



# Transparency, reproducibility, and replicability New initiatives at ICPSR

Margaret C. Levenstein

Director, Inter-university Consortium for Political and Social Research

# Promoting transparency and reproducibility

## ➤ What is ICPSR?

- Preserving and accessing shared data and data-related content
- Journal repositories and journal-related deposits

## ➤ Three new initiatives

- LinkageLibrary
- SOMAR
- Researcher passport

# ICPSR



- Founded in 1962 by 22 universities, now consortium of 800 institutions world-wide
- Focus on social and behavioral science data, broadly defined
- Current holdings
  - 10,000 studies, quarter million files
  - 1500 are *restricted studies*, almost always to protect confidentiality
  - Bibliography of Data-related Literature with 75,000 citations
- Approximately 60,000 active MyData (“shopping cart”) accounts
- Thematic collections of data about addiction and HIV, aging, arts and culture, child care and early education, criminal justice, demography, health and medical care, and minorities

# Preserving and accessing shared data and data-related content

- Make data sharing feasible
  - ICPSR's General Archive
    - Anyone can deposit
    - Curated and preserved
  - Guidance over data life cycle
    - Templates for consent, IRB, DMP consistent with transparent and reproducible access
- Incentivize data sharing
  - Standard citation
  - Bibliography
  - Usage statistics

# Data in the wild

- Organic or non-designed data creates new challenges for reproducibility and transparency
- Data changes in real time
  - Reproducibility requires snapshots, versioning
- No survey instrument or documentation of sampling to provide metadata to assist in re-use or discovery
  - Or even informed use of data the first time
  - Requires development of standards (extending DDI)
  - Citizen scientists necessary for transparency and reproducibility

# Data linkage challenges

- Linked data present challenges for both confidentiality and reproducibility
  - Linkage more accurate with more detailed information
    - Need standards for safe, ethical ways to enhance data with new linkages
  - Linked data easier to re-identify, even after removing unique identifiers
    - Need safe places to analyze linked data
  - Linkage strategies introduce differences in datasets that are often not well documented



# LINKAGE LIBRARY

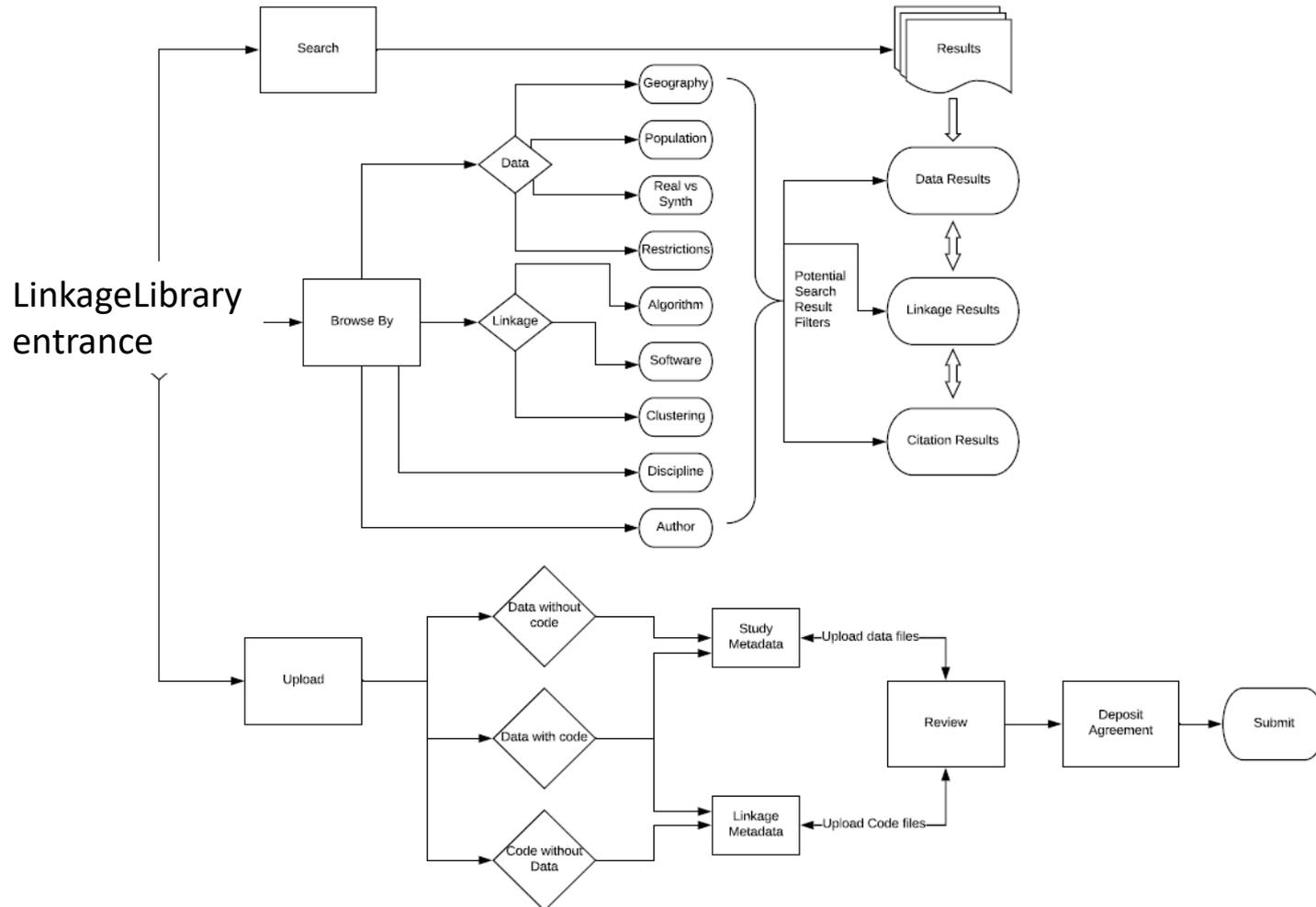
Maintaining datasets to support the data linkage community

The logo for LINKAGE LIBRARY features the text in a bold, blue, sans-serif font. Above the text is a light blue arc with a yellow dot at its left end and a blue dot at its right end, suggesting a connection or link.

