Session Summary: The RDAP12 Data Citation Panel
by Matthew S. Mayernik

Joe Hourclé, from the NASA Solar Data Analysis Center (SDAC), moderated the data citation panel and provided the first presentation. Many institutions across academic disciplines are promoting data citations. As this term suggests, a data citation is a citation included in a reference list of a published article that formally cites data that led to a given research result. In his panel introduction, Joe noted that data users who write research papers currently cite data use in a variety of ways, including 1) acknowledging data providers in the body text or in the acknowledgements section of a paper; 2) citing papers that describe the data collection instruments; 3) citing papers that describe the data and provide findings; and 4) citing the actual data themselves. Formal data citations (#4 in the above list) are becoming more common, but are still more the exception than the rule.

Paul Uhlir, the other speaker in the data citation panel, spoke about his work relating to the International Council for Science Committee on Data for Science and Technology (ICSU CODATA) task group on data citation standards and practices. The ICSU CODATA task group has a number of activities ongoing. They sponsored a workshop in August 2011 on data citation and attribution, the presentations from which can be found at http://sites.nationalacademies.org/PGA/brdi/PGA_064019. A report from this workshop will be published later in 2012. In addition, the task group is developing a white paper that will be based on a survey of data citation and attribution practices and will be published at the end of the 2012 calendar year. The white paper will discuss the importance of data citations, current uses, emerging formal standardization proposals and best practices development, emerging principles for data citation, tool and infrastructure needs and developments, cultural challenges and opportunities, and open research questions.

During the open discussion, the panelists received multiple questions relating to standardization. Can data citation practices be standardized at an international level given the variability of data types, repositories and uses? How important will it be to have standardized identifiers for datasets and dataset authors? These questions have no concrete answers currently, but different organizations are emerging to promote particular approaches, including DataCite (http://datacite.org/) and ORCID (http://about.orcid.org/) for data citations and author identifiers respectively. Other comments focused on the importance of having data citation initiatives integrated into new and existing citation style guides and citation management software.

Matthew Mayernik is a research data services specialist in the library of the National Center for Atmospheric Research (NCAR)/University Corporation for Atmospheric Research (UCAR). He can be reached at mayernik<at>ucar.edu.