

# XX Biennial ICIS in *New Orleans*

Hilton New Orleans Riverside Hotel  
May 26-28, 2016



Science, music, and food meet in NOLA

*Invited speakers include:*

Linda Smith

Nim Tottenham

Paul Bloom

Jim Rehg

Brian Scassellati

Helen Neville

Laurie Santos

Felix Warneken

and many more....

Program Chairs: Lisa Oakes, Vanessa LoBue, Marianella Casasola

For additional details (as they become available):  
<http://www.isisweb.org/view/0/conferences.html>



# Open Video Data Sharing



## The Databrary Project

NSF BCS-1238599  
NICHD U01-HD-076595

# Special promises of video

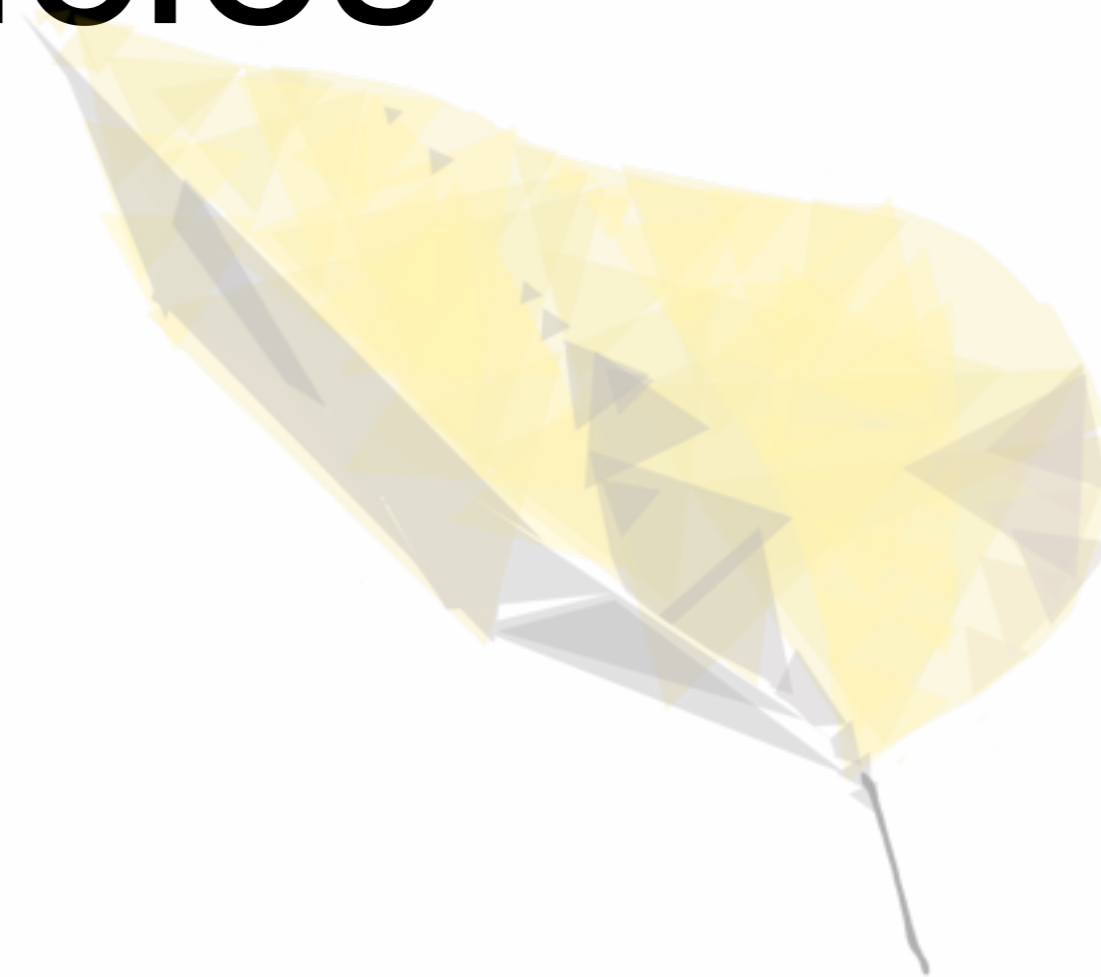
- ▶ Captures richness & complexity of behavior in real time
- ▶ Widely used in developmental, education, social, & behavioral science
- ▶ Cheap & easy to collect
- ▶ Self-documenting
- ▶ High potential for reuse

# Special challenges of video

- ▶ Identifiable data needs special policies
  - ◀ Faces, voices, names, places, classroom
- ▶ Technical issues of sharing
  - ◀ Large files to store/stream, formats
  - ◀ Search for videos, inside videos
- ▶ Cultural barriers to sharing
  - ◀ We haven't done it, we don't know how to do it, no incentives to do it



# Policies



# 3 ethical principles

## Do what you've always done!

- ▶ Ensure that participants' wishes about sharing their data are respected just as you've done for excerpts
- ▶ Treat Databrary data with the same high standard of care that you treat data collected in your own lab
- ▶ Take responsibility for your students' use of Databrary data just as you do for your own data



# Databrary as a walled garden



# Restricted access

- ▶ Authorized researchers
  - ◀ Real researcher/professor with “PI status” at a university or other institution
  - ◀ Governed by an IRB (ethics board)
  - ◀ Developmental psychologists, learning/education researchers, clinical researchers
- ▶ Authorized affiliates
  - ◀ Students/staff of authorized researchers



# Why restrict access?

- ▶ Extends the zone of trust beyond the PI's lab to other PIs in other labs
- ▶ Developmental researchers are more comfortable sharing their data & asking participants to share with other researchers like themselves
- ▶ Participants comfortable sharing with other researchers

