

Data Citation: Giving Credit Where Credit is Due

Susan Davidson, University of Pennsylvania

Most information is now published in complex, structured, evolving datasets or databases. As such, there is increasing demand that this digital information should be treated in the same way as conventional publications and cited appropriately. While principles and standards have been developed for data citation, they are unlikely to be used unless we can couple the process of extracting information with that of providing a citation for it.

I will discuss the problem of automatically generating citations for data in a database given how the data was obtained (the query) as well as the content (the data), and show how the problem of generating a citation is related to two well-studied problems in databases: query rewriting using views and provenance.