

The Role PI & Institutional Characteristics Play in Shaping Data Sharing Behavior

Amy Pienta, ICPSR
University of Michigan



Acknowledgements

Database Development

Library of Congress

NIH/NLM

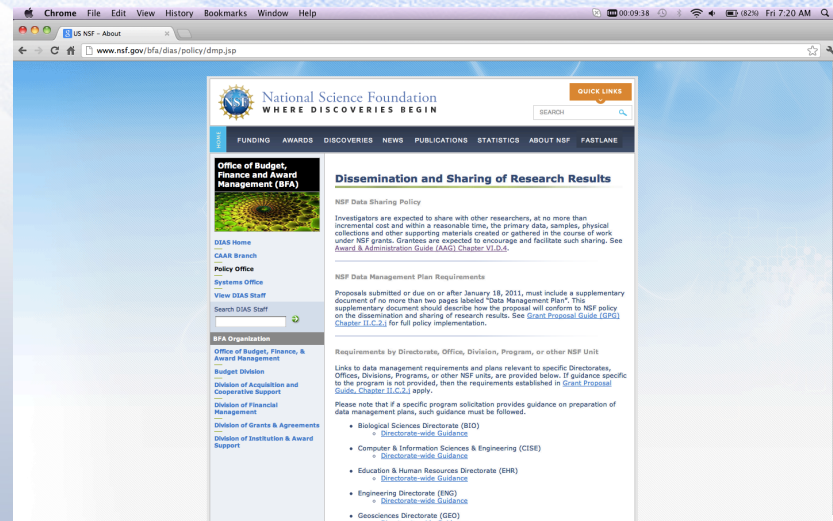
NSF/SciSIP

Collaborators

- Jared Lyle
- Myron Gutmann
- Darrell Donakowski
- JoAnne O'Rourke
- Felicia LeClere
- George Alter

Scope of Our Work

- Open Data Policies
- Increase transparency
- Data viewed as a public good
- Maximize the science investment



ICPSR

Inter-university Consortium for Political and Social Research

GSS General Social Survey

PSID

A national study of socioeconomics and health over lifetimes and across generations