# Access model

- ▶ Unique from other sharing agreements
- ▶ User & contributor privileges to library
  - ◀ Research & non-research uses
  - ◀ IRB approval not required by Databrary for authorization to access data
- ▶ Requires institution to co-sign researcher's agreement
  - ◀ They (not us) are responsible for you

# It is OK to share identifiable data if participants agree!

- ▶ Raw video files
- ▶ Excerpts
- ▶ Metadata (birth date, test date, location, ethnicity, disability status, etc)
- ▶ Codes of behaviors

# 5 Databrary release levels

- ▶ Did not ask
  - ◀ Functions as private

# 5 Databrary release levels

- ▶ Did not ask
- ▶ Private
  - ◀ Only researchers on protocol

# 5 Databrary release levels

- ▶ Did not ask
- ▶ Private
- ▶ Shared
  - ◀ Only authorized researchers



# 5 Databrary release levels

- ▶ Did not ask
- ▶ Private
- ▶ Shared
- ▶ Shared + Excerpts
  - ◀ Authorized researchers can show excerpts for research & educational purposes)

# 5 Databrary release levels

- ▶ Did not ask
- ▶ Private
- ▶ Shared
- ▶ Shared + Excerpts
- ▶ Public
  - ◀ Available to everyone

# 5 Databrary release levels

- ▶ Did not ask
- ▶ Private
- ▶ Shared
- ▶ Shared + Excerpts
- ▶ Public

# Permission to share vs consent to participate

- ▶ Consent to participate *is separate* from permission to share
- ▶ Consent obtained *before* session
- ▶ Release to share obtained *after* session
  - ◀ Clear to participants what was recorded

# Asking participant release to share



[databrary.org/user-guide/getting-started/release-script/example-videos.html](http://databrary.org/user-guide/getting-started/release-script/example-videos.html)

# Just do it!

- ▶ Asking to share does **not** obligate you to share
- ▶ Can't share without asking
- ▶ We'll store all the files, so long as participants are asked
- ▶ Going forward, incorporate Databrary release into your protocol



# Databrary garden

- ▶ Get authorized so you can play in the garden



# In the Databrary garden



# In the Databrary garden non-research uses

Learn about procedures

Browse videos

Curious

Nose around

See displays

Learn about coding rules

See what colleague is up to

Get ideas

Get inspired!