# LINKAGE LIBRARY

- Encourage researchers to share linked (or linkable) data, and linkage strategies
  - Algorithms, code
- Compare approaches across projects, datasets, disciplines
  - Improve linkage practices
  - Improve transparency
- Coming this fall
  - Looking for beta testers now!

# LINKAGE LIBRARY



Publication Date: Sept 19, 2017 [Open Access](#)

This is a Data title, it can be as long or as short as needed. The text can wrap onto multiple lines if needed. Should not overlap with buttons

P1 or Owner Information goes here

Version 3 [How to Cite](#) | [Share This Project](#)

[At A Glance](#)
[Data & Docs](#)
[Bibliography](#)
[Discussion](#)
[Linkages](#)

---

**Project Citation**

Kiser, S. (2016). Action Control of Affordances in an Implicitly Cued Simon Task (Data set). Inter-university Consortium for Political and Social Research (distributor). <https://doi.org/10.3886/E100365V1>

Persistent URL: <http://doi.org/10.3886/E100365V1>

**Project Description**

**Summary:** The present study used covert means to implicitly cue responses in a Simon Task in order to investigate how participants anticipate and resolve conflicts between relevant and irrelevant stimulus information. Participants were randomly assigned to either a Non-predictive or a Predictive Simon Task in which implicit cues predicted the correct response. Results showed that, despite an overall high level of accuracy between the two tasks, when implicit cues were present the average Simon Effect was significantly smaller compared to when cues were absent. Group mean differences in Simon effect scores support that implicit priming mechanisms modulate response selection and action monitoring when conveying information about response outcomes. These results demonstrate that knowledge learned implicitly can be used to resolve conflict between relevant and irrelevant stimulus information in order to avoid non-optimal behavior, providing evidence for the role of implicit learning in action control of response affordances.

**Scope of Project**

**Subject Terms:** [Implicit Learning](#), [Response Inhibition](#), [Simon Effect](#)

**Geographic Coverage:** Washington, DC

**Collection Date(s):** 12/1/2010 – 12/1/2011 (Winter 2010 to Winter 2011)

**Universe:** Undergraduate students from the Catholic University of America ages 18 to 23 years.

**Data Type(s):** experimental data

**Collection Notes:** Mean of median reaction times for correct responses were calculated for each individual and each condition collapsed across trials for each block. Only valid cues predictive of correct responses were taken for the Predictive Simon task in order to make direct comparisons to the Non-Predictive Simon task, which contains no invalid trials.

**Scope of Project**

**Response Rate:** Fifty-three undergraduates from the Catholic University of America were recruited and received course credit to participate in the present study but only forty-eight (32 females, 16 males, aged 18 to 23 years, M = 18.92, SD = 1.22) met the inclusion criteria for the analysis.

**Unit(s) of Observation:** Accuracy, Mean of Median Reaction Times (MMRT), and Calculated Simon Effect scores (Incongruent Trial MMRT - Congruent Trial MMRT)

**Published Versions**

V3 [2017-09-19](#)  
V2 [2017-05-30](#)  
V1 [2016-12-20](#)

---

**Export**

[OAI-PMH](#)

---

**Recent Comments**

Join the discussion

[Leave a comment](#)

---

**Linkage Contributions (2)**

Merge Dataset using Perl [Perl](#)  
Global Recodes [Python](#)

[Submit your linkage code](#)

---

**Report a Problem**

Found a serious problem with the data, such as disclosure risk or copyrighted content? Let us know.

[WATCH](#) 1500
 [LIKE](#) 90
 [DOWNLOAD PROJECT](#)

# LINKAGE LIBRARY

## Others can comment

Publication Date: Sept 19, 2017 [Open Access](#)

This is a Data title, it can be as long or as short as needed. The text can wrap onto multiple lines if needed. Should not overlap with buttons

PI or Owner Information goes here

Version 3 [How to Cite](#) | [Share This Project](#)

[At A Glance](#) [Data & Docs](#) [Bibliography](#) [Discussion](#) [Linkages](#)

**Project Citation**

Kiser, S. (2016). Action Control of Affordances in an Implicitly Cued Simon Task (Data set). Inter-university Consortium for Political and Social Research (distributor). <https://doi.org/10.3886/E100365V1>

Persistent URL: <http://doi.org/10.3886/E100365V1>

**Project Description**

**Summary:** The present study used covert means to implicitly cue responses in a Simon Task in order to investigate how participants anticipate and resolve conflicts between relevant and irrelevant stimulus information. Participants were randomly assigned to either a Non-predictive or a Predictive Simon Task in which implicit cues predicted the correct response. Results showed that, despite an overall high level of accuracy between the two tasks, when implicit cues were present the average Simon Effect was significantly smaller compared to when cues were absent. Group mean differences in Simon effect scores support that implicit priming mechanisms modulate response selection and action monitoring when conveying information about response outcomes. These results demonstrate that knowledge learned implicitly can be used to resolve conflict between relevant and irrelevant stimulus information in order to avoid non-optimal behavior, providing evidence for the role of implicit learning in action control of response affordances.

