

Every Dataset has a Story

Sarah Rush & John Marcotte
ICPSR, University of Michigan
IASSIST 2018



Every Dataset has a Story

- Studies that generate research data often have fascinating histories
- Motivations behind a data collection can enhance understanding
- Stories *about* data can encourage analysis
- Stories can engage the research community



Every Dataset has a Story

1. Background
2. Contributors
3. Engagement
4. Stories
5. Next Steps
6. Questions



1. Background: Past

- Downloading has historically been the primary communication between researchers and data archives
- Bandwidth limitations have been an impediment



1. Background: Present

- Bandwidth limitations are no longer a concern
- People have Internet access in their pockets
- Options for interactive communication have increased (e.g. forums, social media)
- Opportunity for increased knowledge sharing and transparency



1. Background: Archives

Changing role of Archives

- Traditional: Providers of data and codebooks
- Emerging: Active facilitators of research
 - Data discovery tools (e.g., variable search)
 - Online Analysis
 - Bibliography of Data-related Literature
 - Data usage aids (e.g., Data Guides, Data Driven Learning Guides)
 - Virtual Data Enclave for restricted data



1. Background: Archives

Evolving relationships with data providers and data users

- Interactive
- Multiway communication and contributions

Data Providers ↔ Archives ↔ Data Users



2. Contributors

How can we use these evolving relationships to engage data users and enhance secondary data analysis?



2. Contributors: Data Providers

Partner with Principal Investigators to provide:

- Background beyond standard metadata
- Tips on approaching data
- Reasons data were collected



2. Contributors: Data Users

- Encourage sharing of study-related materials that can enhance documentation and metadata:
 - Narratives on how data were selected and analyzed
 - Syntax
 - Scales
 - Notes on working with specific datasets (e.g., overcoming obstacles)



2. Contributors: Benefits

- Builds a support community for data
- Enriches scientific enterprise as well as specific datasets
- Provides a way for data users to give back



2. Contributors: Benefits

- Promotes data analysis and the importance of data
- Brings data “to life,” particularly for novice researchers; helps them see possibilities
- Research builds on research
- Inspire new research



3. Engagement

- Data Sharing for Demographic Research (DSDR) at ICPSR is collecting:
 - Mini interviews with study PIs about their research and why data were collected
 - Stories from data users about their experience analyzing data
- Interviews and stories will be incorporated into the website and promoted on social media
- Stories are just the start!



3. Engagement: Goal

To encourage and enable end-user contributions, thereby *facilitating* and *enhancing* secondary data analysis



4. Stories

Initiatives

- Genomic Data and Biomarkers
- Deductive Disclosure Risk
- Data Harmonization
- Big Data

Discover


Our Mission

Increase the impact of NICHD-funded research within the scientific mission of the NICHD-Population Dynamics Branch by providing research infrastructure to:

- Promote data sharing
- Support the development of procedures and technologies for data sharing
- Disseminate best practices in data sharing
- Provide a resource that catalogs NICHD-funded data available for secondary analysis
- Promote the secondary analysis of data collected through NICHD grants to research teams outside the original grantees

DSDR Data Stories

Featured "Five Questions with a PI" Profile

 John Marcotte, DSDR Director
[more PI Profiles](#)

Data User Stories

[Learn how DSDR data are being used](#)

News and Events

- New Data Available: Continuity and Change in Contraceptive Use, United States, 2012-2014
- Visit DSDR during the 2018 Population Association of America

[More News >>](#)

[Subscribe to DSDR News](#)


Sarah Reah ([logout](#)) [Privacy Policy](#) [Contact Us](#) [Questions or Comments?](#) [f](#) [t](#)

The Data Sharing for Demographic Research (DSDR) project is housed within Interdisciplinary Connections for Political and Social Research (ICPSR) and is fully funded through the Population Dynamics Branch (PDB) of NICHD.
© The Regents of the University of Michigan

- Collected stories will be featured on the DSDR home page as way to engage data users
- Potential to entice website visitors to learn more about studies and data usage

DSDR Data Stories

Featured "Five Questions with a PI" Profile

 John Marcotte, DSDR Director
[more PI Profiles](#)