# In the Databrary garden pre-research uses

Decide whether a study  
is worth doing

Collect preliminary  
data for grant proposal

Replicate, expand, or review  
work based on viewing the  
procedure or codes

# In the Databrary garden teaching & didactic uses



Find excerpts for teaching

Practice coding

Show excerpt in talk

Illustrate an idea

Share excerpts to  
promote your own work

# In the Databrary garden data management

Use Databrary as your  
lab server

Share videos with lab  
staff & collaborators

Use Databrary to back  
up & preserve your data

Complete enrollment for  
grant progress report



# In the Databrary garden reuse data for research



Reuse data to ask  
new questions

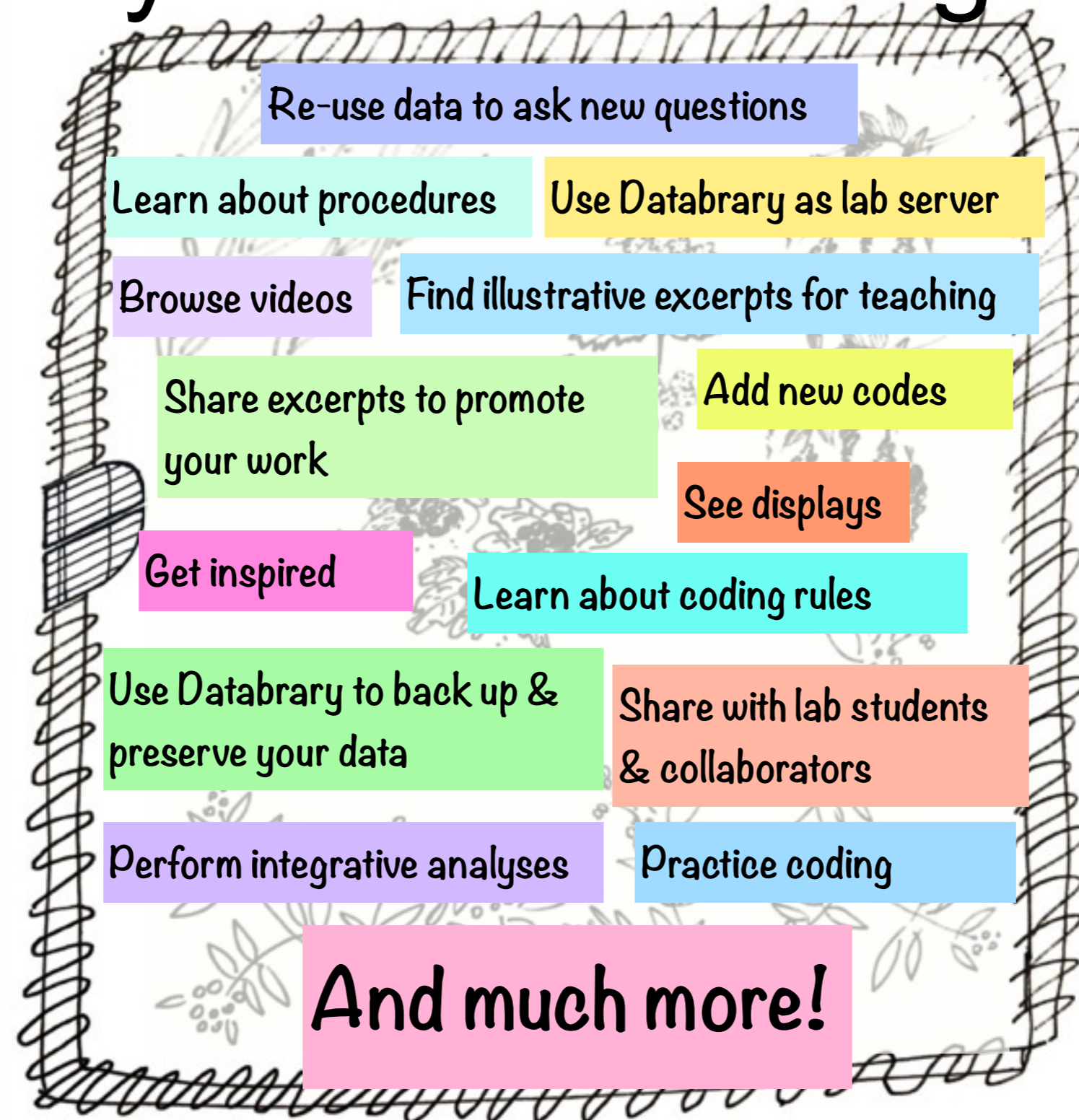
Perform integrative  
analyses

Add new codes

Include Ss from  
other populations

Grow sample size

# In the Databrary garden so many wonderful things to do!



# Curation? Upload as you go

- ▶ Don't wait to upload until study is done