**Scope of Project**

**Subject Terms:** [Implicit Learning](#), [Response Inhibition](#), [Simon Effect](#)

**Geographic Coverage:** Washington, DC

**Collection Date(s):** 12/1/2010 – 12/1/2011 (Winter 2010 to Winter 2011)

**Universe:** Undergraduate students from the Catholic University of America ages 18 to 23 years.

**Data Type(s):** experimental data

**Collection Notes:** Mean of median reaction times for correct responses were calculated for each individual and each condition collapsed across trials for each block. Only valid cues predictive of correct responses were taken for the Predictive Simon task in order to make direct comparisons to the Non-Predictive Simon task, which contains no invalid trials.

**Scope of Project**

**Response Rate:** Fifty-three undergraduates from the Catholic University of America were recruited and received course credit to participate in the present study but only forty-eight (32 females, 16 males, aged 18 to 23 years, M = 18.92, SD = 1.22) met the inclusion criteria for the analysis.

**Unit(s) of Observation:** Accuracy, Mean of Median Reaction Times (MMRT), and Calculated Simon Effect scores (Incongruent Trial MMRT - Congruent Trial MMRT)

WATCH 1500 LIKE 90 DOWNLOAD PROJECT

**Published Versions**

V3 2017-09-19  
V2 2017-05-30  
V1 2016-12-20

**Export**

[OAI-PMH](#)

**Recent Comments**

**User 1** - 09/25/2017  
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim...  
[Read more](#)

**User 2** - 09/25/2017  
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor

**Code Contributions (2)**

Merge Dataset using Perl [Perl](#)  
Global Records [Python](#)

[Submit your linkage code](#)

**Report a Problem**

Found a serious problem with the data, such as disclosure risk or copyrighted content? Let us know.

# LINKAGE LIBRARY

Share data,  
documents,  
code

Publication Date: Sept 19, 2017 [Open Access](#)

This is a Data title, it can be as long or as short as needed. The text can wrap onto multiple lines if needed. Should not overlap with buttons

P1 or Owner Information goes here

Version 3 [How to Cite](#) | [Share This Project](#)

[At A Glance](#) [Data & Docs](#) [Bibliography](#) [Discussion](#) [Linkages](#)

[Breadcrumbs](#) > [Will Appear](#) > [As Users Navigate](#) > Folders

Name	File Type	Size	Last Modified	Actions
<a href="#">Example folder</a>	Folder	-	09/13/2017 01:23:PM	...
<a href="#">Codebook</a>	application/pdf	624 KB	09/13/2017 01:23:PM	...
<a href="#">Spss Data</a>	application/spss-sav	212.8 KB	09/13/2017 01:23:PM	...
<a href="#">Stata Data</a>	application/stata-do	2.5 MB	09/13/2017 01:23:PM	...
<a href="#">Excel File</a>	text/csv	13 KB	09/13/2017 01:23:PM	...
<a href="#">PDF</a>	application/pdf	500 KB	09/13/2017 01:23:PM	...
<a href="#">Image File</a>	application/png	3 MB	09/13/2017 01:23:PM	...

# LINKAGE LIBRARY

## Share ideas

Publication Date: Sept 19, 2017 [Open Access](#)

This is a Data title, it can be as long or as short as needed. The text can wrap onto multiple lines if needed. Should not overlap with buttons

P1 or Owner Information goes here  
Version 3 [How to Cite](#) | [Share This Project](#)

[At A Glance](#) [Data & Docs](#) [Bibliography](#) [Discussion](#) [Linkages](#)

[Breadcrumbs](#) > [Will Appear](#) > [As Users Navigate](#) > Folders

Search Files and Folders

Name	File Type	Size	Last Modified	Actions
<a href="#">Example folder</a>	Folder	-	09/13/2017 01:23:PM	...
<a href="#">Codebook</a>	application/pdf	624 KB	09/13/2017 01:23:PM	...
<a href="#">Spss Data</a>	application/spss-sav	212.8 KB	09/13/2017 01:23:PM	...
<a href="#">Stata Data</a>	application/stata-do	2.5 MB	09/13/2017 01:23:PM	...
<a href="#">Excel File</a>	text/csv	13 KB	09/13/2017 01:23:PM	...
<a href="#">PDF</a>	application/pdf	500 KB	09/13/2017 01:23:PM	...
<a href="#">Image File</a>	application/png	3 MB	09/13/2017 01:23:PM	...

- Download
- Preview
- Add Comment
- Add Code
- View Citation

WATCH 1500 LIKE 90 DOWNLOAD PROJECT

# LINKAGE LIBRARY

## Protect data and files

Publication Date: Sept 19, 2017 Restricted Access

This is a Data title, it can be as long or as short as needed. The text can wrap onto multiple lines if needed. Should not overlap with buttons

PI or Owner Information goes here  
Version 3 [How to Cite](#) | [Share This Project](#)

WATCH 1500 LIKE 90 REQUEST ACCESS

[At A Glance](#) [Data & Docs](#) [Bibliography](#) [Discussion](#) [Linkages](#)

**One or more of the files in this project is restricted**

**Restricted Access**  
You may request access to the files in this upload provided that you fulfill the conditions below. The decision whether to grant or deny access is solely the responsibility of the PI/ Project owner

These data include personal identifiable information and confidential content and therefore cannot be open access except to approved/credentialed researchers. If you would like access to this project, please submit your request via the ICPSR Restricted Use Contracting System.

