

## Why Data Citation Needs Provenance

In today's era of big data-driven science, an increasing amount of information is being published as curated online databases and retrieved by queries, raising the question of how query results should be cited. Citations traditionally contain information about the provenance of the research product (e.g. the authors) among other things. The connection between provenance and citation becomes even more interesting in the context of data citation, where it is infeasible to associate citations with every possible query to the database. In this case, we must reason over the citations associated with a small set of frequent queries ("citation views") to construct citations to general queries. The manner in which this is done has interesting parallels to "where" and "why" provenance reasoning in databases, and will be the focus of this talk.