  - ◀ After the fact curation is painful
  - ◀ We will help, but it will hurt at a little
- ▶ Upload as you go is easy curation
  - ◀ Upload videos & tag with basic metadata at end of each session
  - ◀ You've got to put it somewhere anyway
  - ◀ Data management & backup
- ▶ Databrary = your lab server

# Best practices for sharing

- ▶ Upload as you go spreadsheet requests info useful for search
  - ◀ Child's language(s), session language
  - ◀ Data collection setting, location
  - ◀ Disability status
  - ◀ Even if you don't use this info
- ▶ You have this info at your fingertips at end of session
  - ◀ No work to enter it now, is work later

# Best practices for sharing

- ▶ Accessibility helps everyone
  - ◀ Avoid lab-specific acronyms & terms

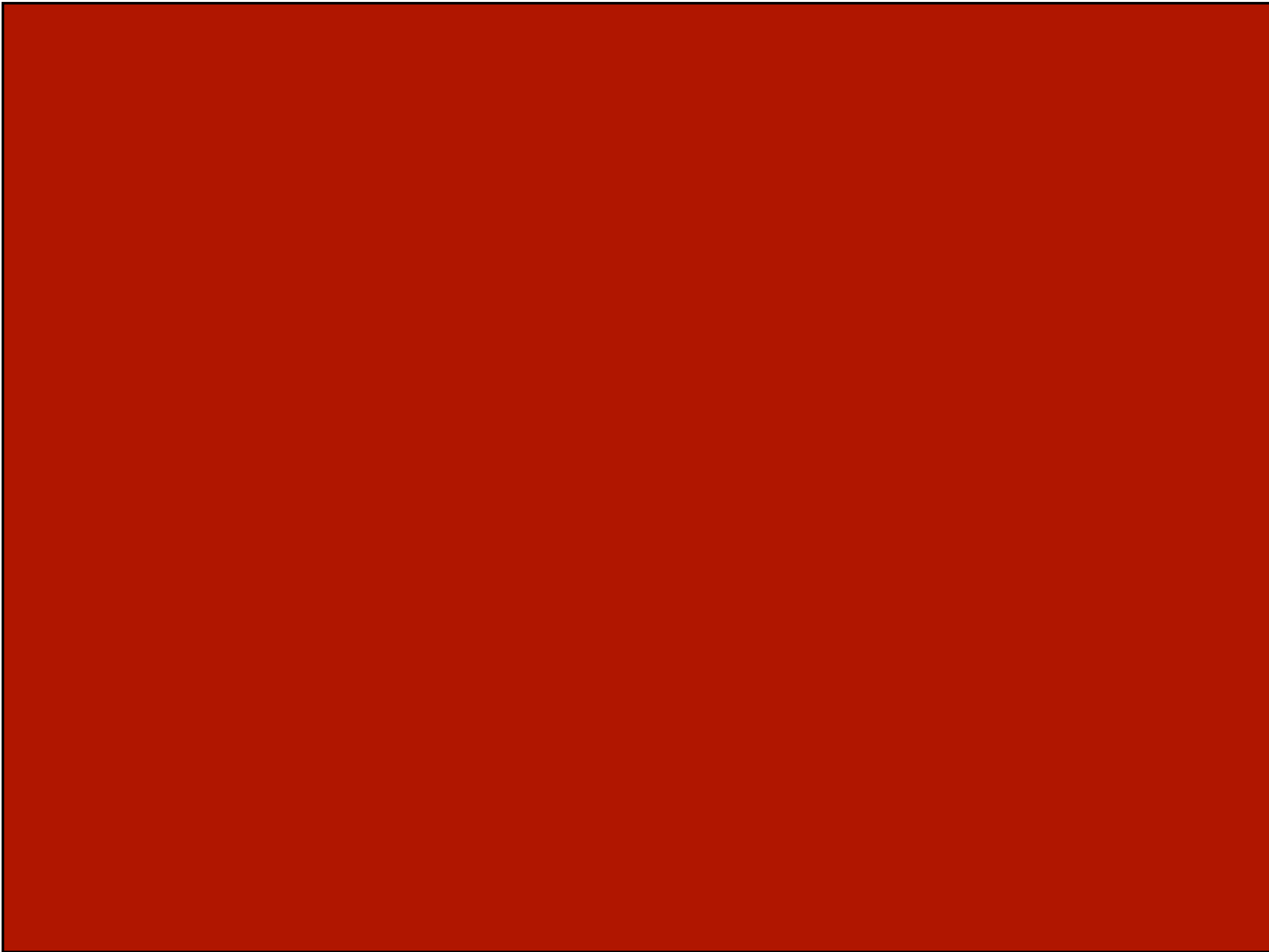
ET CW First Steps Baby OH  
CLB HK SFR emb VT14  
18MA Pedestal  
SOM SE  
18MA-NoCurve Emo Lights