REQUEST ACCESS

Publication Date: Sept 19, 2017 Open Access

This is a Data title, it can be as long or as short as needed. The text can wrap onto multiple lines if needed. Should not overlap with buttons

PI or Owner Information goes here  
Version 3 [How to Cite](#) | [Share This Project](#)

At A Glance
Data & Docs
**Bibliography**
Discussion
Linkages

More published analyses of the data in this study may exist. If you know of any not listed here, please [suggest a new citation](#).

Export Sort by: Title A-Z

Pub. Type	Pub. Year	Citation
	2014	Substance Abuse and Mental Health Services Administration <b>10.8 Million Full-Time Workers Have a Substance Use Disorder.</b> <i>NSDUH Report</i> . Rockville, MD: Substance Abuse and Mental Health Services Administration. Full Text Options: <a href="#">PDE</a> <a href="#">Original source</a> Export Options: <a href="#">RIS EndNote</a> <a href="#">related studies/articles</a>
	2017	Dandu, Kartik V., Carniot, Eric T., Sanghvi, Saurin, Barocas, Soly, Eloy, Jean A. <b>A 10-year analysis of head and neck injuries involving nonpowder firearms.</b> <i>Otolaryngology - Head and Neck Surgery</i> . Full Text Options: <a href="#">DOI</a> <a href="#">WorldCat</a> <a href="#">Google Scholar</a> Export Options: <a href="#">RIS EndNote</a> <a href="#">related studies/articles</a>
	2014	Richards, Tara N., Jennings, Wesley G., Tomsich, Elizabeth, Gover, Angela <b>A 10-year analysis of rearrests among a cohort of domestic violence offenders.</b> <i>Violence and Victims</i> . 29, (6), 887-906 Full Text Options: <a href="#">DOI</a> <a href="#">WorldCat</a> <a href="#">Google Scholar</a> Export Options: <a href="#">RIS EndNote</a>
	1990	Williamson, D.F., Kahn, H.S., Remington, P.L., Anda, Robert F. <b>The 10-year incidence of overweight and major weight gain in US adults.</b> <i>Archives of Internal Medicine</i> . 150, (3), 665-672. Full Text Options: <a href="#">DOI</a> <a href="#">WorldCat</a> <a href="#">Google Scholar</a> Export Options: <a href="#">RIS EndNote</a> <a href="#">ICPSR's Standard Bibliography Display</a> <a href="#">related studies/articles</a>
	2014	Goodwin, R.D., Scheckner, B., Peria, L., Feldman, J.M., Taha, F., Lipsitz, J.D. <b>A 10-year prospective study of respiratory disease and depression or anxiety in adulthood.</b> <i>Annals of Allergy, Asthma and Immunology</i> . Full Text Options: <a href="#">DOI</a> <a href="#">WorldCat</a> <a href="#">Google Scholar</a> Export Options: <a href="#">RIS EndNote</a>
	1996	Madans, Jennifer H., Cox, Christine S., Kleinman, Joel C., Makuc, Diane, Feldman, Jacob J., Finucane, Fanchon F., Barbano, Helen E., Comoni-Huntley, Joan <b>10 Years After NHANES I: Mortality Experience at Initial Followup, 1982-1994.</b> <i>Public Health Reports</i> . 101, (5), 474-481. Full Text Options: <a href="#">WorldCat</a> <a href="#">Google Scholar</a> Export Options: <a href="#">RIS EndNote</a>
	1996	Madans, Jennifer H., Kleinman, Joel C., Cox, Christine S., Barbano, Helen E., Feldman, Jacob J., Cohen, B., Finucane, Fanchon F., Comoni-Huntley, Joan <b>10 years after NHANES I: Report of Initial Followup, 1982-1984.</b> <i>Public Health Reports</i> . 101, (5), 460-473. Full Text Options: <a href="#">WorldCat</a> <a href="#">Google Scholar</a> Export Options: <a href="#">RIS EndNote</a> <a href="#">related studies/articles</a>
	1991	Williamson, D.F., Kahn, H.S., Byers, T. <b>The 10-y incidence of obesity and major weight gain in black and white US women aged 30-55 y.</b> <i>American Journal of Clinical Nutrition</i> . 53, (6 Suppl.), 1515S-1516S. Full Text Options: <a href="#">WorldCat</a> <a href="#">Google Scholar</a>



# LINKAGE LIBRARY

## Threaded discussion

Publication Date: Sept 19, 2017 Open Access

This is a Data title, it can be as long or as short as needed. The text can wrap onto multiple lines if needed. Should not overlap with buttons

PI or Owner Information goes here

Version 3 [How to Cite](#) | [Share This Project](#)

[At A Glance](#) [Data & Docs](#) [Bibliography](#) [Discussion](#) [Linkages](#)

**TOPICS** + A more popular topic thread. With 5 comments in the thread.

search

This is a topic thread 1

**A more popular topic thread. With 5 comments in the thread.** 5

Another thread 2

**User 1** - 09/25/2017  
15  
Liked Reply Report  
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

**User 2** Owner - 09/25/2017  
-  
Like Reply Report  
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

**User 1** - 09/25/2017  
2  
Like Reply Report  
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

**User 3** - 09/25/2017  
-  
Like Reply Report  
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

**User 4** Collaborator - 09/25/2017  
-  
Like Reply Report  
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

[WATCH](#) 1500 [LIKE](#) 90 [DOWNLOAD PROJECT](#)

# LINKAGE LIBRARY

## Linkage algorithms

Publication Date: Sept 19, 2017 [Open Access](#)

