Collecting Datasets by Analyzing References in the zbMATH Database

<u>Moritz Schubotz</u>[1,*], **Olaf Teschke**[1], **Fabian Müller**[1], **Klaus Hulek**[2]

[1]*Department of Mathematics, FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Berlin, Germany*
[2]*Institut für Algebraische Geometrie, Gottfried Wilhelm Leibniz Universität Hannover, Hannover, Germany*
[*]Email: `moritz.schubotz@fiz-karlsruhe.de`

References to non-traditional forms of publications like software packages, datasets, and the like, are often constructed in a non-normalized ad-hoc fashion, hurting both the findability of the result as well as the attribution of research output to a researcher. In this talk, we present a preliminary analysis of the references used in the publications indexed by zbMATH. In particular, we inspect references that do not point to classical research papers. Therefore, we apply a twofold approach. For one, we use random sampling, for explorative analysis of the dataset. In a second step, we apply heuristics and clustering to quantify the findings of the explorative analysis.