



# Open Video Data Sharing



## The Databrary Project

NSF BCS-1238599  
NICHD U01-HD-076595

# Afternoon Schedule

- ▶ 12:30 pm - 1:30 pm
  - ◀ Lunch
  - ◀ Community forum on open data sharing

# Afternoon Schedule

- ▶ 1:45 pm - 3:15 pm
  - ◀ Introducing Databrary
  - ◀ Databrary functionality
- ▶ 3:15 pm - 3:30 pm
  - ◀ Break
- ▶ 3:30 pm - 5:00 pm
  - ◀ Advanced Datavyu OR
  - ▶ Advanced Databrary

# Logistics

- ▶ Wifi
  - ◀ ID: Databrary
  - ◀ Password: Datavyu1
- ▶ Lunch is free
- ▶ Databrary/Datavyu booth
- ▶ Karen's talk
  - ◀ "Good Behavior: Coding, Sharing and Repurposing Video", Saturday 9:55 am, Level 2, 204 C



## SRCD PRECONFERENCE WORKSHOP: CODING AND SHARING VIDEO WITH DATABRARY AND DATAVYU

Event Date: **March 18, 2015**

Databrary will hold a pre-conference workshop before SRCD on Wednesday March 18th, 2015 at the Philadelphia Convention Center. Our pre-conference will have a Datavyu and Databrary workshop with one-on-one training and a Databrary luncheon demonstrating our newest features.

*Event, posted March 18, 2015.*

### UPCOMING EVENTS

MARCH 18, 2015

**SRCD Preconference Workshop:  
Coding and Sharing Video with  
Databrary and Datavyu**

<http://databrary.org/post/pre-conference.html>

# Databrary Team

- ▶ Karen Adolph, Project Director
- ▶ Rick Gilmore, Associate Director
- ▶ David Millman, Assistant Director

# Databrary Team

- ▶ Dylan Simon, Technical Director & Systems Architect
- ▶ Lisa Steiger, Community Liaison & Project Manager

# Databrary Team

- ▶ Gladys Chan, UI/UX Designer
- ▶ Vicky Foo, Databrary & Datavyu Quality Assurance
- ▶ Drew Gordon, Information Engineer
- ▶ Tom Gebert, Software Developer
- ▶ Omi Majumder, Databrary Quality Assurance & Curation
- ▶ Lina Wictorén Roy, Project Coordinator