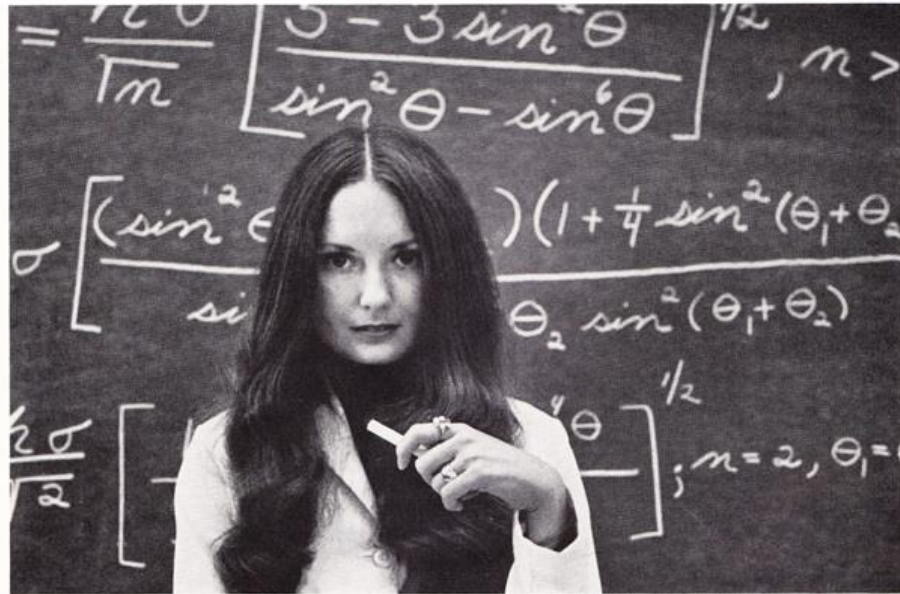
Data User Stories

[Learn how DSDR data are being used](#)

Dewar's Example:

DEWAR'S PROFILES

(Pronounced Do-ers "White Label")



BLENDED SCOTCH WHISKY • 86.8 PROOF • © SCHENLEY IMPORTS CO., N.Y., N.Y.

SHEILA ANN T. LONG

HOME: Hampton, Virginia

AGE: 28

PROFESSION: Physicist

HOBBIES: Ballet, Sailing, Car Racing, Chess

LAST BOOK READ: "Beyond Freedom and Dignity"

LAST ACCOMPLISHMENT: Member of the team of international scientists who are mapping earth's electromagnetic field for the first time

QUOTE: "Scientific research in all fields has been a prime contributor to America's greatness. Let us not forget this in our concern for the dying environment, for Technology holds the very means to save it."

PROFILE: Brilliant, beautiful, in love with life. Involved, and unintimidated by difficult challenge. Saluted by *New Woman* magazine as one of the 26 women who "made it big in their twenties."

SCOTCH: Dewar's "White Label"



Authentic. There are more than a thousand ways to blend whiskies in Scotland, but few are authentic enough for Dewar's "White Label." The quality standards we set down in 1846 have never varied. Into each drop goes only the finest whiskies from the Highlands, the Lowlands, the Hebrides. **Dewar's never varies.**

Similar Profile to Dewar's:



Profiles

John E. Marcotte, Ph.D.

Home State: Michigan

Position(s): Archivist & Clinical Adjunct

Research Focus: Data Security, Disclosure, Nursing Staffing

Profile: Director of [Data Sharing for Demographic Research \(DSDR\)](#) at ICPSR

Currently working on: Disclosure Risk with Administrative Data

Last Book Read: Big Data: A Revolution That Will Transform How We Live, Work, and Think

Quote: "Show me the data!"

How did you become involved with the study you deposited with DSDR?

My project has had a series of principal investigators. When the project needed a new P.I., my experience and expertise were a good fit.

What are your major findings?

Disclosure protections for organizations is a growing concern.

Now that these data are available to other researchers, what advice would you give for those analyzing the contents of the study?

The best research builds on previous work. Team science is better science. Reach out to collaborators in different disciplines.

What excites you about being a researcher?