monitoring the future



a continuing study of American youth

 **ANES**
American National Election Studies

surveys of consumers
THOMSON REUTERS UNIVERSITY OF MICHIGAN

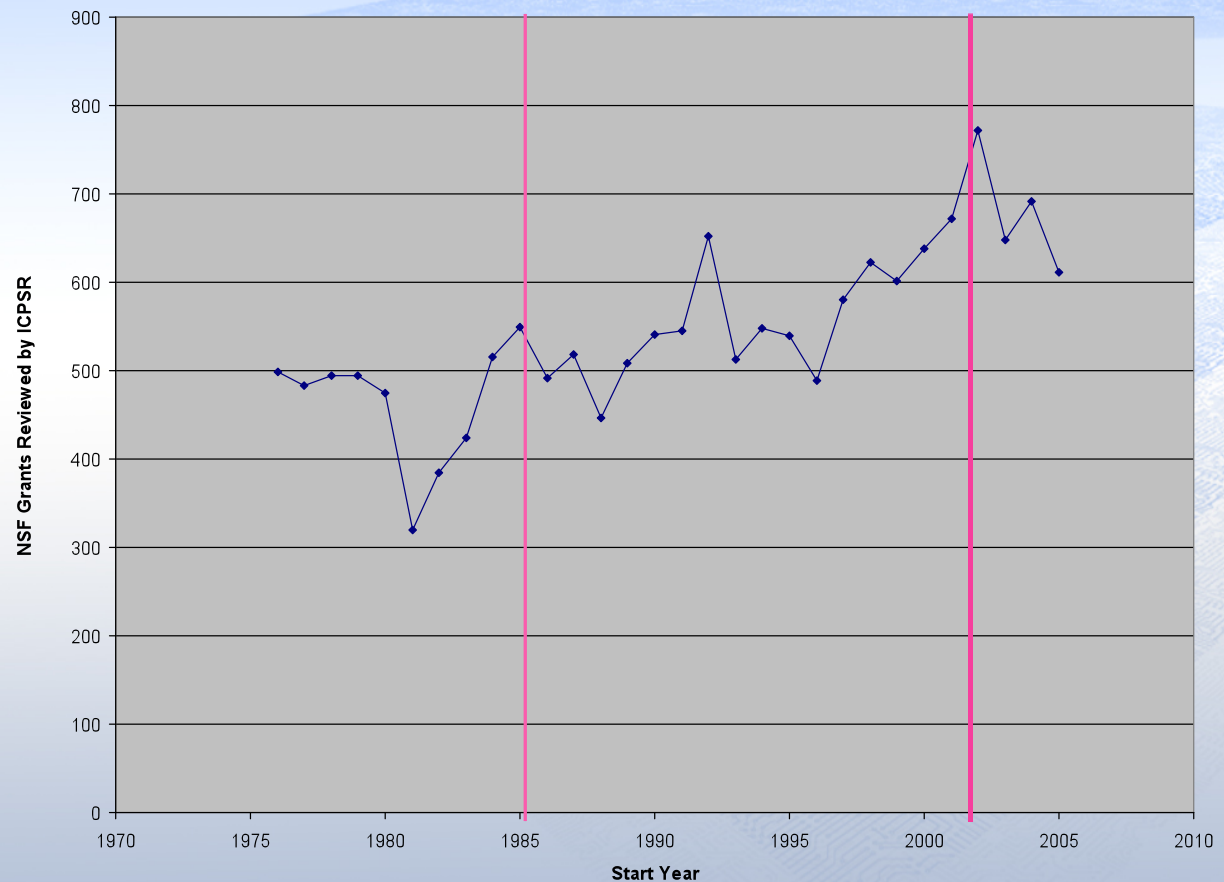
Awards in a database held at ICPSR

- Social science and/or behavioral science awards made by NSF and NIH
- Original or primary data collection proposed, including assembling a database from existing (archival) sources

NSF Grant Awards in LEADS Database

LEADS originally
contained
17,194 awards
made by NSF

LEADS spans 35
years of awards -
1976 to 2005



LEADS Database at ICPSR

	# Records Reviewed	# Social Science Data
Recent NSF (1976+)	17,194	2,537
Historic NSF (Pre-1976)	96,403	4,019

National Survey of Social Science PIs

- 1,499 NSF PIs emailed a survey
 - living, email address located
- 411 responded (27.4% response rate)
- 316 collected data (77% screened in correctly)
- 283 after dissertation awards dropped

Results: Award Characteristics

- 88.8% funded by SBE (CSE=15, OPP=16)
- Awards made between 1985-2001
- Awards were 1-8 years long (more than ½ were 2-3 years long)
- Mean award amount 312K
 - Archived - 591K
 - Informal - 231K
 - Not Shared - 225K

Results: Prevalence of Data Sharing among Social/Behavioral PIs

	Archived Data	Shared Formally	Not Shared
NSF (N=277)	23.1%	43.3%	33.6%

Results: Prevalence of Data Sharing among Social/Behavioral PIs

	Archived Data	Shared Formally	Not Shared
NSF N=277	23.1%	43.3%	33.6%
NIH N=732	7.7%	46.2%	46.2%

