                            | 949                                                          | 149                                         | 186                             | 111                      | 4                                            |
| <b>6. Humanities</b>                 | 33,250                           | 23,287                                                       | 4,946                                       | 11,346                          | 4,590                    | 522                                          |
| 6.01 History and archaeology         | 15,952                           | 11,601                                                       | 2,095                                       | 5,528                           | 2,312                    | 183                                          |
| 6.02 Languages and literature        | 17,037                           | 11,642                                                       | 2,577                                       | 5,798                           | 2,260                    | 311                                          |
| 6.03 Philosophy, ethics and religion | 2,808                            | 1,962                                                        | 546                                         | 960                             | 404                      | 39                                           |
| 6.04 Art                             | 1,205                            | 776                                                          | 186                                         | 342                             | 174                      | 8                                            |
| 6.05 Other Humanities                | 911                              | 626                                                          | 121                                         | 232                             | 110                      | 1                                            |



- Regarding SSH content
- Empirical approaches: missing or scarce metadata, citation behavior does not allow for field delineation
- Mapping—based approaches: available mappings reflect the characteristics of national communities instead of field delineations)
- Pragmatic attempts (combining manual methods and data-mining heuristics): missing coherence