

Using a Data Repository

Jessica Logan, Ph.D.

April 10, 2017

logan.251@osu.edu

Associate Director of Research
Crane Center of Early Childhood Research and Policy

Decisions, decisions...

- Decide TO Archive
- Decide WHICH repository to use
- Decide WHAT to store
- Decide you have a final, releasable dataset

Decide to use a Repository

- Some researchers are very uncomfortable with data sharing.
- Three basic arguments:
 - It was my idea to collect these data
 - I did all the work
 - It's against the ethos of science



Brian Nosek
@BrianNosek



Lawyers argue against open data arguing it is "inconsistent with accepted scientific methodology" There's work to do

Decide to use a Repository

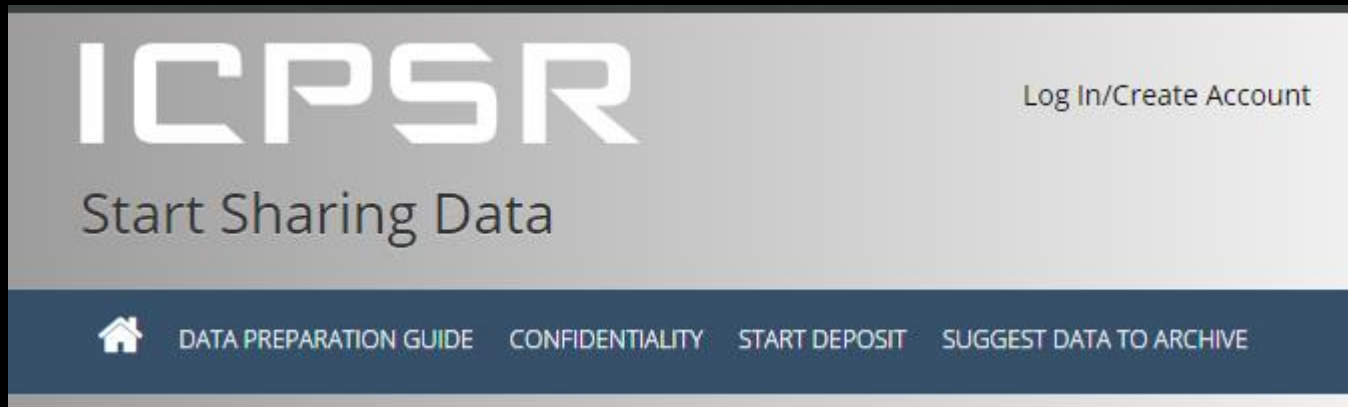
- Data sharing is the moral high ground
 - Data should be owned by the public that funded it
 - Much more can be done with it.
 - Science culture is changing
- If you or another co-investigator are hesitant, lots of resources available:
 - ICPSR Guide:
<https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/>
 - Langat et al., 2011
<https://academic.oup.com/phe/article/4/1/4/1484135/Is-There-a-Duty-to-Share-Ethics-of-Sharing>

Decide to use a Repository

- Anticipation of changes in federal policy
 - All federal grants we apply for require a data sharing and storage plan.
 - We should start now with old projects!
- Sharing data from closed projects
 - a) Allows data to have a wider reach
 - b) Makes sharing with other authors easier
 - c) Keeps data archived and accessible long-term

Decide WHICH repository

- We chose Inter-university Consortium for Political and Social Research (ICPSR):



- Lots of resources on how to share
- Good confidentiality and maintenance policies
- OSU already pays for a site license so it's free!

Decide WHICH repository

ICPSR: Share Your Data:

<https://www.icpsr.umich.edu/icpsrweb/deposit/index.jsp>

How to Share Your Data

ICPSR welcomes and encourages deposits of digital data. Deposits are made using a secure data deposit form to describe the data collection and upload content.

[Start Deposit](#)

For deposits that involve physical materials, please contact us (deposit@icpsr.umich.edu).

ICPSR is enhancing our data deposit system. Are you interested in providing feedback about the process of sharing data with ICPSR? If so, [sign up here](#).

What Should My Deposit Include?