# Best practices for sharing

- ▶ Accessibility helps everyone
  - ◀ Avoid lab-specific acronyms & terms
  - ◀ Use plain English
  - ◀ Establish file naming conventions
- ▶ Databrary will keep you organized
- ▶ You will never lose a file

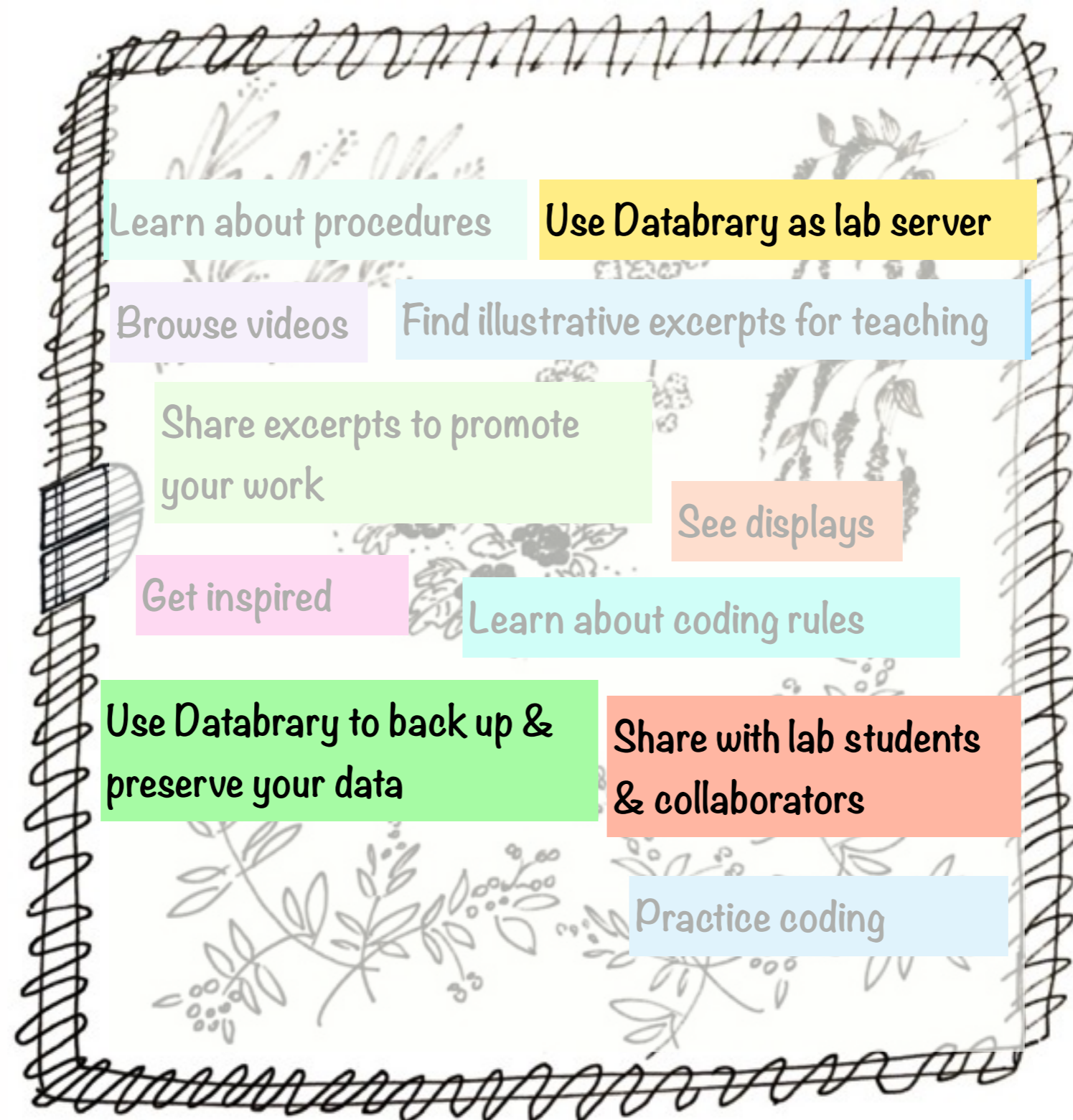




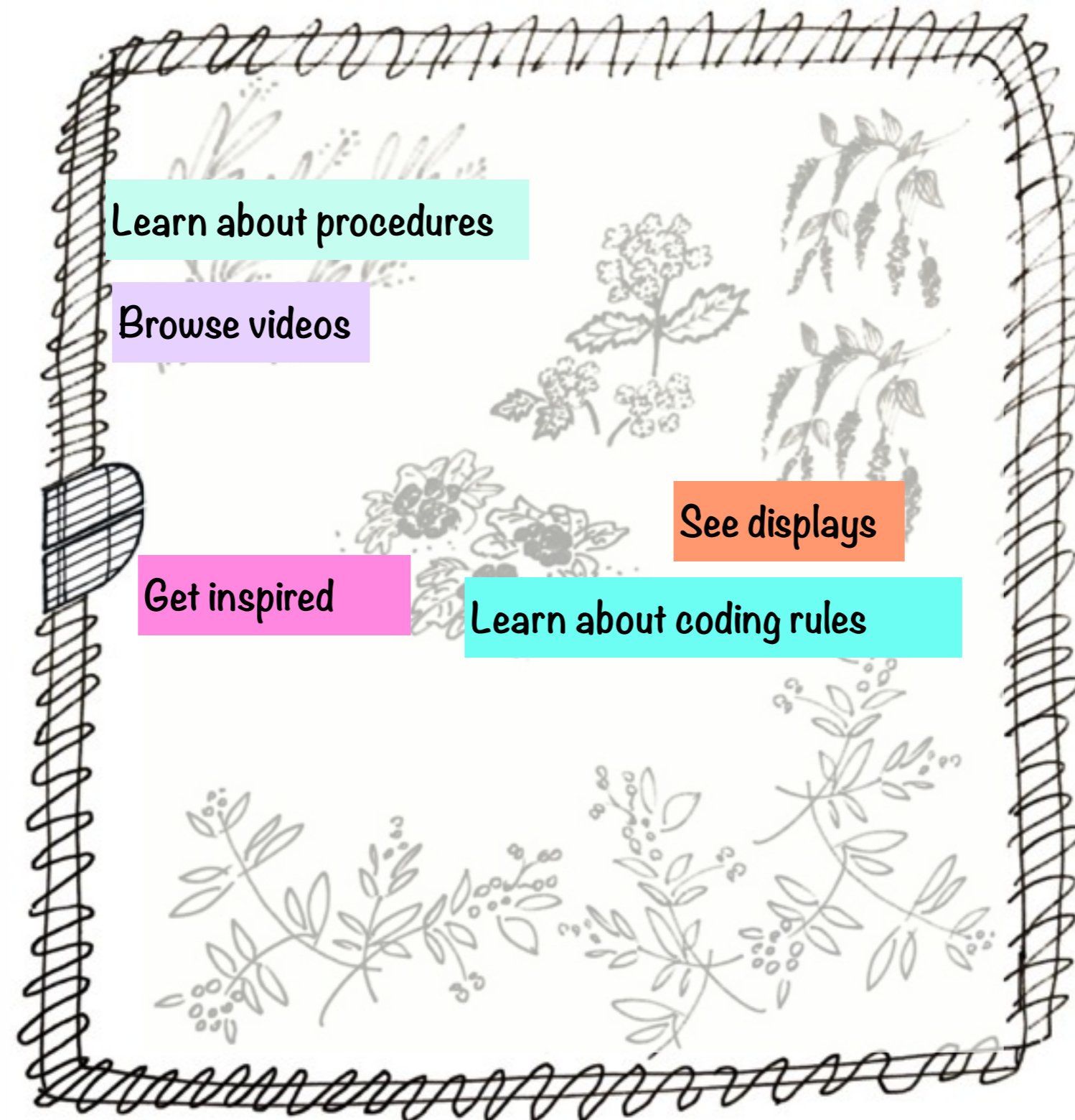
# Curation? Upload as you go

- ▶ Describe study design (conditions, groups, etc.)
- ▶ Enter (type/paste/import) session metadata
- ▶ Upload new video, associate it with session
- ▶ Keep track of which data were entered
- ▶ Export previously entered data for

