

Public Use Datasets

The National Institutes of Health (NIH) expects researchers and investigators to make their data widely and freely available to the public. "Sharing data reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, makes possible the testing of new or alternative hypotheses and methods of analysis ..." In general, NIH expects the timely release and sharing of data to occur no later than the acceptance for publication of the main findings from the final dataset. However, there are several factors to consider when creating public use datasets. Many investigators do not have the expertise or resources to create them, or the knowledge of how to release them to the public. Moreover, the intricacies of de-identifying participant data and safeguarding protected health information, as well as creating user-friendly data files and assembling the corresponding documentation, can be surprisingly difficult.

This session will discuss the NIH requirements and expectations for sharing data and creating public use datasets. It will present case studies outlining the obstacles faced during their creation and the solutions found to overcome them. It will also review repositories for public use datasets and policy issues surrounding their use. Speakers will include experts from the NIH and academia who have the perspective and the practical experience in satisfying the NIH policy.

Speakers: Annette McDaniel, Duke Clinical Research Institute
Rebekah Rasooly, NIDDK/NIH
Paul Van Veldhuisen, The EMMES Corporation
Paul Wakim, NIDA/NIH

Organizer: James Rochon, Rho