

Making Dynamic Data Citable: Approaches to Data Citation within the Context of the RDA Working Group

RAUBER, Andreas^{1*}

¹Vienna University of Technology

Being able to reliably and efficiently identify entire or subsets of data in large and dynamically growing or changing datasets constitutes a significant challenge for a range of research domains. In order to repeat an earlier study, to apply data from an earlier study to a new model, we need to be able to precisely identify the very subset of data used. While verbal descriptions of how the subset was created (e.g. by providing selected attribute ranges and time intervals) are hardly precise enough and do not support automated handling, keeping redundant copies of the data in question does not scale up to the big data settings encountered in many disciplines today. Furthermore, we need to be able to handle situations where new data gets added or existing data gets corrected or otherwise modified over time. Conventional approaches, such as assigning persistent identifiers to entire data sets or individual subsets or data items, are thus not sufficient.

In this talk we will review the challenges identified above and discuss solutions that are currently elaborated within the context of the working group of the Research Data Alliance (RDA) on Data Citation: Making Dynamic Data Citable. These approaches are based on versioned and time-stamped data sources, with persistent identifiers being assigned to the time-stamped queries/expressions that are used for creating the subset of data. We will further review examples of how these can be implemented for different types of data and see how this fits into the larger context of activities on Data Citation.

Keywords: Research Data Alliance, data citation, dynamic data, information technology, interoperability