Being able to provide data that informs health outcomes and public policy

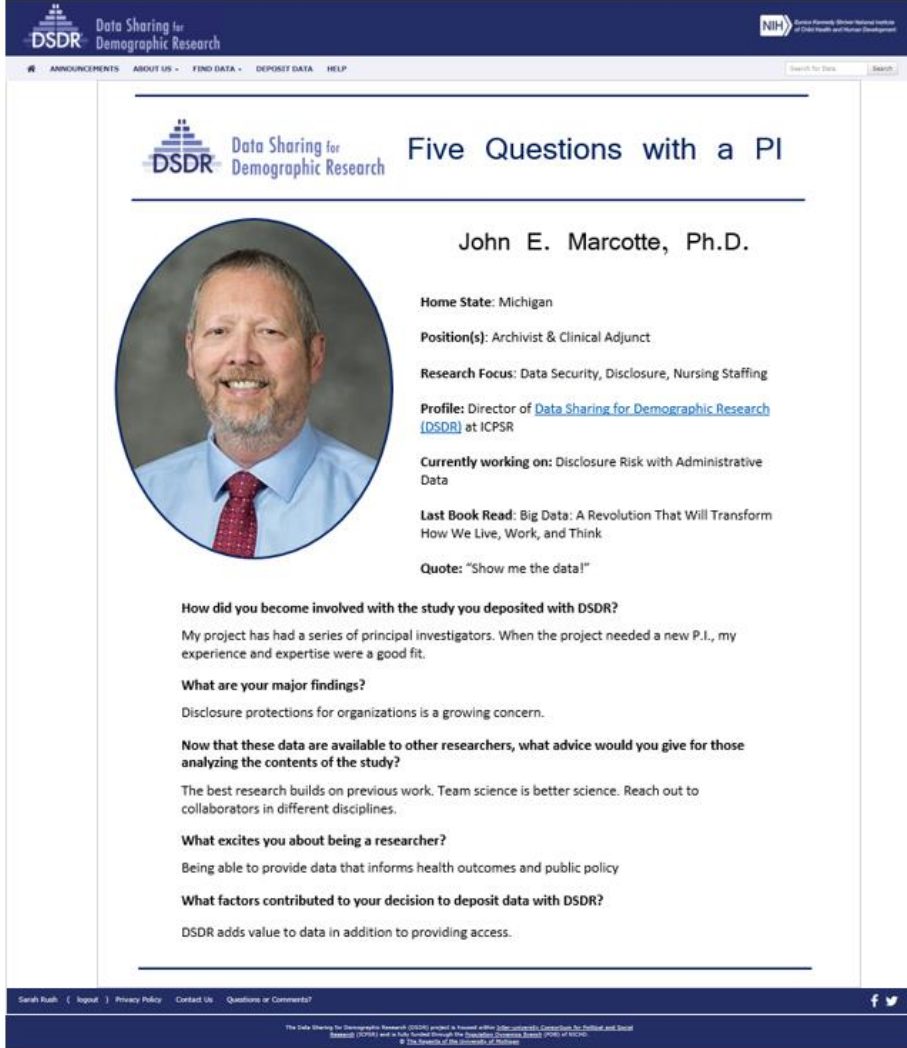
What factors contributed to your decision to deposit data with DSDR?

DSDR adds value to data in addition to providing access.

Data Archive: *DSDR!*

4. Stories: Data Providers

- Features work of Data Provider in interview format
- Opportunity for Data Provider to promote study to new and repeat users of data
- To be incorporated into DSDR website in multiple areas: home page, study page, interviews page



The screenshot shows the DSDR website interface. At the top, there is a navigation bar with the DSDR logo and the text "Data Sharing for Demographic Research". Below this, there are links for "ANNOUNCEMENTS", "ABOUT US", "FIND DATA", "DEPOSIT DATA", and "HELP". The main content area features a circular profile picture of John E. Marcotte, Ph.D., a man with a beard wearing a light blue shirt and a red tie. To the right of the photo, his name and title are listed: "John E. Marcotte, Ph.D.". Below the photo, several key details are provided: "Home State: Michigan", "Position(s): Archivist & Clinical Adjunct", "Research Focus: Data Security, Disclosure, Nursing Staffing", "Profile: Director of Data Sharing for Demographic Research (DSDR) at ICPSR", "Currently working on: Disclosure Risk with Administrative Data", "Last Book Read: Big Data: A Revolution That Will Transform How We Live, Work, and Think", and "Quote: 'Show me the data!'". The interview questions and answers are listed below: "How did you become involved with the study you deposited with DSDR?" (My project has had a series of principal investigators. When the project needed a new P.I., my experience and expertise were a good fit.), "What are your major findings?" (Disclosure protections for organizations is a growing concern.), "Now that these data are available to other researchers, what advice would you give for those analyzing the contents of the study?" (The best research builds on previous work. Team science is better science. Reach out to collaborators in different disciplines.), "What excites you about being a researcher?" (Being able to provide data that informs health outcomes and public policy), and "What factors contributed to your decision to deposit data with DSDR?" (DSDR adds value to data in addition to providing access.). At the bottom of the page, there is a footer with social media icons for Facebook and Twitter, and a small copyright notice.

4. Stories: Data Providers

Home State: Michigan

Position(s): Archivist & Clinical Adjunct

Research Focus: Data Security, Disclosure, Nursing Staffing

Profile: Director of [Data Sharing for Demographic Research \(DSDR\)](#) at ICPSR

Currently working on: Disclosure Risk with Administrative Data

Last Book Read: Big Data: A Revolution That Will Transform How We Live, Work, and Think

Quote: "Show me the data!"