# In the Databrary garden



# In the Databrary garden



Learn about procedures

Browse videos

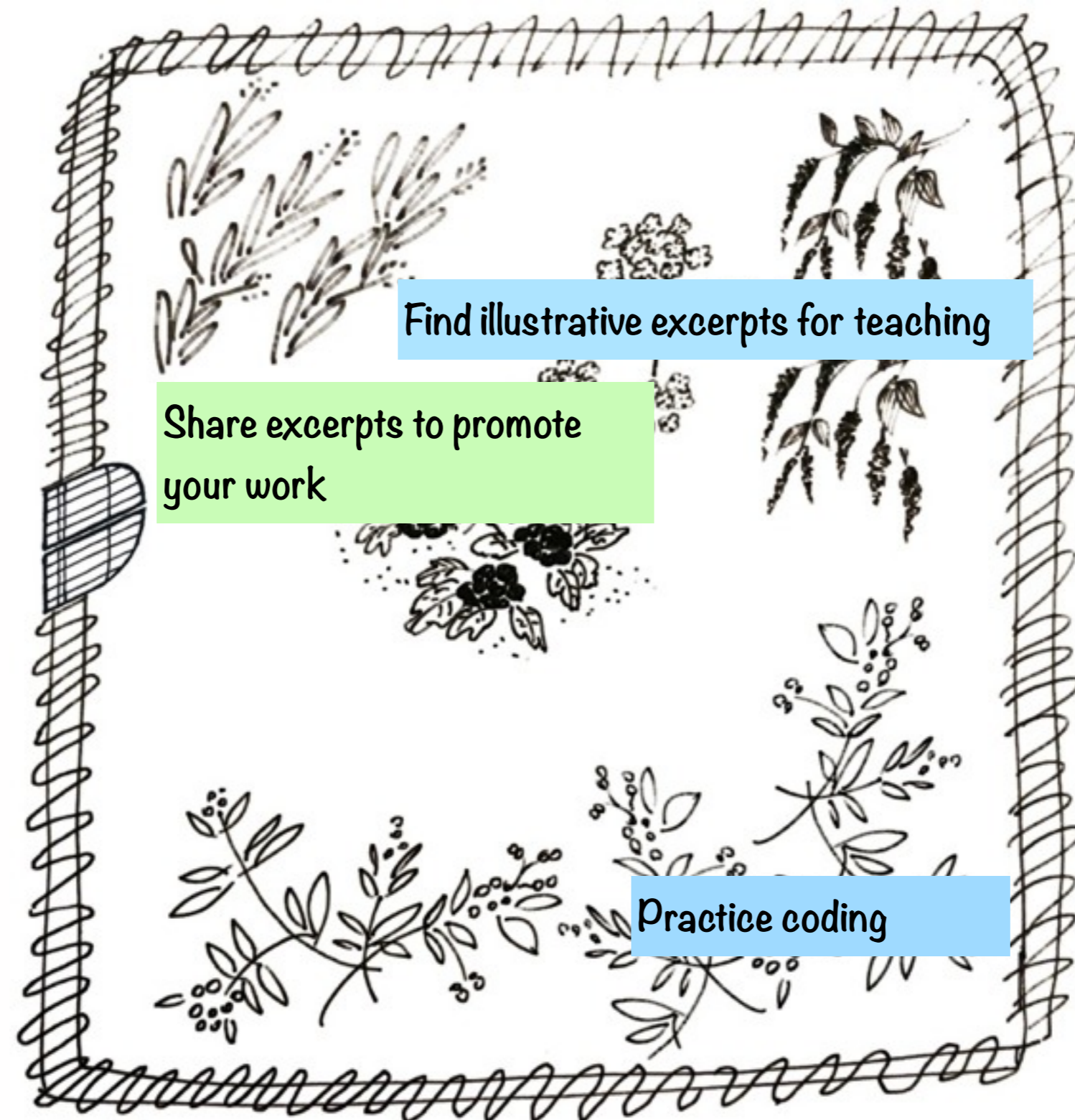
Get inspired

See displays

Learn about coding rules



# In the Databrary garden

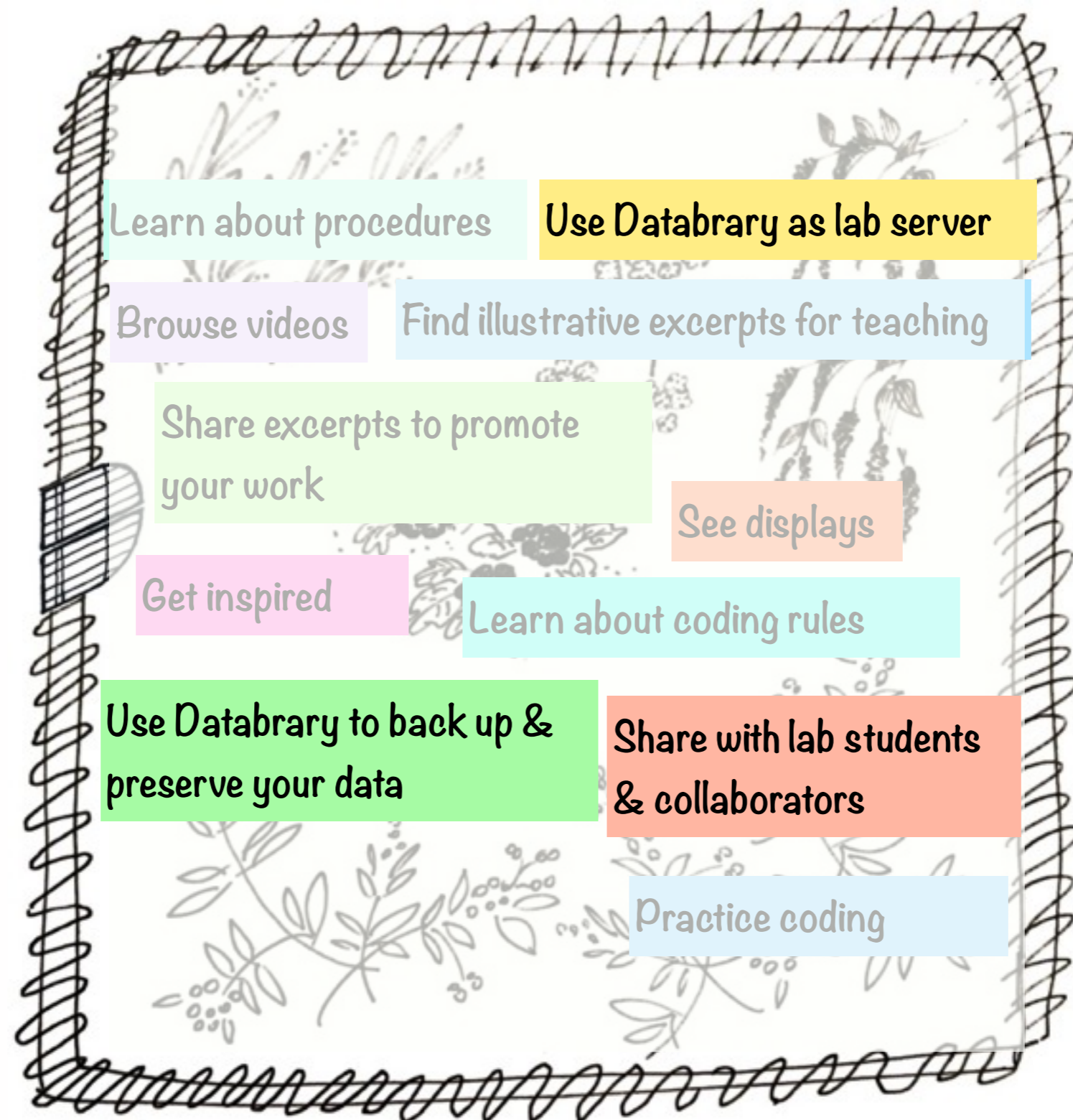


Find illustrative excerpts for teaching

Share excerpts to promote your work

