

# MANAGING RESEARCH DATA FOR TRANSPARENCY AND REUSABILITY

Scout Calvert, PhD  
Michigan State University  
[calvert4@msu.edu](mailto:calvert4@msu.edu)

## WHY SHARE DATA?

- It's easy and cheap (maybe?) because of development of computers, networks
- Citation boost: papers that have available data are cited more
- Public access to research: Expectation in Europe and North America, globally
- Potential for data reuse for more and better knowledge claims
- Prevent and detect fraud
- Prevent errors and retraction due to error
- Transparency is central to replication and reproducibility
- Tools and standards for sharing keep getting better

## POLITICAL SCIENCE

# When contact changes minds: An experiment on transmission of support for gay equality

Michael J. LaCour<sup>1</sup> and Donald P. Green<sup>2</sup>

Can a single conversation change minds on divisive social issues, such as same-sex marriage? A randomized placebo-controlled trial assessed whether gay ( $n = 22$ ) or straight ( $n = 19$ ) messengers were effective at encouraging voters ( $n = 972$ ) to support same-sex marriage and whether attitude change persisted and spread to others in voters' social networks. The results, measured by an unrelated panel survey, show that both gay and straight canvassers produced large effects initially, but only gay canvassers' effects persisted in 3-week, 6-week, and 9-month follow-ups. We also find strong evidence of within-household transmission of opinion change, but only in the wake of conversations with gay canvassers. Contact with gay canvassers further caused substantial change in the ratings of gay men and lesbians more generally. These large, persistent, and contagious effects were confirmed by a follow-up experiment. Contact with minorities coupled with discussion of issues pertinent to them is capable of producing a cascade of opinion change.

## POLITICAL SCIENCE


# Durably reducing transphobia: A field experiment on door-to-door canvassing

David Broockman<sup>1\*</sup> and Joshua Kalla<sup>2</sup>

Existing research depicts intergroup prejudices as deeply ingrained, requiring intense intervention to lastingly reduce. Here, we show that a single approximately 10-minute conversation encouraging actively taking the perspective of others can markedly reduce prejudice for at least 3 months. We illustrate this potential with a door-to-door canvassing intervention in South Florida targeting antitransgender prejudice. Despite declines in homophobia, transphobia remains pervasive. For the intervention, 56 canvassers went door to door encouraging active perspective-taking with 501 voters at voters' doorsteps. A randomized trial found that these conversations substantially reduced transphobia, with decreases greater than Americans' average decrease in homophobia from 1998 to 2012. These effects persisted for 3 months, and both transgender and nontransgender canvassers were effective. The intervention also increased support for a nondiscrimination law, even after exposing voters to counterarguments.

# DETECTING FRAUD

Holsteins Favor Heifers, Not Bulls: Biased Milk Production Programmed during Pregnancy as a Function of Fetal Sex

Katie Hinde , Abigail J. Carpenter, John S. Clay, Barry J. Bradford  
Published: February 3, 2014 • <http://dx.doi.org/10.1371/journal.pone.0086169>

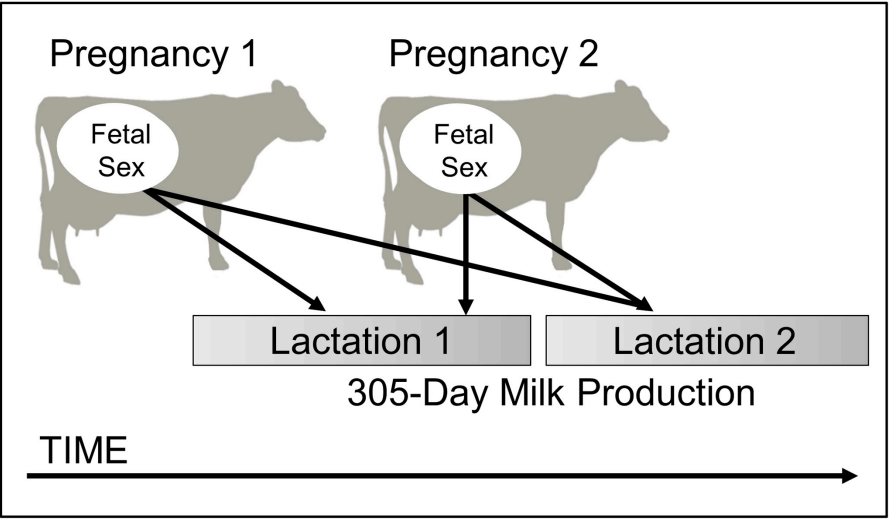
Article	Authors	Metrics	Comments	Related Content
⌵				

Abstract

- Introduction
- Methods
- Results
- Discussion
- Acknowledgments
- Author Contributions
- References

Abstract

Mammalian females pay high energetic costs for reproduction, the greatest of which is imposed by lactation. The synthesis of milk requires, in part, the mobilization of bodily reserves to nourish developing young. Numerous hypotheses have been advanced to predict how mothers will differentially invest in sons and daughters, however few studies have addressed sex-biased milk synthesis. Here we leverage the dairy cow model to investigate such phenomena. Using 2.39 million lactation records from 1.49 million dairy cows, we demonstrate that the sex of the fetus influences the capacity of the mammary gland to synthesize milk during lactation. Cows favor daughters, producing significantly more milk for daughters than for sons across lactation.