How did you become involved with the study you deposited with DSDR?

My project has had a series of principal investigators. When the project needed a new P.I., my experience and expertise were a good fit.

What are your major findings?

Disclosure protections for organizations is a growing concern.

Now that these data are available to other researchers, what advice would you give for those analyzing the contents of the study?

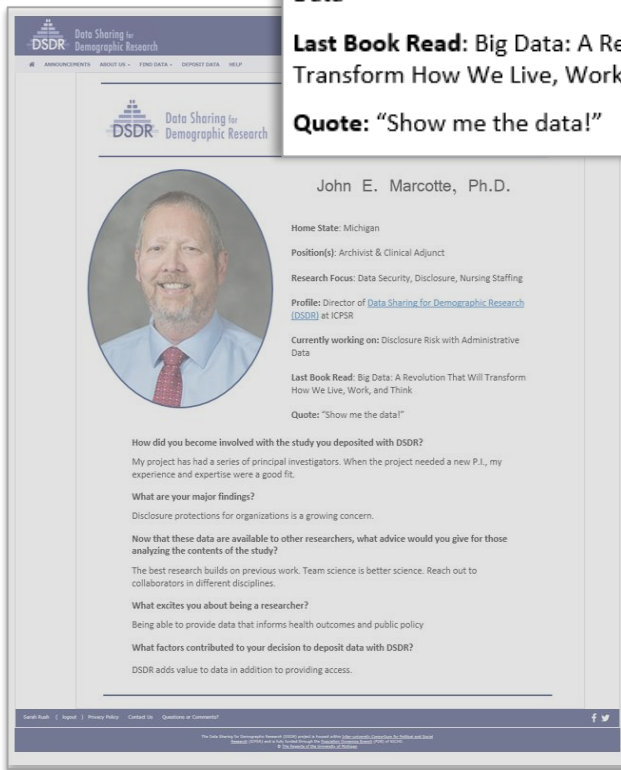
The best research builds on previous work. Team science is better science. Reach out to collaborators in different disciplines.

What excites you about being a researcher?

Being able to provide data that informs health outcomes and public policy

What factors contributed to your decision to deposit data with DSDR?

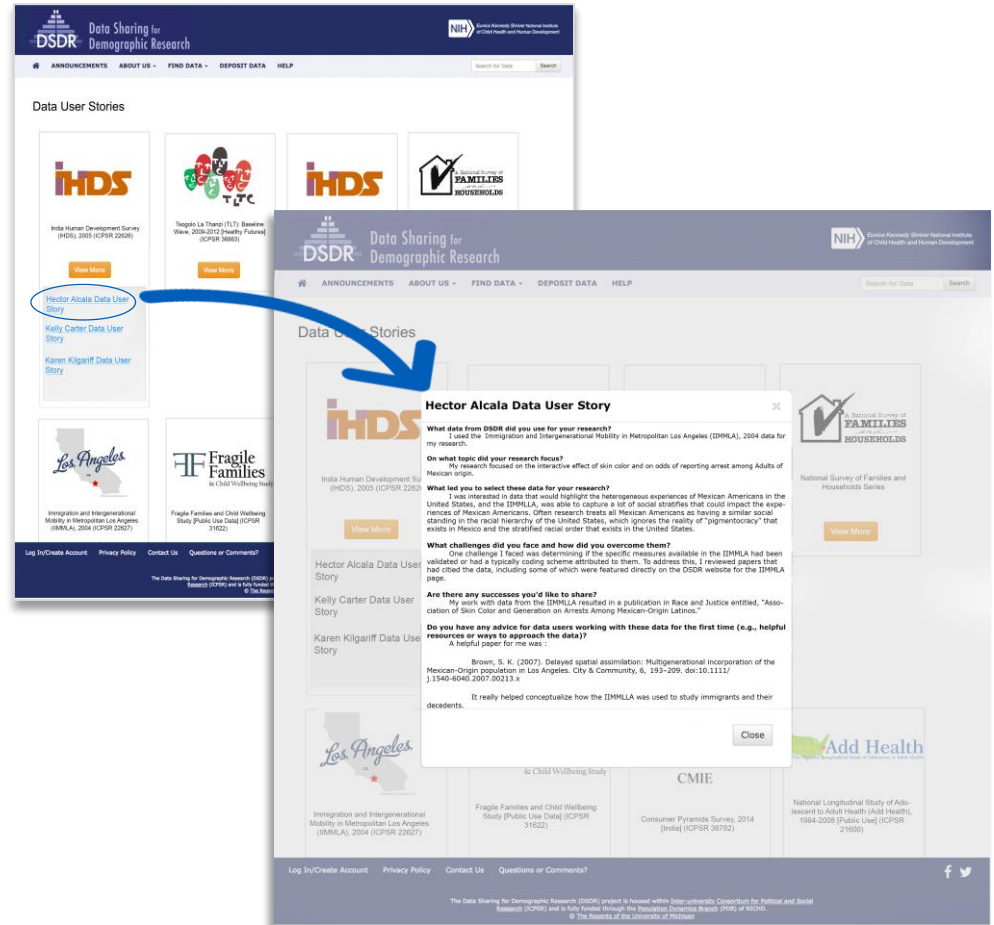
DSDR adds value to data in addition to providing access.