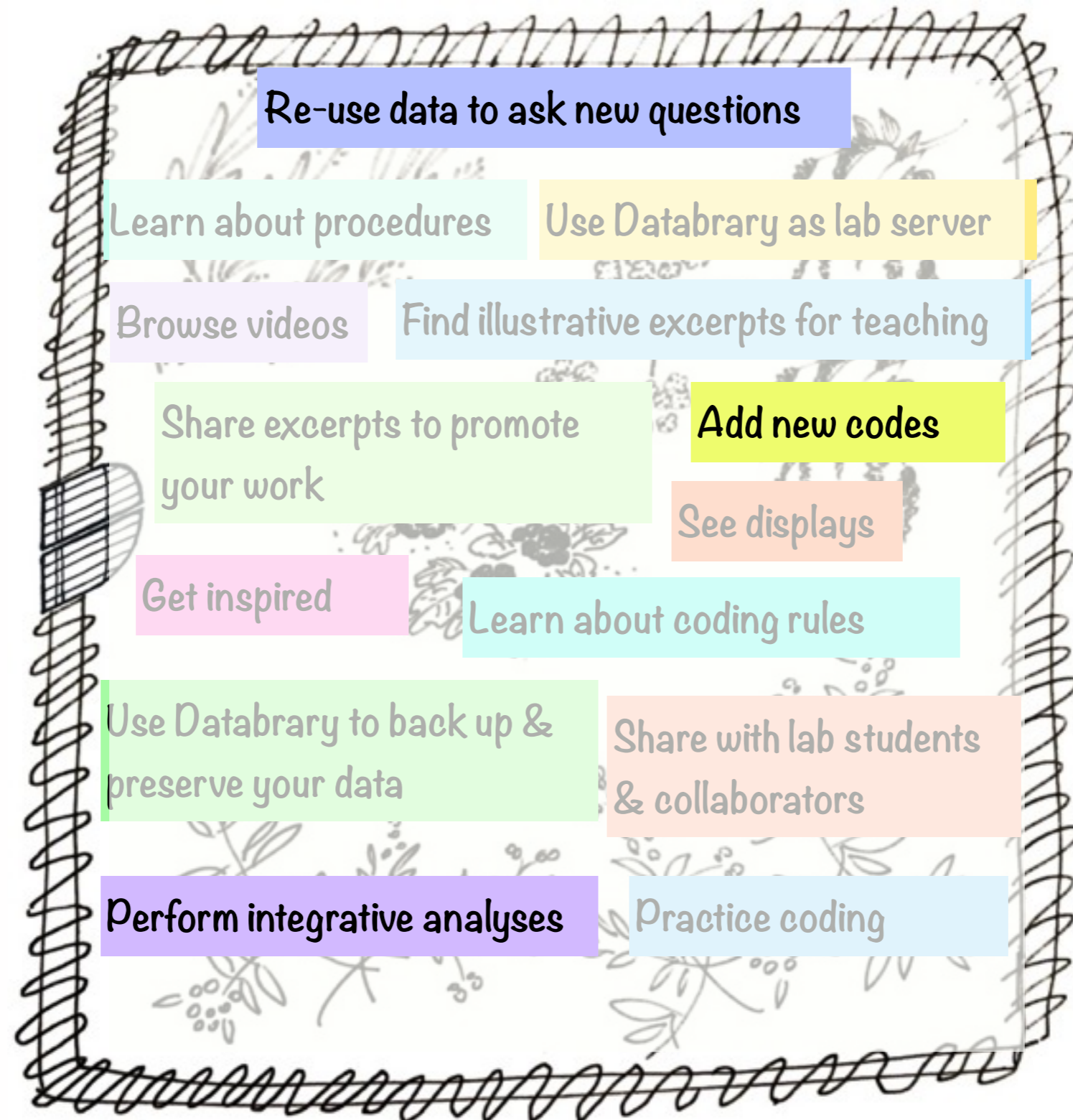
Practice coding

# In the Databrary garden



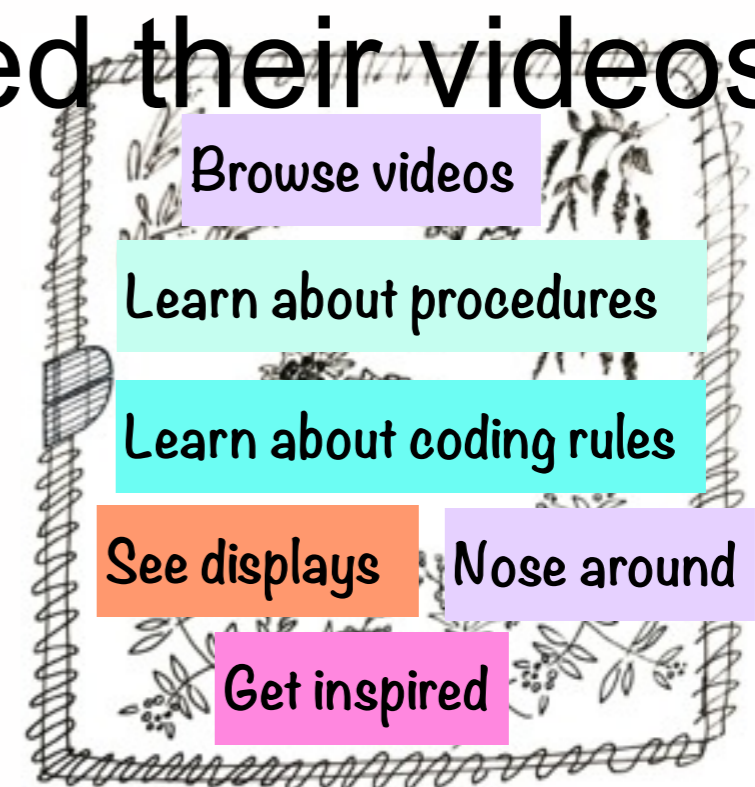


# In the Databrary garden



# Pre-research uses

- ▶ Browse videos, nose around
- ▶ See what colleagues are up to
- ▶ See how to do a procedure
- ▶ Learn how someone coded their videos
- ▶ See displays
- ▶ Get inspired