Deposits should include all data and documentation necessary to independently read and interpret the data collection:

Decide WHAT to store

ICPSR deposit should include:

- Data file
 - Submit either an SPSS, SAS, or STATA file(s)
 - Mutually exclusive variable names
 - Must be de-identified
- Documentation
 - Codebooks / Variable Names / Data dictionary
 - Document missing codes
- Description
 - Study summary: Design, methodology, procedures

Decide WHAT to share

- Everything?
 - Probably not
- Think about:
 - Include item-level information?
 - Good documentation
 - De-identification

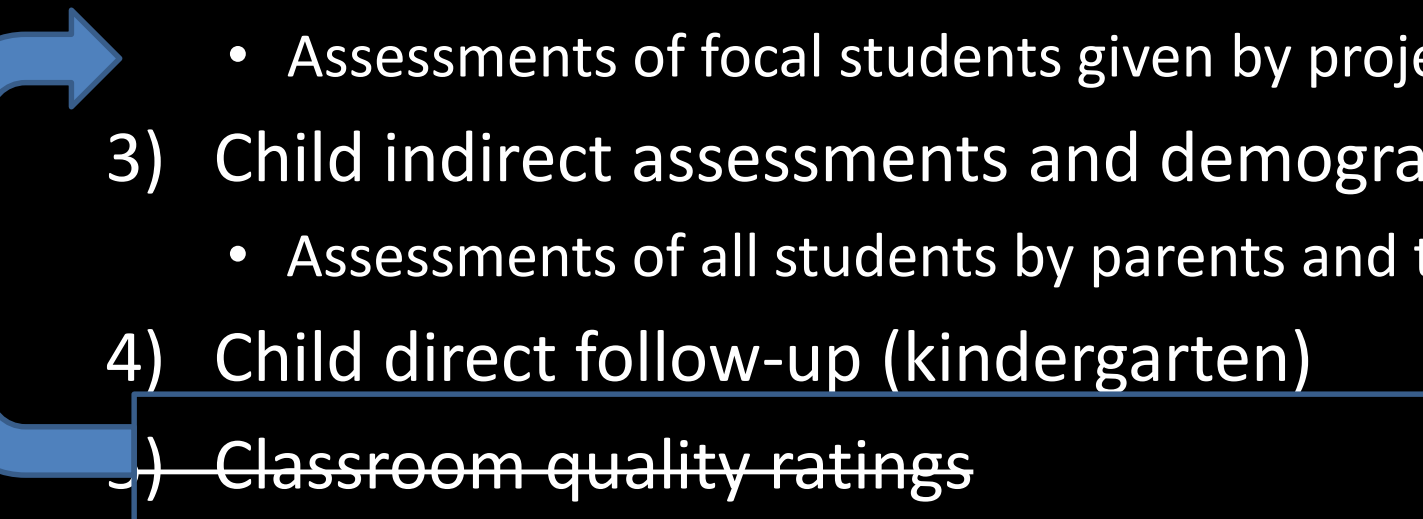
Decide WHAT to share: CASE

- STAR2: Book reading intervention for teachers, parents, and their preschoolers.
 - Focal students and non-focal got different batteries
 - Three intervention conditions
 - Two cohorts and Two sites
 - Followed focal children for 2 years

Decide WHAT to share: CASE

- Started with these internally released datasets:
 - 1) Teacher demographics and condition
 - 2) Child direct assessments Fall and Spring
 - Assessments of focal students given by project staff
 - 3) Child indirect assessments and demographics
 - Assessments of all students by parents and teachers
 - 4) Child direct follow-up (kindergarten)
 - 5) Classroom quality ratings
 - 6) Parent Intervention implementation

Decide WHAT to share: CASE

- Started with these internally released datasets:
 - 1) Teacher demographics and condition
 - 2) Child direct assessments Fall and Spring
 - Assessments of focal students given by project staff
 - 3) Child indirect assessments and demographics
 - Assessments of all students by parents and teachers
 - 4) Child direct follow-up (kindergarten)
 - ~~5) Classroom quality ratings~~
 - ~~6) Parent Intervention implementation~~
- 

Decide WHAT to share: CASE

- Started with these internally released datasets:
 - 1) Teacher demographics (~75 variables)
 - 2) Child direct assessments Fall and Spring
 - ~100 variables
 - 3) Child indirect assessments and demographics
 - ~500 variables
- Went through each and decided what to keep:

Decide WHAT to share: CASE

- Construct level first.
- *Excerpt* from measures table in grant application...