The screenshot shows the DSDR profile page for John E. Marcotte, Ph.D. The page includes a circular profile picture of a man with a beard and glasses, wearing a blue shirt and a red tie. The text on the page matches the information provided in the callouts above. At the bottom of the page, there is a footer with navigation links: Sarah Rusk, Home, Privacy Policy, Contact Us, and Questions or Comments? There are also social media icons for Facebook and Twitter.

4. Stories: Data Users

- Highlights secondary analysis of data in DSDR
- Opportunity for Data Users to promote research and talk about experience
- To be incorporated into DSDR website in multiple areas: home page, study page, data stories page



4. Stories: Data Users

Hector Alcalá Data User Story

What data from DSDR did you use for your research?

I used the Immigration and Intergenerational Mobility in Metropolitan Los Angeles (IIMMLA), 2004 data for my research.

On what topic did your research focus?

My research focused on the interactive effect of skin color and on odds of reporting arrest among Adults of Mexican origin.

What led you to select these data for your research?

I was interested in data that would highlight the heterogeneous experiences of Mexican Americans in the United States, and the IIMMLA, was able to capture a lot of social stratifies that could impact the experiences of Mexican Americans. Often research treats all Mexican Americans as having a similar social standing in the racial hierarchy of the United States, which ignores the reality of "pigmentocracy" that exists in Mexico and the stratified racial order that exists in the United States.

What challenges did you face and how did you overcome them?

One challenge I faced was determining if the specific measures available in the IIMMLA had been validated or had a typically coding scheme attributed to them. To address this, I reviewed papers that had cited the data, including some of which were featured directly on the DSDR website for the IIMMLA page.

Are there any successes you'd like to share?

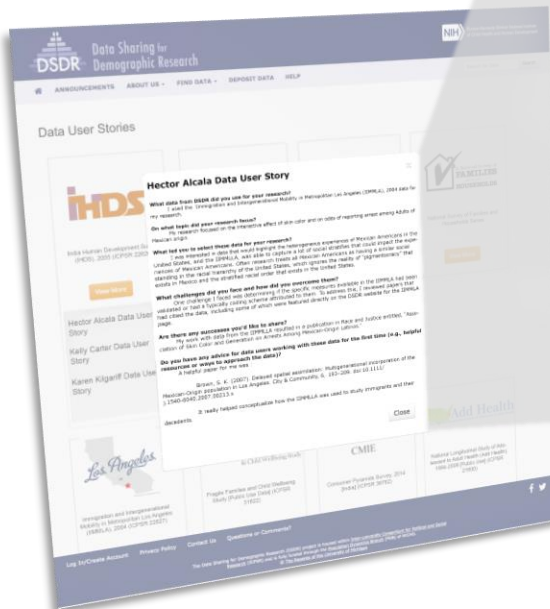
My work with data from the IIMMLA resulted in a publication in Race and Justice entitled, "Association of Skin Color and Generation on Arrests Among Mexican-Origin Latinos."

Do you have any advice for data users working with these data for the first time (e.g., helpful resources or ways to approach the data)?

A helpful paper for me was :

Brown, S. K. (2007). Delayed spatial assimilation: Multigenerational incorporation of the Mexican-Origin population in Los Angeles. *City & Community*, 6, 193–209. doi:10.1111/j.1540-6040.2007.00213.x

It really helped conceptualize how the IIMMLA was used to study immigrants and their decedents.



5. Next Steps

- Launch Data Stories feature on DSDR website DSDR.ICPSR.UMICH.EDU
- Promote feature and acquire additional PI interviews and data user stories
- Seek additional user contributions such as syntax, scales, and notes on working with specific datasets



6. Questions



Contact Information

Sarah Rush: shawtin@umich.edu

John Marcotte: jemarcot@umich.edu

DSDR: DSDR.ICPSR.UMICH.EDU



@DSDRdata



Contact Us