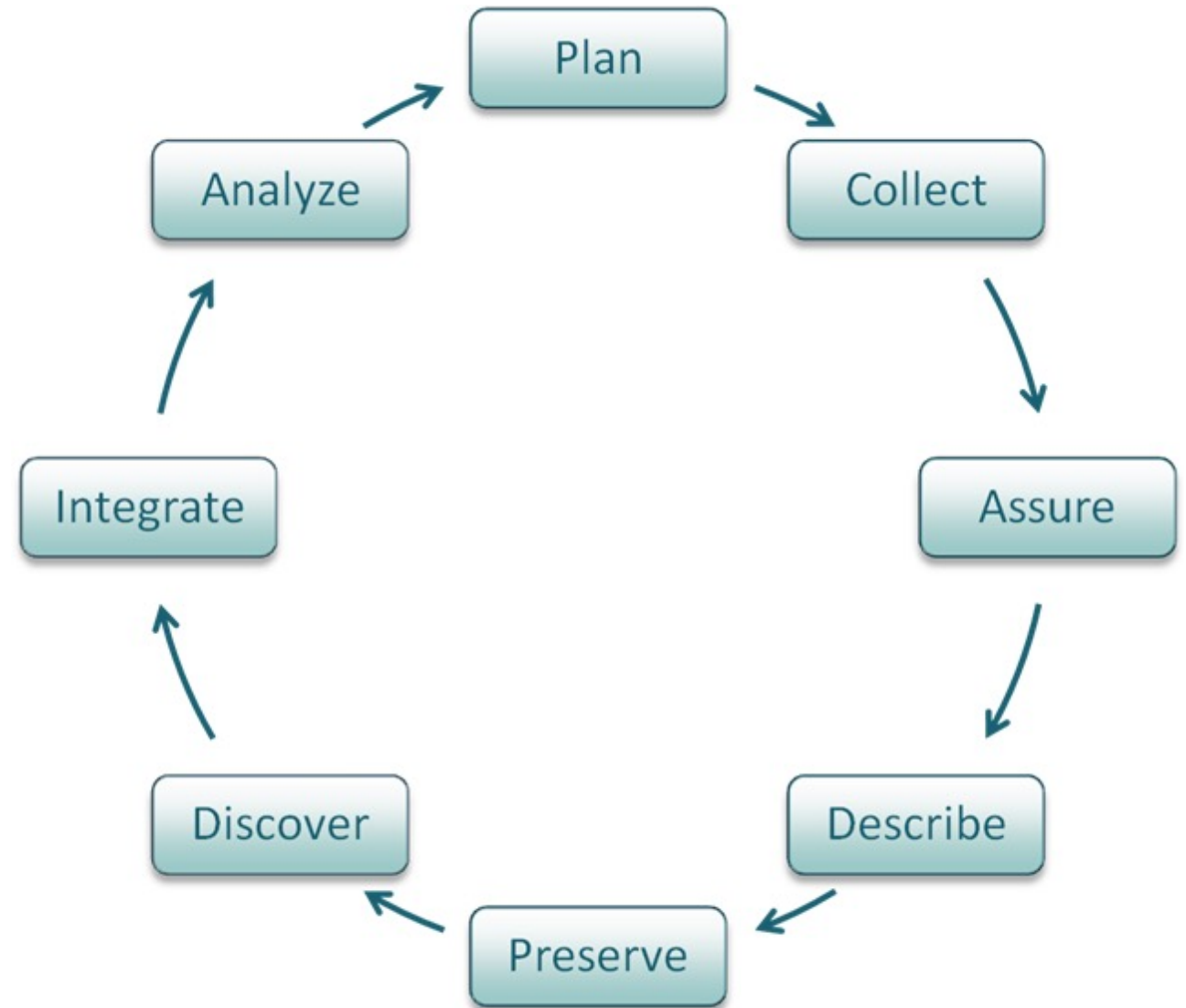
DATA REUSE

# PREVENT RETRACTION

- Most retractions are due to simple errors
- Good data management prevents accidental errors
- Data package for peer review requires documentation of entire analytical context