This is a Data title, it can be as long or as short as needed. The text can wrap onto multiple lines if needed. Should not overlap with buttons

P1 or Owner Information goes here

Version 3 [How to Cite](#) | [Share This Project](#)

[At A Glance](#) [Data & Docs](#) [Bibliography](#) [Discussion](#) [Linkages](#)

[Add Linkage](#) Showing 1 to 2 of 2 entries. Sort by: [Date Added \(newest first\)](#)  [more options](#)

[Select](#) [Actions](#)

**50** **4**  
[Like](#) [Comments](#)

[Global Recodes](#)  
Contributor Name | Dec 9th 2017  
This linkage is a part of the process of data integration – linking combines the input sources (census, sample surveys and administrative data) into a single population, but integration also processes this population to remove duplicates/mis-matches. The first step in data linkage is to determine needs, the data availability, and whether a unique identifier exists  
**Method:** Probabilistic  
[Python](#) [Related Data Sets](#)

**30** **15**  
[Liked](#) [Comments](#)

[Merge Dataset using Perl](#)  
Contributor Name | Dec 9th 2017  
In this code, we synthesized the evidence identifying participant or population characteristics that can influence the validity and completeness of data linkage and may be associated with systematic bias in reported outcomes. The purpose of this code was to synthesize the evidence through a structured narrative review of population characteristics that may be associated with changes in sensitivity and specificity of data linkage, thereby introducing systematic bias into reported outcomes...  
**Method:** Deterministic  
[Perl](#) [Sas](#) [Related Data Sets](#)

[WATCH](#) 1500 [LIKE](#) 90 [DOWNLOAD PROJECT](#)

# LINKAGE LIBRARY

## Linkage algorithms

Publication Date: Sept 19, 2017 [Open Access](#)

This is a Data title, it can be as long or as short as needed. The text can wrap onto multiple lines if needed. Should not overlap with buttons

PI or Owner Information goes here

Version 3 [How to Cite](#) | [Share This Project](#)

[At A Glance](#) [Data & Docs](#) [Bibliography](#) [Discussion](#) [Linkages](#)

[Add Linkage](#) Showing 1 to 2 of 2 entries. Sort by: [Date Added \(newest first\)](#)  [more options](#)

Select Actions

[50](#) [4](#)  
Like Comments

[Global Recodes](#)  
Contributor Name | Dec 9th 2017  
This linkage is a part of the process of data integration – linking combines the input sources (census, sample surveys and administrative data) into a single population, but integration also processes this population to remove duplicates/mis-matches. The first step in data linkage is to determine needs, the data availability, and whether a unique identifier exists  
Method: Probabilistic  
[Python](#) [Related Data Sets](#)

**Data sets related to this linkage:**  
[Dataset 1 in DLRep Collection \(DLRep Study Number\)](#)  
[Dataset 2 in ICPSR's Collection \(ICPSR Study Nuber\)](#)  
[Data Set that has been linked outside of DLRep](#)

[30](#) [15](#)  
Like Comments

[Merge Dataset using Perl](#)  
Contributor Name | Dec 9th 2017  
In this code, we synthesized the evidence identifying participant or population characteristics that can influence the validity and completeness of data linkage and may be associated with systematic bias in reported outcomes. The purpose of this code was to synthesize the evidence through a structured narrative review of population characteristics that may be associated with changes in sensitivity and specificity of data linkage, thereby introducing systematic bias into reported outcomes...  
Method: Deterministic  
[Perl](#) [Sas](#) [Related Data Sets](#)

**Data sets related to this linkage:**  
[Dataset 1 in DLRep Collection \(DLRep Study Number\)](#)

# LINKAGE LIBRARY

## Related data & documents

Publication Date: Sept 19, 2017 Open Access

**This is a Data title, it can be as long or as short as needed. The text can wrap onto multiple lines if needed. Should not overlap with buttons**

P1 or Owner Information goes here

Version 3 [How to Cite](#) | [Share This Project](#)

[At A Glance](#) [Data & Docs](#) [Bibliography](#) [Discussion](#) [Linkages](#)

**Linkage** +

Global Recodes Leave a comment Download Code

Creator Information goes here

Merge Dataset using Perl Perl

**Global Recodes** Python

**Description:** This linkage is a part of the process of data integration – linking combines the input sources (census, sample surveys and administrative data) into a single population, but integration also processes this population to remove duplicates/mis-matches. The first step in data linkage is to determine needs, the data availability, and whether a unique identifier exists

**Format:** Python

**Method:** Deterministic

**Statistical Linkage Key (SKL):** Company ID

**True Positive Rate:** 20%

**Positive Predictor Value:** 30%

**Related Data Sets:** [\(+ Add/Upload Data\)](#)

[Perfect Data Set 1](#)

[Perfect Data Set 2](#)

[Data Set that has been linked outside of DLRep](#)

Name	File Type	Size	Last Modified	Actions
<a href="#">App</a>	Folder	-	09/13/2017 01:23:PM	...
<a href="#">Read Me</a>	application/pdf	624 KB	09/13/2017 01:23:PM	...
<a href="#">Merge Tolerance</a>	application/python	212.8 KB	09/13/2017 01:23:PM	...
<a href="#">Sensitivity Matrix</a>	application/python	2.5 MB	09/13/2017 01:23:PM	...

