Speaker: Prof. Dr. Ad Backus

Title: Diversity and inclusion in traditional multilingualism research

My primary research has largely been in a relatively traditional field of which the research questions and methods seem little affected by digitalization. Multilingualism research largely wants to know how bi- and multilinguals learn and use their languages and what the implications are for these languages. There is a certain sense of being involved in social justice work because multilinguals tend to be on the wrong end of the privilege continuum, often belonging to dispossessed, displaced and colonized communities. However, this sense often only translates into raising awareness and recognition ('multilingualism is the norm and small or immigrant languages are just as valuable as major languages'), sometimes laying groundwork for social justice work ('low-status languages are sometimes not transmitted to children anymore and this injustice can to be combatted through revitalization efforts'), but more often than not the research questions remain firmly rooted in fundamental issues of cognition ('how is multilingualism represented in the mind?') and linguistics ('how do languages change as a result of foreign influence?'). This makes the field ill-prepared, I think, for the algorithmic era. Multilingual data collection tends to be costly: there are no corpora, there are no written data as language mixing is typically a spoken phenomenon, there are few opportunities for automatic transcription, etc. At the same time, methodological requirements tend to favor ever larger data collections, which makes people gravitate towards the larger languages, where data can be found on the Internet or where at least analysis can benefit from many of the digital tools made available through NLP. As an editor of an international journal, I have noticed an enormous representation of languages such as English, Chinese, Spanish, Japanese and French. I will sketch this background and then list some open questions regarding this issue, in the interest of the more general question whether the ever-increasing hegemony of major languages is a necessary by-product of digitalization.