



Scientific Data Management: Options for Research Projects

M. Kaminski, C. Judy, F. Fetterer and D. Scott

National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA (marilynk@nsidc.org / Fax: +1 303-492-2468)

Management of digital scientific data is an increasingly complex and costly issue for all data providers and data archives. Small research projects face special challenges when they have a mandate for data management but limited or no budget for that purpose. Since these projects cannot know the ultimate value of their resulting data products a priori it is difficult to initially justify funding for comprehensive management of the data. Several steps can be taken to assure short-term data preservation along with basic data discovery and access while working to establish need and requirements for longer-term preservation.

The National Snow and Ice Data Center (NSIDC) provides data management services for a wide variety of projects. Data management can encompass a range of services depending on the requirements of each project. At a minimum these services must include basic data archival, discovery and access, ensuring preservation of all data resulting from documentable scientific efforts. At that level, supportive information and services may be dependent upon the scientist and the user community interacting in an unfunded, Web-based wiki or forum. At higher levels a broader set of requirements covering data packaging, format definition, archive media, documentation, and metadata can be defined for a wide range of available services. Services may include user support, accessibility, and end-user transformations such as gridding, reprojection, subsetting, and mapping. Effective use of existing standards, hardware and software technology, best practices, and automation can frequently mitigate the concern of smaller projects that their data will not be adequately served within their budgetary guidelines.

Planning for data management should be a collaboration between the data provider and the data manager. NSIDC's experience with data management and scientific research projects encompassing archival, documentation, data discovery and access strategies, and user support, as well as development and use of many data tools and services, has led to identification of a variety of options that can be effective in guiding this collaboration. By carefully examining options for management of the data in the short term (during the life of the research project), preservation and access is assured. As projects progress and growing interest in specific data is observed, level of service may increase. Further collaboration leads to selection of effective strategies for long-term preservation, enabling scientific extension of the data and compounding its inherent value.