# Outline:

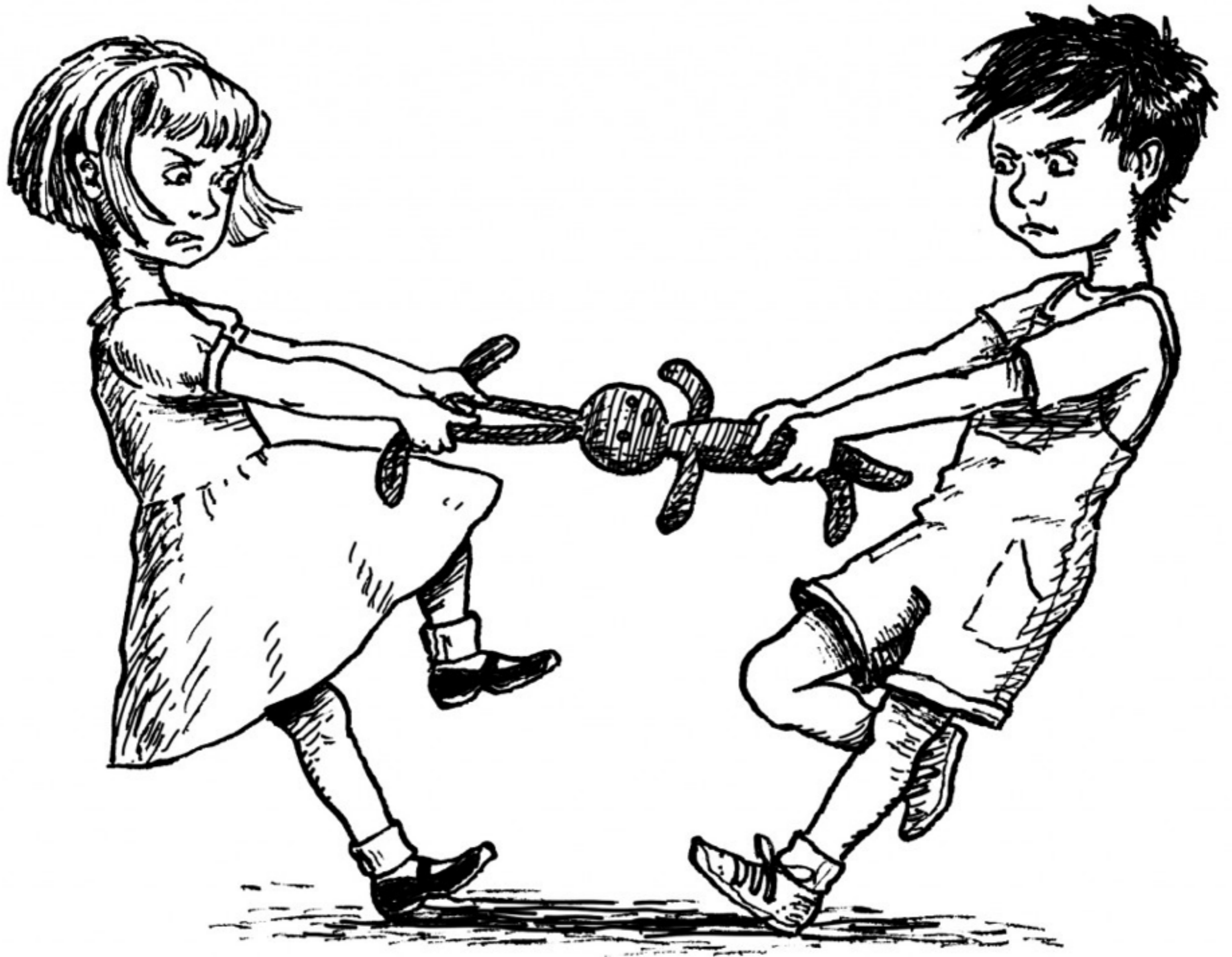
## A community forum on open data sharing

- ▶ What is open data sharing?
- ▶ Why everyone should share openly
- ▶ Why everyone doesn't
- ▶ Moving forward
- ▶ The Databrary vision



# What is open data sharing?

- ▶ What is sharing?











