Managing Dataset DOIs and Versions in a Changing Archive

Steven Worley, Bob Dattore, and Zaihua Ji
NCAR, Boulder, United States (worley@ucar.edu)

The Digital Object Identifier (DOI) has a proven value in textual publications, because it enables rapid access to cited references. There is an important and growing movement to do the same with reference worthy datasets. If datasets are static (fixed content) the application of a DOI is relatively straightforward and very similar to textual publications. However, datasets are often dynamic, with new versions, archive fixes, and time series extensions as examples. This is a complicating challenge for data management systems. For a data DOI to be effective as a citable reference in publications, the data used for any particular research study must be precisely reproducible regardless of the dynamic adjustments that occur to the dataset.

This paper will discuss the implementation for dataset DOIs in the Research Data Archive at NCAR. We will present the methods used to manage the dataset DOIs and versions as well as the use cases that formed the foundation for the system requirements.