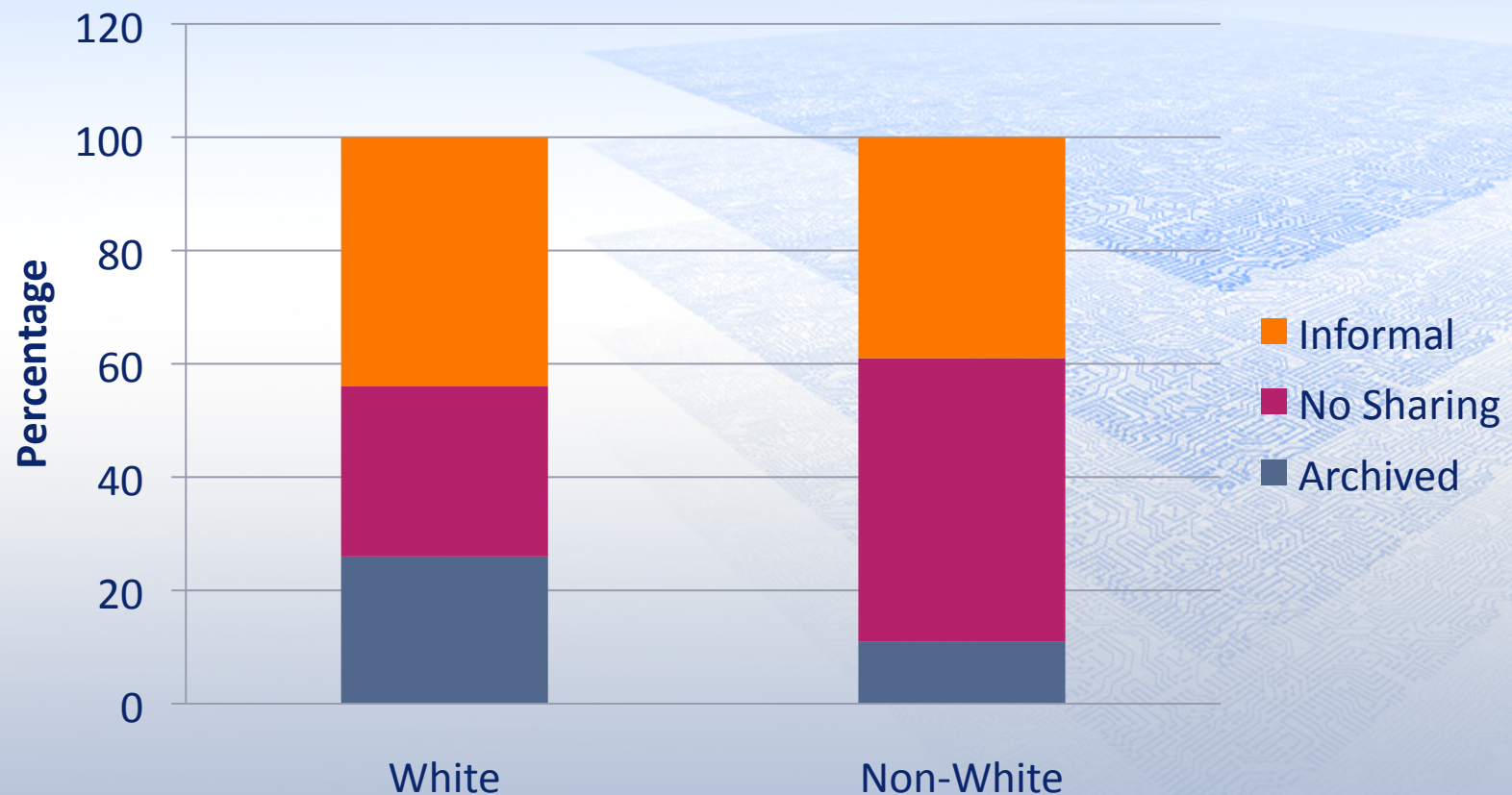
Data sharing rates in greater depth...

- Archived includes:
 - self-reported archived (verified most)
 - placed in an Institutional Repository
- Informal includes:
 - Department and personal website
 - Data upon request (95%)
 - Other