**User 2** - 09/25/2017

Like Report

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

# LINKAGE LIBRARY

## More about linkage strategy

Date Added: Dec 9, 2017

### Merge Dataset using Perl

Contributor Information

Version 1 [How to Cite](#) | [Share This Linkage](#)

[At A Glance](#) | [Linkage Files](#) | [Related Data Sets](#) (1) | [Bibliography](#) (2) | [Discussion](#) (5)

**WATCH** 10 | **LIKE** 30 | **DOWNLOAD LINKAGE**

#### Linkage Citation

Smith, J (2011) GraphicsDrawer source code (Version 2.0) [Source code](#) <http://www.graphicsdrawer.com> Inter-university Consortium for Political and Social Research (distributor) <https://doi.org/10.3886/E100365V1>

Persistent URL: <http://doi.org/10.3886/E100365V1>

#### Linkage Description

**Summary:** In this code, we synthesized the evidence identifying participant or population characteristics that can influence the validity and completeness of data linkage and may be associated with systematic bias in reported outcomes. The purpose of this code was to synthesize the evidence through a structured narrative review of population characteristics that may be associated with changes in sensitivity and specificity of data linkage, thereby introducing systematic bias into reported outcomes. A priori criteria for abstract inclusion were that the article empirically evaluated methodological issues relating to data linkage and reported on patient characteristics in matched versus unmatched records. Articles were excluded if they did not involve data linkage, such as a discussion of another form of health information technology, if they did not involve data, such as a commentary, letter or discussed a data linkage methodology without an empirical evaluation or if they presented a linkage project without comparing characteristics in matched and unmatched records.

**File Type(s):** [Perl](#), [Sas](#)

#### Methodology

**Linkage Method:** Deterministic

**Statistical Linkage Key (SKL):** Company ID

**Blocking/Clustering:** Bigram Indexing

**Data preparation:** Data linkers should provide analysts with record-level indicators of the data linkage process to enable adjustments for linkage error in the analyses. Indicators could include the pass-ID (the step in a rule-based linkage process when a pair of records linked), or match scores.

#### Linkage Results

**True Positive Rate:** 20%

**Positive Predictor Value:** 30%

**Errors:** While the separation principle might reduce the risk of identification, it can increase the risk of biased analyses. Linkers and analysts may be unaware of important groups who are disproportionately affected by linkage error if information is not shared between them.

#### License

This work is licensed under a [Creative Commons Attribution 3.0 Unported License](#)

#### Published Versions

V1 [2017-12-9](#)

#### Export

[OAI-PMH](#)

#### Recent Comments

Join the discussion

[Leave a comment](#)

#### Related Data Sets (1)

[Dataset 1 in DLRep Collection \(DLRep Study Number\)](#)

[Link a data set](#)

[Report a Problem](#)

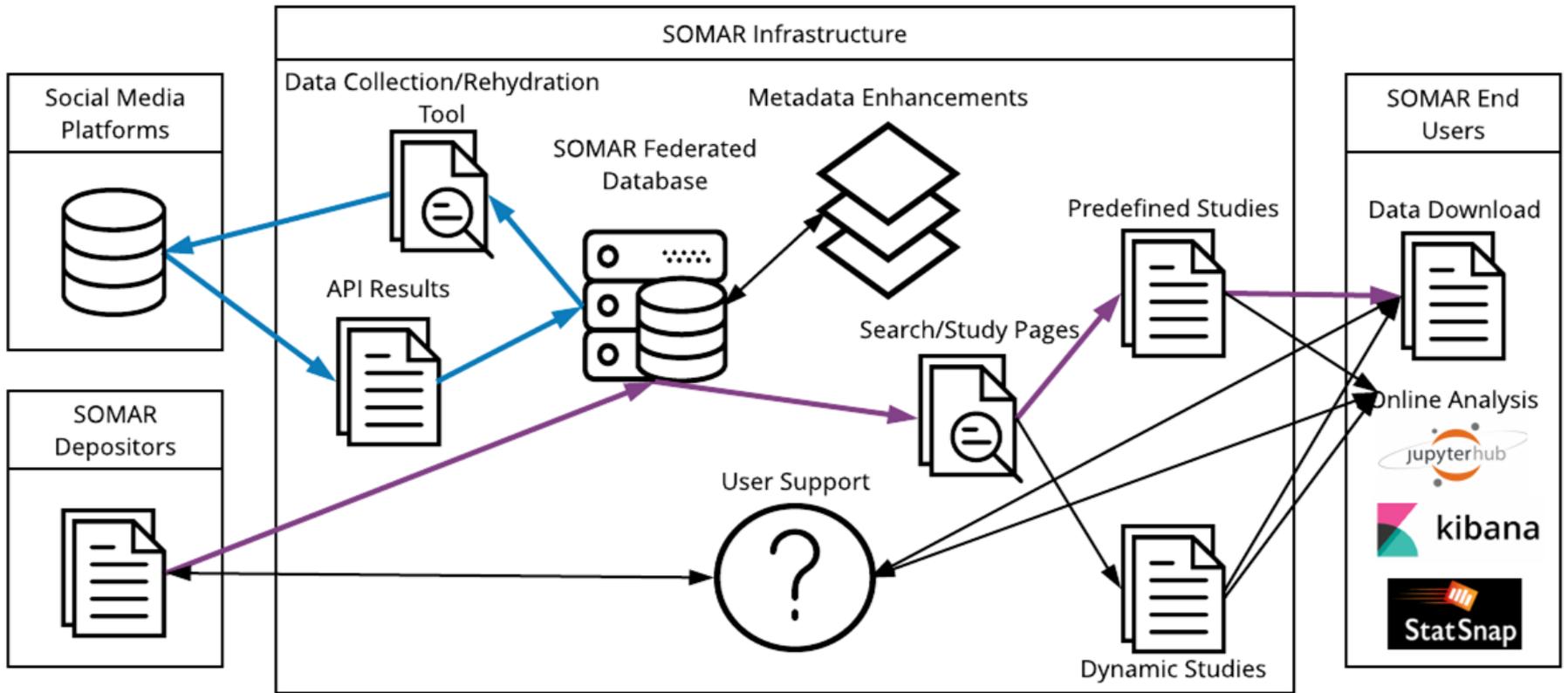
Found a serious problem with the data, such as disclosure risk or copyrighted content? Let us know.