Measure, STAR 2	Sample	When Administered	Construct Being Measured	Psychometric Qualities
<i>Child Indirect Assessments: Other</i>				
Social Skills Rating System: Teacher Form (SSRS; Gresham & Elliot, 1990)	Primary/Peer	T1 T3 T4	Social skills, behavior problems	Stability coefficient (test/retest) across subtests = .84 - .93; Internal consistency coefficient across subtests (Cronbach's alpha) = .82 - .94
Student IEP & Speech/Language Service	Primary	T3	IEP services	
Child Language Sample	Primary	T1		

Decide WHAT to share: CASE

- Social Skills Rating Scale (SSRS)
 - 49 item rating scale
 - Filled out by both parents and teachers
 - Completed in fall, spring, and (for a subset of kids) one year post intervention.
 - Two subscales (Social Skills and Problem Behaviors), with four and two subparts.
 - Can calculate: Raw scores, standard scores, percentile ranks, confidence bands.

SSRS Codebook

Fall + Spring

Date Given PSRS#_DATE Teacher ID TEACH ID
Date of Birth CHILD_BIRTHDATE Child ID CHILD ID
Age PSRS_AGE#A years Month M#B Grade PSRS#G Gender: GENDER

- | | | |
|---|--|--|
| 1 | Follows your instructions | <u>PSRS#_001</u> <u> </u> <u>1A</u> |
| 2 | Helps you with household tasks without being asked | <u> </u> <u>002</u> <u> </u> <u>2A</u> |
| 3 | Appropriately questions household rules that may be unfair | <u> </u> <u>003</u> <u> </u> <u>3A</u> |
| 4 | Attempts household tasks before asking for your help | <u> </u> <u>004</u> <u> </u> <u>4A</u> |
| 5 | Gives compliments to friends or other children in the family | <u> </u> <u>005</u> <u> </u> <u>5A</u> |
| 6 | Participates in organized group activities | <u> </u> <u>006</u> <u> </u> <u>6A</u> |
| 7 | Politely refuses unreasonable requests from others | <u> </u> <u>007</u> <u> </u> <u>7A</u> |

SSRS Codebook

	PSRS3_DATE	Date Given
	PSRS_AGEM3A	Age
	PSRS3G	Grade
	PSRS3001	Follows your instructions
	PSRS3002	Helps you with household tasks without being asked
Date Given	PSRS3003	Appropriately questions household rules that may be unfair
	PSRS3004	Attempts household tasks before asking for your help
Date of Birth	PSRS3005	Gives compliments to friends or other children in the family
Age <i>PSRS_</i>	PSRS3006	Participates in organized group activities
	PSRS3007	Politely refuses unreasonable requests from others
	PSRS3008	Introduces herself or himself to new people without being told
1 Follows	PSRS3009	Uses free time at home in an acceptable way.
2 Helps y	PSRS3010	Asks permission before using another family member's property

3	Appropriately questions household rules that may be unfair	<u>003</u> <u>3A</u>
4	Attempts household tasks before asking for your help	<u>004</u> <u>4A</u>
5	Gives compliments to friends or other children in the family	<u>005</u> <u>5A</u>
6	Participates in organized group activities	<u>006</u> <u>6A</u>
7	Politely refuses unreasonable requests from others	<u>007</u> <u>7A</u>

Decide WHAT to share: CASE

- Started with these internally released datasets:
 - 1) Teacher demographics (~75 variables)
 - 2) Child direct assessments Fall and Spring
 - ~100 variables
 - 3) **Child indirect assessments and demographics**
 - ~500 variables
- Too many items... instead decided to keep construct-level variables

Decide WHAT to share: CASE

- Final inclusion of SSRS in Publishable Database:
 - Total Scores and standard scores
 - For each of the two subscales
 - At each of the three time-points
 - For each of the two raters (parents and teachers)

PSRS1SSSS	Parent: Standard Score Social Skills T1
PSRS1SST	Parent: Total Social Skills T1
PSRS1PBT	Parent: Total Problem Behavior T1
PSRS1PBSS	Parent: Standard Score Problem Behavior T1

Decide WHAT to share:

De-Identification

- Using ICPSR guidelines, we removed:
 - Birthdates
 - converted into age in months at testing
 - Any address information (zip codes)
 - Building-level information
 - Text fields
- Made new ID variables that are not linked to our data collection procedures.