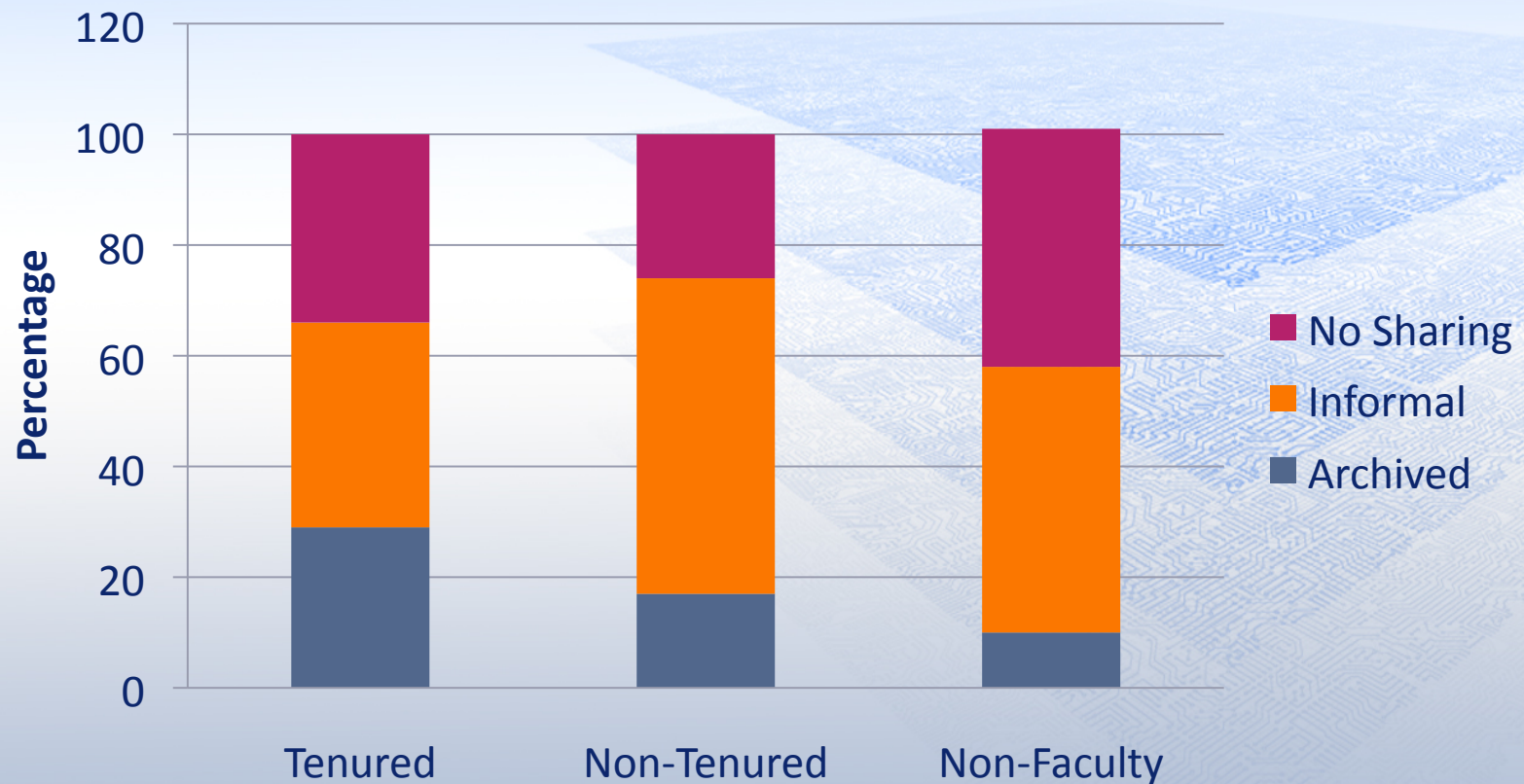
Results: Characteristics of NSF PIs

- 35% Female
- 83.4% White
- 60.7% Tenured (v. non-tenured & non-faculty)
- Mean Age = 43 (range, 27-75)
- Mean # Lifetime Federal Awards 6.6 (range, 1-100)

Results: Race and Data Sharing Behavior among NSF PIs (n=277)



