



Use cases for identifiers and data citation — Separating concerns.

Peter Fox (1) and Mark Parsons (2)

(1) Rensselaer Polytechnic Institute, Tetherless World Constellation, Troy, United States (pfox@cs.rpi.edu), (2) Rensselaer Polytechnic Institute, Institute for Data Exploration and Applications, Troy, United States (parsom3@rpi.edu)

Over the past two decades, the data management community has worked to codify many aspects modern data management. Our interest in clarifying and promoting data citation policies and practices naturally involves associated technologies such as persistent identifiers and automated indices. While the use of dataset identifiers is on the rise; the datasets identified are rarely cited formally for myriad reasons. Data managers often expect too much from data citation, assuming false parallels to literature citation; i.e, the use cases for literature citation and data citation are often different, and thus identifier utilization differs. We call for a closer examination of the purpose and practice of data citation, and relation to identifiers. We will present several relevant use cases that convolve citation and identifiers and suggest possible approaches for addressing these use cases. We seek to engage the community in a more exploratory conversation on the topic that separates concerns appropriately.