



# Session 5: Public Use Datasets

*James Rochon, PhD  
Rho Federal Systems  
Chapel Hill, NC*



**FINAL NIH STATEMENT ON SHARING RESEARCH DATA**

**RELEASE DATE: February 26, 2003**

**NOTICE: NOT-OD-03-032.**

**National Institutes of Health**

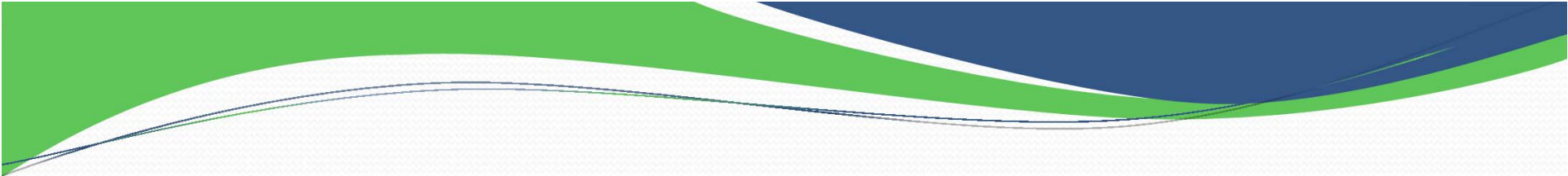
**As part of NIH's long-standing policy to share and make available to the public the results and accomplishments...**

**[http://grants.nih.gov/grants/policy/data\\_sharing/](http://grants.nih.gov/grants/policy/data_sharing/)**

## Background and Rationale:

- **“... long-standing policy to share ... the results of the activities that it funds.”**
- **“We believe that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health.”**
- **“Sharing data reinforces open scientific inquiry, encourages diversity of analysis and ... makes possible the testing of new or alternative hypotheses and methods of analysis ...”**



- 
- **“NIH recognizes that the investigators who collect the data have a legitimate interest in benefiting from their investment of time and effort.”**
  - **“NIH continues to expect that the initial investigators may benefit from first and continuing use but not from prolonged exclusive use.”**

## Policy:

- **“The NIH expects and supports the timely release and sharing of final research data.”**
- **“... to be no later than the acceptance for publication of the main findings from the final data set.”**
- **“... investigators submitting an NIH application seeking \$500,000 or more in direct costs in any single year are expected to include a plan for data sharing ...”**





## Implementation:

- **How do you create public-use datasets?**
- **How do you release them to the public?**
- **How do you de-identify participant data and safeguard confidential health information?**
- **How do you create user-friendly data files and assemble the corresponding documentation?**
- **Where do you post them?**



# **NIH Data Sharing Policy & Example of Its Implementation**

**Paul Wakim**  
**National Institute on Drug Abuse**



# **NIDA CTN Data Share: a Case Study**

**Paul VanVeldhuisen  
The Emmes Corporation**





# **In the Trenches – De-identifying the CALERIE Database from a Programming Perspective**

**Annette McDaniel**  
**Duke Clinical Research Institute**



# **The NIDDK Repositories – Adding Value to Shared Resources**

**Rebekah Rasooly  
National Institute of Diabetes  
and Digestive and Kidney Diseases**