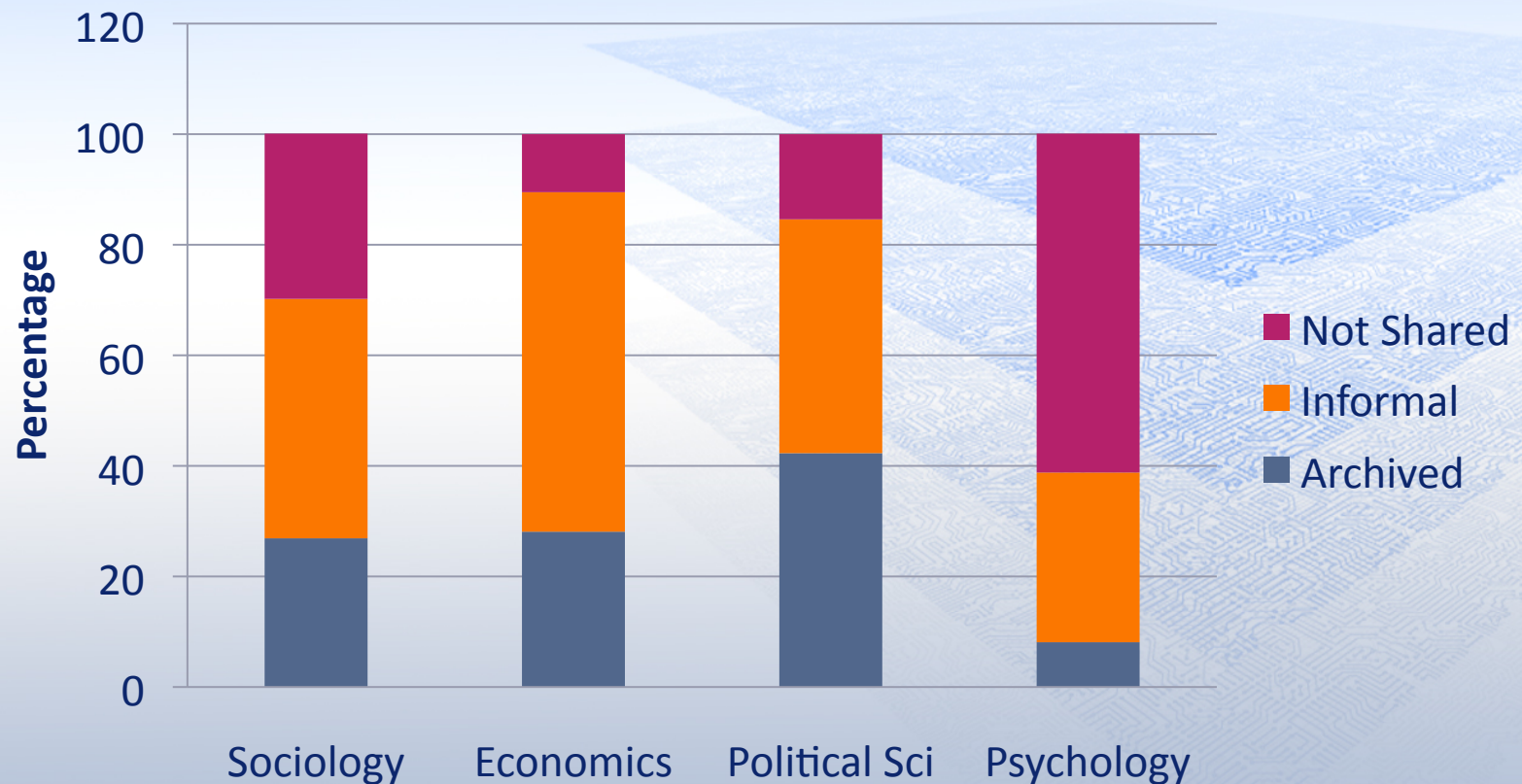
Results: Rank and Data Sharing Behavior among NSF PIs (n=277)



Results: Discipline of NSF PIs

	Percentage
Sociology	24.2
Economics	20.6
Political Science	9.4
Psychology	22.4
Other	23.5

