

# Opportunities and hurdles in WoS raw data

Workshop for the BMBF funding line *Quantitative research on the science sector*  
by KU Leuven, Humboldt-Universität zu Berlin and DZHW

<https://wiho-fdm.github.io/>

## Motivation and Scope:

Bibliographic databases constitute (partial) mappings of the science system and as such reduces the complexity and heterogeneity of the underlying science system via normalisation into predefined categories. Simultaneously none of the bibliographic databases feature complete coverage. Rather, all of them are to some extent biased, e.g. towards coverage of a certain set of disciplines, languages or regions. These differences in coverage are not random, but rather the result of different ways and positions “what” should be included in such databases. Collection and curation methods of these databases have changed over time, which, for instance, can introduce issues when performing longitudinal analyses. Furthermore, the emergence and inclusion of identifiers in these databases, such as DOIs, ORCID IDs, ResarcherIDs etc., may lure researchers into assuming near complete coverage of such identifiers or that these identifiers are not prone to error. However, knowledge about these and other characteristics mostly results from experience with the data and, given the large volume of data conveniently arranged in rectangular (SQL) tables of bibliographic databases such as the Web of Science, might be hidden at first sight affecting both analyses as well as conclusions drawn from the data.

The workshop will offer an opportunity to share experience on the observed imperfections in light of the research interests of the funded projects. In the first part we will offer an inner perspective on the Web of Science focusing especially on structural breaks, an outer perspective comparing the Web of Science with the Scopus database and last an outlook on how the outcome of bibliometric analysis is shaped by measurement decisions defined by the researcher. In the second part we like to invite the participating projects to present their research interest given any observed data issues and discuss jointly in small groups potential remedies.

## Organisation:

- Time: 12 March 2020, 10:00h – 17:00h
- Location: DZHW Berlin  
(Schützenstr. 6a, 10117 Berlin, <https://www.dzhw.eu/en/kontakt>)
- Workshop language: *English*
- Places: *max. 12*
- Registration deadline: 20 February 2020
- Registration: [stahlschmidt@dzhw.eu](mailto:stahlschmidt@dzhw.eu) (via *First come, first served*)
- Travel expenses: [wihoforschung@dlr.de](mailto:wihoforschung@dlr.de)

## Programme:

10:00h – 12:00h	Input by KU Leuven, Humboldt-Universität zu Berlin and DZHW
12:00h – 12:45h	Lunch break
12:45h – 14:30h	Input by participants
14:30h – 14.45h	Coffee break
14:45h – 16:45h	Breakout groups
16:45h – 17:00h	Concluding discussion