# Different kinds of sharing





# Different kinds of sharing



# Different kinds of sharing



SPM

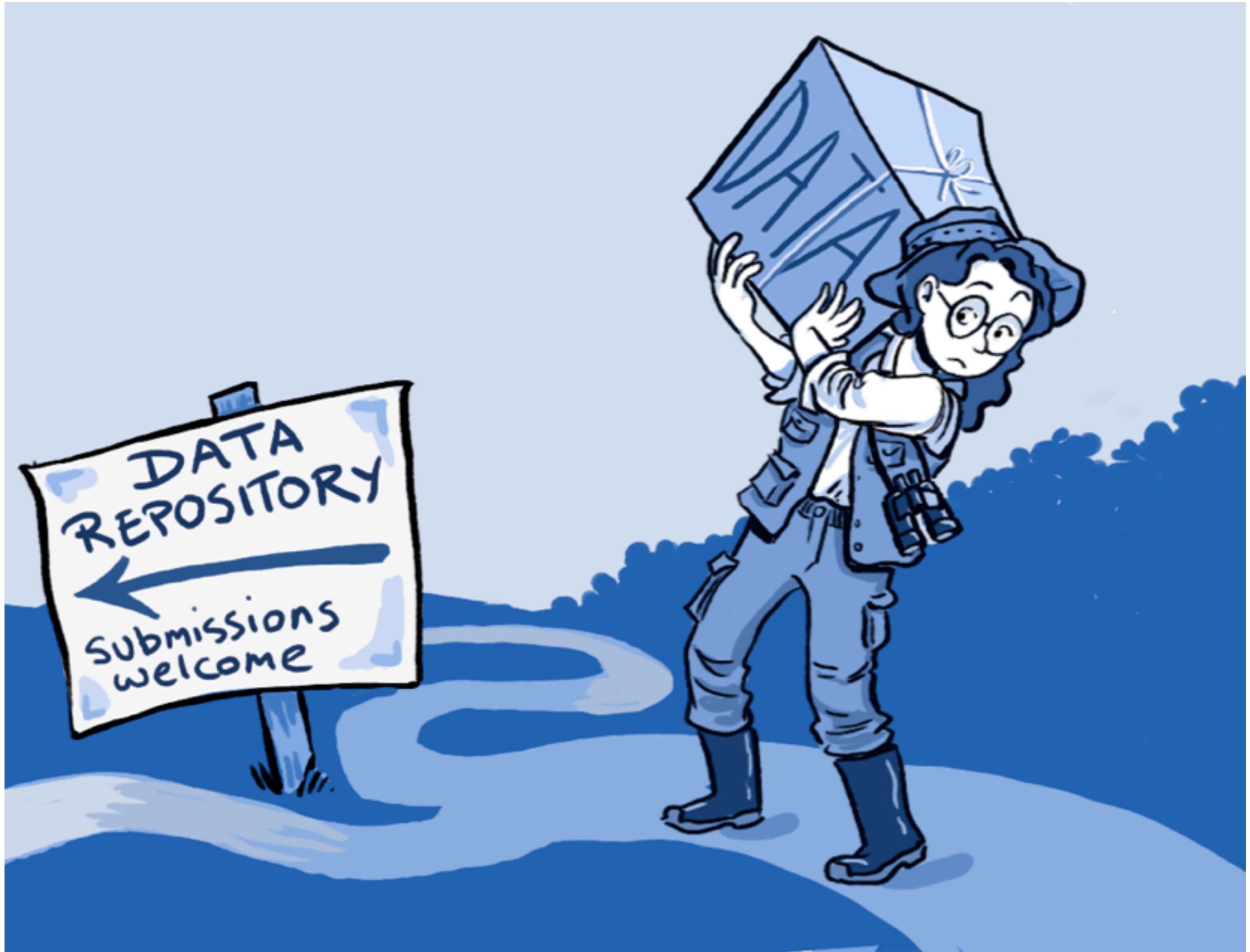


*By members & collaborators of the Wellcome Trust Centre for Neuroimaging*  
[Introduction](#) | [Software](#) | [Documentation](#) | [Courses](#) | [Email list](#) | [Data](#) | [Extensions](#)

**SPM Courses**

**London, May & October 2015**





# Sharing *data* with

- ▶ Lab colleagues
- ▶ Collaborators on project
- ▶ Collaborators *not* on project
- ▶ Other researchers by request
  - ◀ co-authorship required

# *Open* data sharing

- ▶ With other researchers
  - ◀ Without your approval or permission
  - ◀ People you might not know
  - ◀ or like
  - ◀ or agree with
  - ◀ No requirement for co-authorship
  - ◀ Expectation of proper credit, citation

# Your turn

- ▶ What do you think about data sharing?
- ▶ Have you shared? Why or why not?
- ▶ What do you think about *open* data sharing?
- ▶ Have you shared openly? Why or why not?
- ▶ Is open data sharing desirable?



# Why everyone should *openly* share data



# Why everyone should

- ▶ It's the right thing to do
- ▶ Shows leadership
- ▶ Others will benefit, and so will you
- ▶ Truth, justice, and the scientific way

# Why everyone should

- ▶ Maximizes your hard efforts
- ▶ The more you believe in your data, the more you should want to share
- ▶ More citations, credit, renown
- ▶ More invitations to co-author, collaborate
- ▶ Your data are better than your arch rival's

# Why everyone doesn't



<https://www.youtube.com/watch?v=N2zK3sAtr-4>



# Articles are only a sample of what was done

- ▶  $n=1$
- ▶ Supplemental materials may not help
- ▶ Some journals don't accept them

# Top 10 Reasons Everyone Doesn't Share Data

# 10. Data mining means...

- ▶ mine and only mine
- ▶ I am still analyzing the data to write more papers
- ▶ I have applied for a grant, and have to wait for the response
- ▶ I am planning a textbook/autobiography/art gallery, and sharing the data would preempt it

# 9. Arrogance

- ▶ I like to be selective in my choice of collaborators
- ▶ Other shared data (not mine) is of poor quality; I don't want to be associated with such low standards
- ▶ I don't have to
- ▶ Data mining/re-use isn't real science

# 8. It's a tough world

- ▶ It took me 5 years to collect! Why should I give it away?
- ▶ People should acquire their own data. It's good experience for them/you
- ▶ Data sharing doesn't help me one bit