*Goal: other researchers can open and reuse your data without additional communication!*

MANAGING  
RESEARCH  
DATA FROM  
THE OUTSET  
OF THE  
PROJECT



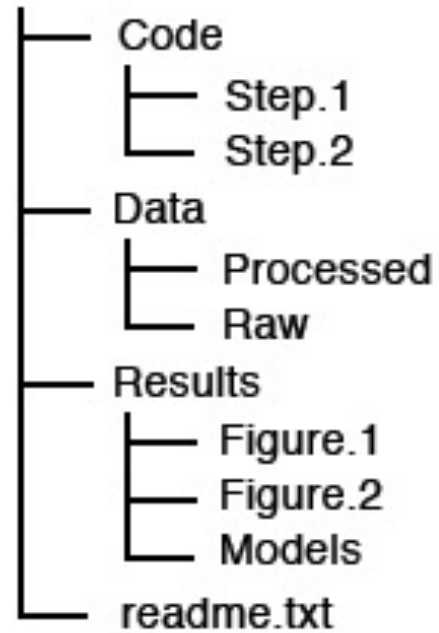
# THE DATA PACKAGE

- Goal: other researchers can open and reuse your data without additional communication!
- Data AND Documentation: [Metadata](#)
- Data – Raw? Cleaned? Subsetted?
- Documentation of data collection, variables, measurements, parameters, missing variables, codebooks.
  - DDI (Data Documentation Initiative) is one standard – embedded in [Colectica](#) for Excel
- Documentation of all cleaning / subsetting steps
- Analysis steps, do files, scripts, code used in analysis
  - `##` Comment in the code or use built in tools
- Documentation of software, version, packages, etc.

# DATASET ORGANIZATION

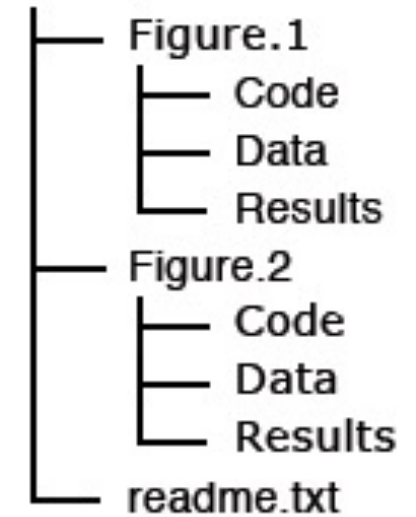
## A) Organized by File type

### Example.A



## B) Organized by Analysis

### Example.B





# FILE ORGANIZATION

- Include a README file to include information about the project
- Use human readable file names that distinguish types and versions of files
- Use consistent folder and file naming conventions
- ISO 8601 dates help with organization and version control YYYY-MM-DD
- Examples:
  - ProjectCode\_analysis001\_2021-06-21.txt
  - ProjectCode\_data001\_2021-05-14.csv
  - ProjectCode\_manuscript\_2021-09-23.doc

# FILE TYPES

- Ensure your data can be seen and reused long into the future:
  - Use widely available and open file types whenever possible
  - Maintain a copy in the original software (SPSS, Stata, etc.)
  - But export to Excel or CSV for sharing, archiving, and depositing
  - UK Data Service has a [list of recommended formats](#).

# PROTECT YOUR DATA

- Get organized starting now, and stay organized.
- Back up your data automatically or on a schedule.
- Keep 3 copies: Two local, one offsite (or in the cloud).
- Practice online safety. Learn about phishing and other exploits.

# DEPOSIT YOUR DATA

- Depositing your data promotes your research, citation boost for papers where data is shared
- Reliable, independent infrastructure
- Good options for repositories that promote FAIR data
- Findable, Accessible, Interoperable, Reusable
- Get a DOI to make sharing and citation easier

## What is FAIR DATA?



Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

**FINDABLE**



Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.

**ACCESSIBLE**



Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.

**INTEROPERABLE**



Data and collections have a clear usage licenses and provide accurate information on provenance.

**REUSABLE**

# DATA REPOSITORIES

- [Registry of Research Data Repositories](#)
- [FAIRsharing.org list of databases and repositories](#)
- [Open Science Framework](#) also provides badges for good practice
- [Zenodo](#) is free and open, provided by CERN

## WRAPPING UP

- Get an [ORCID](#) researcher ID to disambiguate your work from others and promote research transparency
- Resources to help you get started:
  - [Data One >> Learning](#)
  - [UK Data Service](#)
  - [University of California Curation Center \(UC3\)](#)
  - [Digital Curation Centre](#)
  - [Fairsharing.org](#)

# THANK YOU

Scout Calvert, PhD  
[calvert4@msu.edu](mailto:calvert4@msu.edu)