# SOMAR: Social Media Archive

- Addresses 4 communities who:
  - study social media use specifically
  - leverage social media data to understand people and society
  - study social science methods
  - Investigate new methods for curation, publication, confidentiality and quality assessment, and long-term management of research data
- Archive enables historical and longitudinal analyses often missing from rapidly changing social medial platforms

# SOMAR: Social Media Archive

- Archive data where possible
- Archive workflows and code where data sharing is prohibited
  - Eg: Twitter IDs and code for rehydrating
- Curation and metadata
  - Provenance, dates, hashtags, confidentiality protection



# **SOMAR Challenges**

- **Technical infrastructure**
- **Ethical and legal infrastructure**
- **Metadata enhancements**
- **Adoption**

# Researcher Passport

RAAPID | HOME

Home | About | FAQs

Join now/Sign in

The Researcher Passport:  
Improving Data Access  
and Confidentiality  
Protection

Learn more

**Let your credentials speak for you!**

A researcher passport is a durable and transferable digital identifier issued by a central, community-recognized data steward. It provides credentials that identifies a trusted researcher to multiple repositories and other data custodians.

**Different levels for different needs**

The passport level (tin, copper, bronze, silver, gold, or platinum) reflects the level of trust achieved by the researcher and status with regard to those characteristics.

**PASSPORT**  
T  
TIN LEVEL  
Tin Details

**PASSPORT**  
C  
COPPER LEVEL  
Copper Details

**PASSPORT**  
B  
BRONZE LEVEL  
Bronze Details

**PASSPORT**  
S  
SILVER LEVEL  
Silver Details

**PASSPORT**  
G  
GOLD LEVEL  
Gold Details

**PASSPORT**  
P  
PLATINUM LEVEL  
Platinum Details

*Establishing shared understanding of what it means to be a trusted researcher*

# Researcher Passport

- Researcher Passport: Improving Data Access and Confidentiality Protection
  - ICPSR's Strategy for a Community-normed System of Digital Identities of Access
    - <https://deepblue.lib.umich.edu/handle/2027.42/143808>
    - Identifies inconsistent language and policies that impede access
    - Facilitate sharing of *proprietary* data
- Passports for safe people
  - Verified identities, institutional affiliation
  - Training
  - Experience (good and bad)
- Visas to control access
  - Permission to “enter” (access) specific data specifying
    - Passport holder
      - Project, Place, Period

# Passport levels

PI SCORE	DATA SECURITY LEVEL				
	LOW	MODERATE	HIGH	HIGHEST	
8+	unrestricted	secure download	secure download	VDE / physical enclave	Platinum
7	unrestricted	secure download	virtual enclave	physical enclave	Gold
6	unrestricted	secure download	virtual enclave	no access	Silver
5	unrestricted	virtual enclave	no access	no access	Bronze
4	unrestricted	virtual enclave	no access	no access	Copper
0-3	unrestricted	no access	no access	no access	Tin

# Researcher passport

- Implementation plan
  - Issue passports as digital identifiers associated with ICPSR MyData Accounts
    - Other passport issuers?
    - Blockchain
  - Repositories issue visas to access restricted data
    - Visa controls access to secure download or VDE
      - Visa issued after signed DUA
  - Establishing content of training requirements
    - Determining which existing training include that content
  - Governance system to provide due process for allegations of misconduct

# On-line passport application

## Verified identification

RAAPID HOME

http://raapid.icpsr.umich.edu/home

**RAAPID** Home | My Passport | My Visas | About | FAQs Jane Smith (log out)

**Upgrade Passport**

**TIN LEVEL**  
View Details

**Jane Smith**  
Total points: 2  
Passport level: Tin  
Access Level: Low

**Approved**  
As of: 6/13/2018  
\*pres: 6/13/2020

Personal Professional Research Experience Training Additional Details My Visas

First Name:

Last Name:

Email:   verified

Address Line 1:   
Street address, P.O. box, Company name, etc

Address Line 2:   
Apartment, suite, unit, floor, building, etc

City:

State/Province/Region:

ZIP:

Country:

# Professional qualifications and affiliations

The screenshot shows the RAAPID HOME web application interface. At the top, the browser address bar displays "http://raapid.icpsr.umich.edu/home". The RAAPID logo is on the left, and navigation links for Home, My Passport, My Visas, About, and FAQs are in the center. A user profile for Jane Smith is shown on the right with a "log out" link. Below the navigation is a blue banner with a circuit-like background, featuring a "TIN LEVEL" badge, a "View Details" button, and an "Upgrade Passport" button. The user's profile information includes: Jane Smith, Total points: 2, Passport level: Tin, Access Level: Low, and an "Approved" status as of 6/13/2018, with an expiration date of 6/13/2020. Below this is a tabbed interface with tabs for Personal, Professional, Research Experience, Training, and Additional Details. The "Professional" tab is selected, showing a "My Visas" button. The "Education" section includes fields for Highest Degree earned (Doctoral/terminal degree), Year granted (2018), Institution, and Focus/Major(s), with an "Add degree (optional)" link. The "Professional Work/Experience" section includes fields for Current Employer Type (Academic), Current Role/Appointment (Full/Associate Professor), Employer, and Years In This Role (0-2 years), with an "Add work experience (optional)" link.