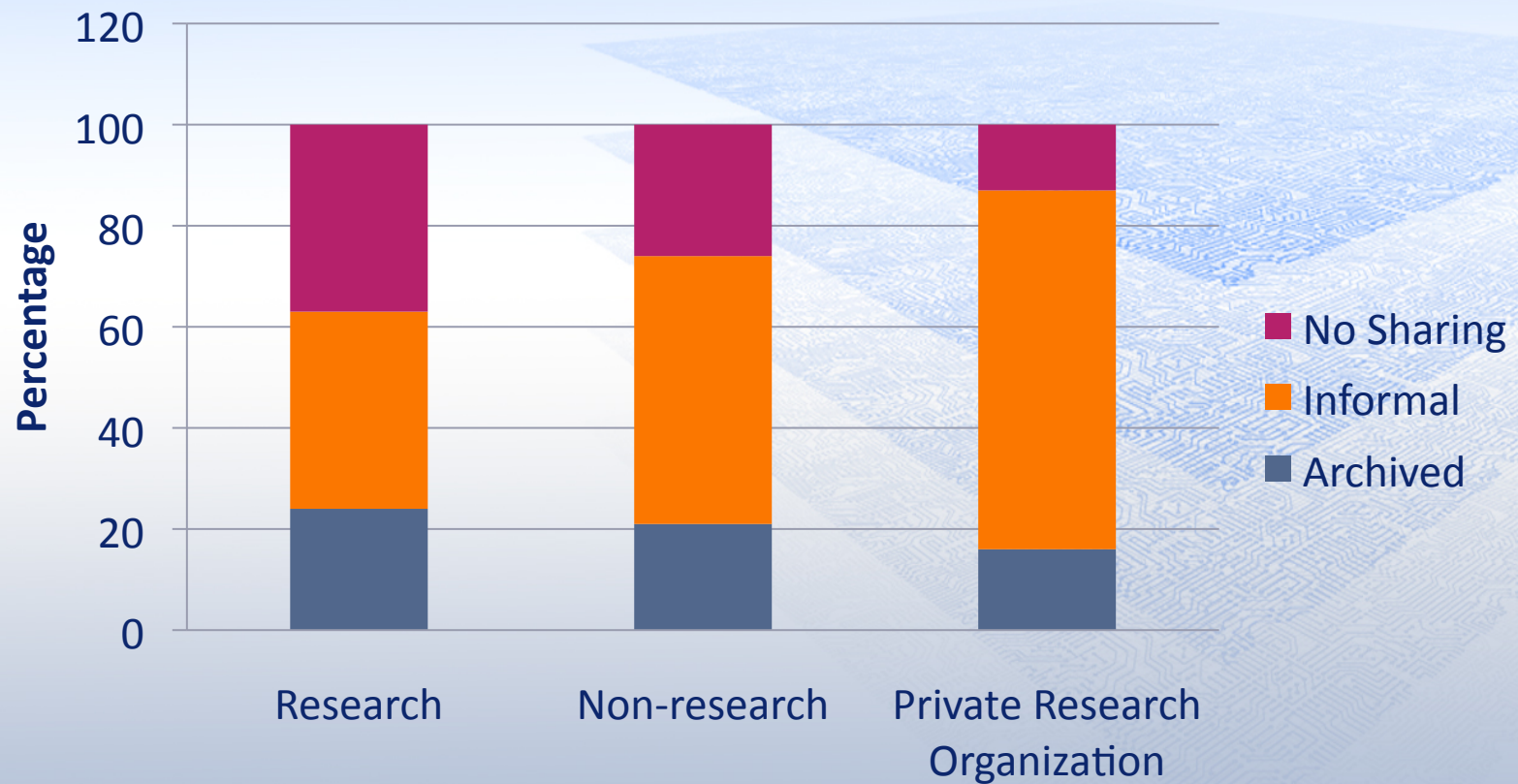
Results: Disciplinary Differences in Data Sharing Behavior



Contextual Factors

- Geographic region of the institution
 - Northeast, Mid-West, Southern, West
 - Similar data sharing practices across
- Carnegie classification of the institution
 - Private Research Organizations
 - Research Institutions
 - Non-research Institutions

Type of Institution



Results: Perceived Barriers to Sharing Data from NSF Award

Do not have adequate time to prepare data	49.4
Difficult to prepare documentation	45.3
Concerned about protecting confidentiality	31.4
Concerned about potential for others misinterpreting data	24.0
Concerned others will publish before me	21.6
Others not interested in topic	20.9
Informed consent language prevents it	14.9
IRB prevents me from sharing	10.2

Median Publications by Data Sharing

	Data Archived (n=111)	Data Shared Informal (n=415)	Data Not Shared (n=409)	Total (n=935)
Median # Primary PI Pubs	6	6	3	4
Median # Secondary Pubs, No PI	8	6	3	5
Median # Pubs with Students	4	3	1	2

Source: Pienta, Amy M., George Alter, and Jared Lyle. 2010. "The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data." Presented at the BRICK, DIME, STRIKE Workshop, The Organisation, Economics, and Policy of Scientific Research, Turin, Italy, April 23-24, 2010 (<http://hdl.handle.net/2027.42/78307>)

Recapping the Highlights

- Disciplinary differences result in the greatest differences in data sharing behaviors.
 - Economists and Political Scientists share the most. Informal sharing is greater among the economists.
 - Psychologists share data less often than all others social/behavioral scientists awarded projects by NSF.

Recapping the Highlights

- White PIs are also more likely to share.
- Tenured faculty archive data more often than non-tenured faculty.
 - But, non-tenured faculty share more informally.
 - Other individual characteristics matter less.

Recapping the Highlights

- Institutional differences exist as well.
 - PIs located at private research organizations are most successful at sharing data – especially informally.
- The most important barriers to data sharing among NSF PIs is time to prepare the data and documentation.

Pienta, Amy M. and Jared Lyle. National Study of Social Science Principal Investigators. ICPSR29941-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], forthcoming. doi: 10.3886/ICPSR29941.v1