# 7. Personal Commitments

- ▶ My boss/collaborator won't let me
- ▶ I have an exclusive agreement to share only with X
- ▶ I promised the data to a competitor of yours

# 6. Creative Lies

- ▶ My hard drive crashed
- ▶ Data is encoded, and I forgot the password
- ▶ Give me a few more weeks. I am thinking about the best way to distribute the files

# 5. Mistrust

- ▶ Others will misuse/overinterpret the data
- ▶ Data only good for the analysis I already published
- ▶ I'll give it to you, but not to Dr. X



# 4. Magical thinking

- ▶ I am setting up my own database — I will put the data in it
- ▶ Ok, I'll give it to you. Will you send me a reminder next month?

# 3. Fear of embarrassment

- ▶ My data aren't organized
- ▶ My data are organized, but not to outsiders
- ▶ My file/variable names are idiosyncratic
- ▶ I wrote my own software
- ▶ Yes, I still use Excel 1.0 on Windows 3.1

## 2. Time & money

- ▶ I have a series of urgent deadlines. I will send it to you when I have time/retire/get a sabbatical/win the lottery
- ▶ I have to publish new papers; I don't have time or staff to find/clean/convert/rename old files
- ▶ People will have issues with the data, and I don't have time to rebut them

# 1. Privacy, confidentiality, ethics

- ▶ I collect identifiable data
- ▶ My participants are infants, children, and adolescents
- ▶ I don't have permission from my participants, their parents, collaborators, ethics board/IRB

# Your turn

- ▶ Why don't we share openly?
- ▶ What reasons for *not* sharing openly should the community reject?
- ▶ What reasons for *not* sharing openly should the community acknowledge and ameliorate?



**We the People** of the  
insure domestic Tranquility, provide for the common defence,  
and our Posterity, do ordain and establish this Constitution

Section 1. All legislative Powers herein  
Section 2. The House of Representatives  
shall be composed of Members chosen every second Year by the People of the several States, and the Electors in each State shall have the Qualifications requisite for Electors in that State.







# Databrary

Access ▶ Sharing Data ▶ Policies ▶

## DATA SHARING MANIFESTO

*This document is also available as  
.pdf or .md.*

### **Data Sharing Manifesto**

- Databrary's primary purpose is to foster open data-sharing among a community of researchers in developmental science in particular, and in psychological and neural science more generally.

<http://databrary.org/access/policies/data-sharing-manifesto.html>

# Moving forward

## ▶ Ownership

- ▶ We don't own our data
- ▶ We are its stewards
- ▶ Data stewards responsible for what, when, and with whom to share
- ▶ But, should share data and materials as widely as possible



# Moving forward

## ▶ Credit

- ▶ Citation is the currency of scholarship
- ▶ Data should be easy to cite
- ▶ Sharing data early resolves claims about who discovered what first
- ▶ Data re-use and repurposing is awesome

# Moving forward

## ▶ Blame

- ▶ Own your mistakes and fix them
- ▶ Sharing data will help find and fix them faster
- ▶ Science is (or should be) “self-correcting”
- ▶ No publicity is bad publicity

# Moving forward

## ▶ Time

- ▶ Save it by curating data for sharing as you collect it
- ▶ Find old projects you discarded
- ▶ Save it by putting data in one place, common format, preserved for the future
- ▶ Share any time

# Moving forward

## ▶ Money

- ◀ Get more by sharing
- ◀ Save on collection by re-using

# Moving forward

- ▶ Privacy, confidentiality, ethics
  - ◀ Embrace common principles
  - ◀ Enforce community standards
  - ◀ Respect participants' wishes