Decide WHAT to share: CASE

- From Child questionnaire codebook

1 Does this child have an IEP?

SQ#001

OK

2 Does the IEP specify speech/language (S/L) services for this child?

#002

OK

From IEP please list S/L goals (summarize if necessary):

Goal 1

002A

NOT OK

Goal 2

002B

Goal 3

002C

Goal 4

002D

Goal 5

002E

3 How significant is this child's communication impairment

003

Deciding your dataset is done

- Write a study summary
 - Make sure a few different people read it
- Do a very thorough data cleaning
 - Check ranges and missing codes
- Make clear data documentation
 - Data dictionaries including variable names, labels, and ranges.

Variable	Variable Label	Value Labels	Actual Range (if applicable)
PSRS1PBT	Parent: Total Problem Behavior T1		0-20
PSRS1PBSS	Parent: Standard Score Problem Behavior T1		85-145
TSRS1SSSS	Teacher: Standard Score Social Skills T1		40-130
TSRS1SST	Teacher: Total Social Skills T1		0-60
TSRS1PBT	Teacher: Total Problem Behavior T1		0-16

Why did students withdraw: Documented exception codes

0 = typical case; no change

1 = student withdrawn; moved to another class

2 = student withdrawn by parent

3 = deceased

4 = student stopped attending school

5 = student moved

6 = student withdrawn; sibling participant in different condition

7 = teacher withdrew

8 = student withdrawn; Spanish speaking

9 = student withdrawn; incomplete consent

10 = student dropped; parent is a teacher participant

11 = student withdrawn; too young

12 = student withdrawn by parent due to previous participation










13 = student withdrawn; parent unable to keep up with schedule/"simplify life"

14 = student unable to speak

15 = student moved to itinerant services

16 = student dissented

Final Submitted Files:

-  Publications List
-  STAR 2 - Data Dictionary Directs
-  STAR 2 - Data Dictionary Indirects
-  STAR 2 - Data Dictionary Teacher Demo
-  STAR 2 Measures Tables
-  STAR 2 Study Summary and Methods
-  Star2DIRECT_deident_032415
-  Star2INDIRECT_deident_032415
-  Star2TEACHERDEMO_deident_032415

Submit files, and wait

- From decision to archive to submission was about 1 year
- From submission to availability online was about 4 months.

Find & Analyze Data

[FIND DATA](#) [SEARCH/COMPARE VARIABLES](#) [DATA-RELATED PUBLICATIONS](#) [RESOURCES FOR STUDENTS](#) [HELP](#)

[REVISE SEARCH](#)

Studies (1,780)

[Variables](#)

[Publications](#)

Your query returned 628 studies. Series results are grouped under the most relevant study. [Ungroup series results](#) to view all results.

Sort by:

Study Relevance ▼

1. [Sit Together and Read in Early Childhood Special Education Classrooms in Ohio \(2008-2012\) \(ICPSR 36738\)](#)
Justice, Laura

ICPSR Available:

Dataset(s)

DS0: Study-Level Files

Documentation: [Documentation.pdf](#)

All the documentation we made is here

DS1: Child Direct Measures - [Download All Files \(4.9 MB\)](#)

Documentation: [Frequencies.pdf](#)

They ran basic information on the sample

Download: [SAS](#) [SPSS](#) [Stata](#) [R](#) [ASCII](#) [Excel/TSV](#)
[ASCII + SAS Setup](#) [SPSS Setup](#) [Stata Setup](#)

Analyze Online: [simple crosstab/frequency](#) [SDA](#)

Can analyze online

Look at all the formats!