RAAPID HOME

http://raapid.icpsr.umich.edu/home

RAAPID

Home | My Passport | My Visas | About | FAQs

Jane Smith (log out)

Upgrade Passport

TIN LEVEL

View Details

Jane Smith  
Total points: 2  
Passport level: Tin  
Access Level: Low

Approved  
As of: 6/13/2018  
expires: 6/13/2020

Personal Professional Research Experience Training Additional Details

My Visas

Education

Highest Degree earned: Doctoral/terminal degree

Year granted: 2018

Institution:

Focus/Major(s):

+ Add degree (optional)

Professional Work/Experience

Current Employer Type: Academic

Current Role/Appointment: Full/Associate Professor

Employer:

Years In This Role: 0-2 years

+ Add work experience (optional)

# History of data stewardship

The screenshot shows a web browser window with the URL `http://raapid.icpsr.umich.edu/home`. The page header includes the RAAPID logo, navigation links for Home, My Passport, My Visas, About, and FAQs, and a user profile for Jane Smith with a log out option. A large blue banner features a circuit-like background, a 'TIN LEVEL' profile card, and an 'Upgrade Passport' button. Below the banner, the user's name 'Jane Smith' is displayed along with their total points (2), passport level (Tin), and access level (Low). A green checkmark icon indicates an 'Approved' status, valid as of 6/13/2018 and expiring on 6/13/2020. A tabbed interface shows 'Research Experience' as the active section, with other tabs for Personal, Professional, Training, and Additional Details. The 'Research Experience' section contains several form fields: an ORCID field, a '+ Add research publication' button, a 'Do you have experience using restricted data?' question with Yes/No radio buttons, a text area for description, a 'Have you received a federal grant in the last 10 years?' question with Yes/No radio buttons, a grant number input field with a '+ Add more grant numbers (optional)' button, and another 'Have you received a federal clearance in the last 10 years?' question with Yes/No radio buttons and a text area for description. A 'My Visas' button is located in the top right corner of the form area.

http://raapid.icpsr.umich.edu/home

RAAPID Home | My Passport | My Visas | About | FAQs Jane Smith (log out)

**TIN LEVEL**  
View Details

**Upgrade Passport**

**Jane Smith**  
Total points: 2  
Passport level: Tin  
Access Level: Low

Approved  
As of: 6/13/2018  
Expires: 6/13/2020

Personal Professional **Research Experience** Training Additional Details My Visas

ORCID:

+ Add research publication

Do you have experience using restricted data?  Yes  No

If yes, please describe:

Details include inputting doi of publications along with dataset name and providing repository (if secondary data)

Have you received a federal grant in the last 10 years?  Yes  No

If yes, what is the grant number?  + Add more grant numbers (optional)

Have you received a federal clearance in the last 10 years?  Yes  No

If yes, please describe:

RAAPID HOME

http://raapid.icpsr.umich.edu/home

RAAPID Home | My Passport | My Visas | About | FAQs Jane Smith (log out)

**TIN LEVEL**

**Upgrade Passport**

**Jane Smith**  
 Total points: 2  
 Passport level: Tin  
 Access Level: Low

**Approved**  
 As of: 6/13/2018  
([cancel/upgrade passport](#)) expires: 6/13/2020

Personal Professional Research Experience **Training** Additional Details **My Visas**

Have you ever completed any Data Security training?  Yes  No

If yes, please describe:

Have you ever completed any Research Conduct training?  Yes  No

If yes, please describe:

Do you have any experience with Restricted Qualitative Data?  Yes  No

If yes, please describe:

+ Add other training

# Participating repository issue visas

The screenshot shows the RAAPID website interface. At the top, there is a navigation bar with the RAAPID logo, links for Home, My Passport, My Visas, About, and FAQs, and a user profile for Jane Smith (log out). Below the navigation bar is a large blue banner with a circuit-like pattern, featuring a 'TIN LEVEL' badge, a 'View Details' button, and an 'Upgrade Passport' button. The user profile section displays Jane Smith's name, total points (2), passport level (Tin), and access level (Low). A green checkmark icon indicates approval, with a note 'Approved As of: 6/13/2018' and a link to '(show/update passport) expires: 6/13/2020'. Below the profile is a tabbed interface with 'My Visas' selected. The 'Current Visas' section lists three visas: 1. ICPSR: Issued 6/13/2018, expires 6/14/2019, includes Dataset name 1, 2, and 3. 2. The Dataverse Project: Issued 6/13/2018, expires 12/13/2018, includes Dataset name 4. 3. CISER: Issued 6/13/2018, expires 6/30/2018, includes Dataset name 5. The CISER visa is marked as 'EXPIRED' with a red stamp. The 'Past Visas' section is currently empty.

*Note that this is conceptual. These repositories have not agreed to this!*

# Questions

- Can reach ICPSR at [help@icpsr.umich.edu](mailto:help@icpsr.umich.edu)
- Researcher Credentialing
  - Johanna Bleckman at [Bleckman@umich.edu](mailto:Bleckman@umich.edu)
- LinkageLibrary
  - Susan Leonard at [hautanie@umich.edu](mailto:hautanie@umich.edu)
- SOMAR
  - Libby Hemphill at [LibbyH@umich.edu](mailto:LibbyH@umich.edu)