# Databrary vision

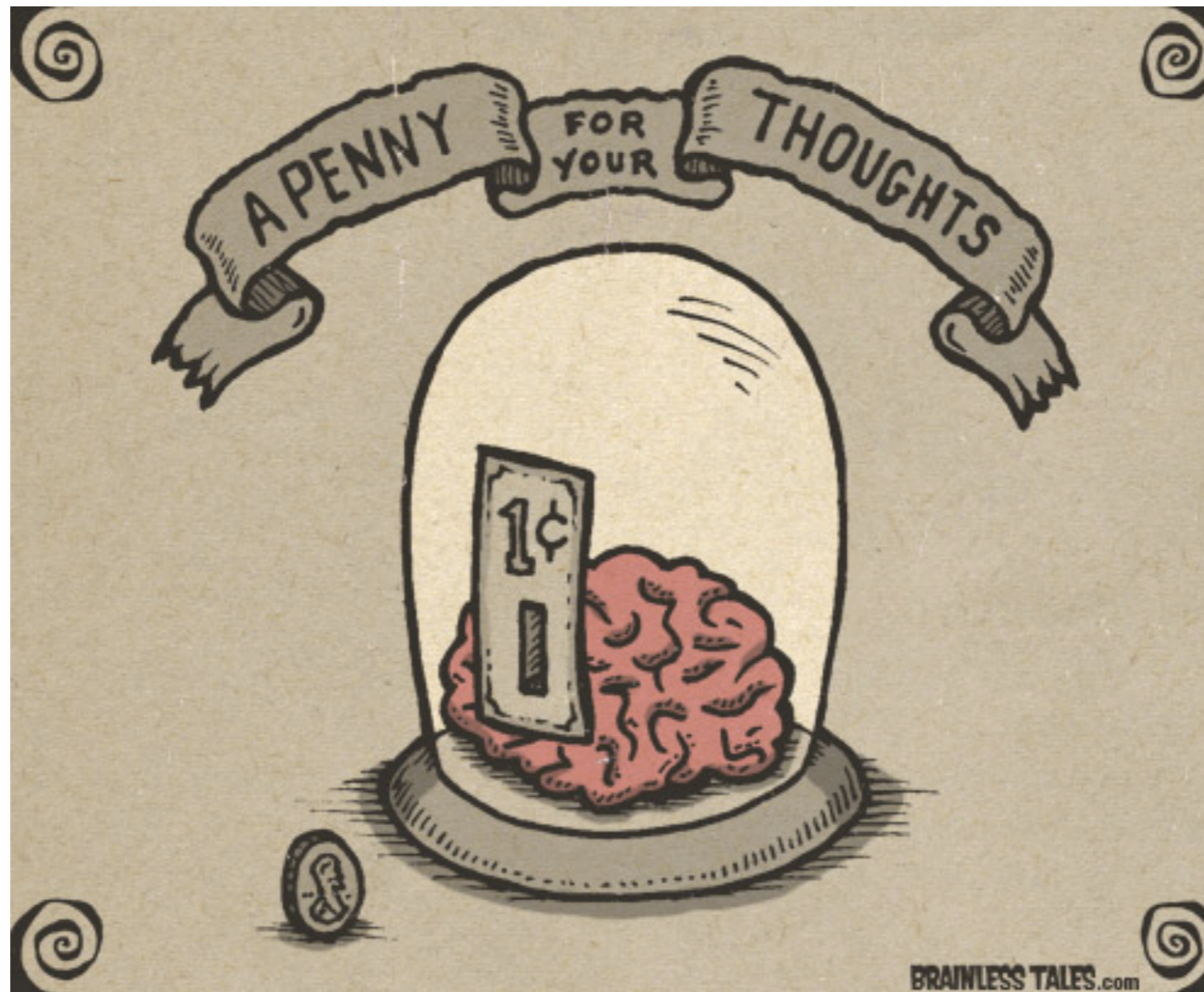
- ▶ Everyone shares *openly*
- ▶ Sharing is secure, easy, fun
- ▶ Sharing saves time, money
- ▶ Data are easy to find, preserved for the future
- ▶ More citations, more impact
- ▶ Discover more about child development faster



# Databrary vision



# Your thoughts?







Register

## Get Databrary Access

### Create Account

Register to browse shared data and comment on volumes.

### Sign Agreement

Learn about the responsibilities of all Databrary members.

### Confirm Email

Confirm your email address to continue.

### Set Password

Pick a secure password.

### Request

#### Authorization

Begin the authorization process.

## Create Account

Please provide information about yourself.

FULL NAME

EMAIL

AFFILIATION



Databrary is an open data library for developmental science.  
Share video, audio, and related metadata. Discover more, faster.

Get started with our [User Guide](#), and get the [latest news](#) here.



[Your Profile](#)

Manage your data and authorizations.



[Browse Data](#)



[Browse People](#)