<https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/36738?searchSource=revise&q=Sit+Together+and+Read+>

Sit Together and Read in Early Childhood Spec (2008-2012), Child Indirect Measures

Row:

select row variable (required) ▼

[recode](#)

Column:

select column variable ▼

[recode](#)

Control:

select control variable (optional) ▼

[recode](#)

Selection Filter(s):

Example: age(18-50)

Percentaging:

Column Row Total

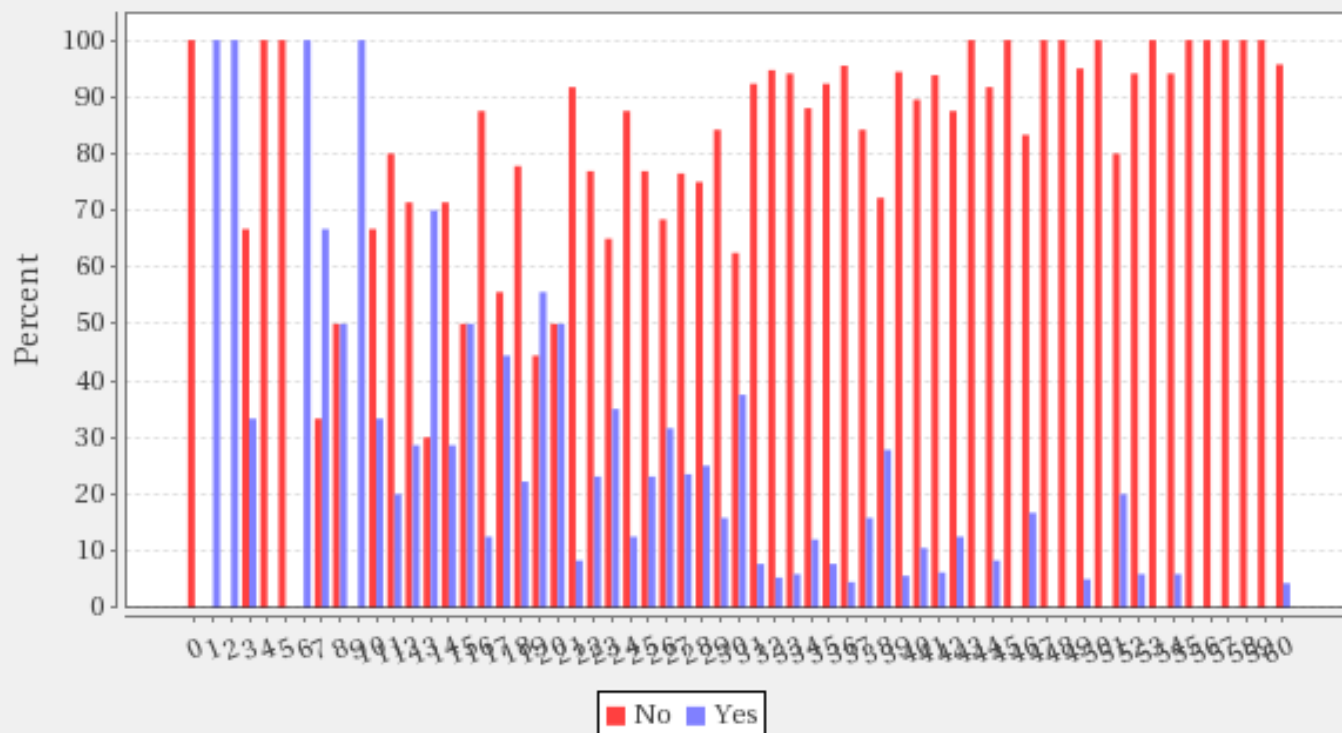
Summary statistics:

Show summary statistics

Build the Table

Row: CSQ1006: Does this child have any known diagnoses (e.g., ... [recode](#)
Column: TSRS1SST: Teacher: Total Social Skills T1 [recode](#)
Control: select control variable (optional) [recode](#)
Selection Filter(s): Example: age(18-50)
Percentaging: Column Row Total
Summary statistics: Show summary statistics

Does this child have any known diagnoses (e.g., autism, Down syndrome, hearing loss, ADHD)? BY Teacher: Total Social Skills T1





Open Data Cambridge

@CamOpenData

Follow

Research process takes so many steps, but the current focus is on the last one 'publish'. Reproducibility important everywhere.

@OSFramework