# 5 Aims

- ▶ Create community of researchers committed to sharing
- ▶ Create policies for sharing
- ▶ Enhance Datavyu video coding tool
- ▶ Provide data management tools
- ▶ Build Databrary repository for sharing & storing video

# Aim 1: Community Building

- ▶ Help at every step
- ▶ Minimize transaction costs
- ▶ Solve pain points
- ▶ Incentives/disincentives matter
- ▶ The customer is always right

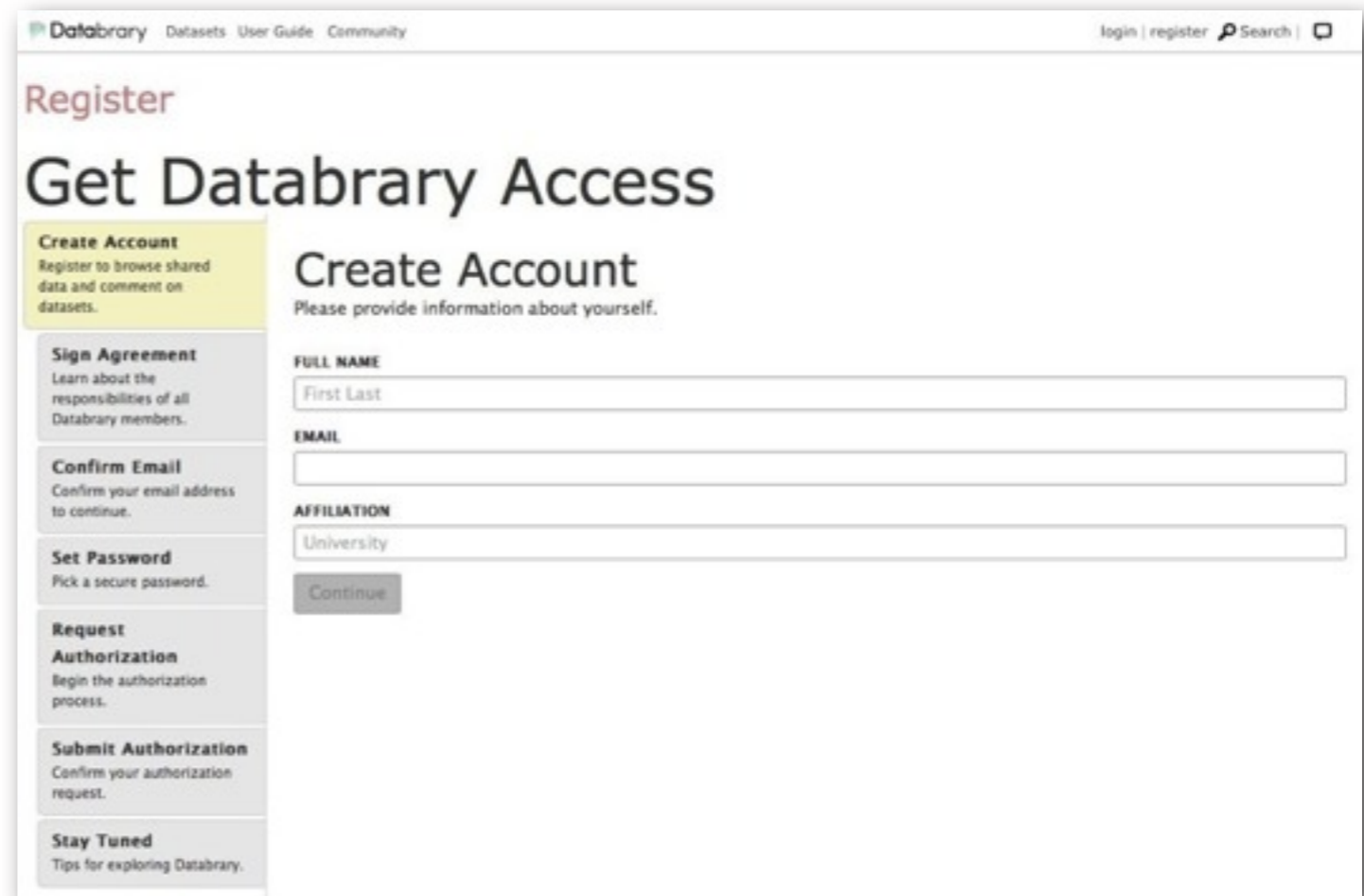
# Aim 1: Community Building

- ▶ *Data* will follow the people
- ▶ Get *people* in first



# Getting people in

- ▶ Registration takes 2 min
- ▶ Staff handles authorization with institutions



The screenshot shows the Databrary registration page. The header includes the Databrary logo and navigation links for Datasets, User Guide, and Community. In the top right corner, there are links for login, register, and a search icon. The main heading is "Register" followed by "Get Databrary Access". On the left side, there is a vertical list of steps: "Create Account" (highlighted in yellow), "Sign Agreement", "Confirm Email", "Set Password", "Request Authorization", "Submit Authorization", and "Stay Tuned". The "Create Account" section is active, showing a form with fields for "FULL NAME" (with a "First Last" placeholder), "EMAIL", and "AFFILIATION" (with a "University" placeholder). A "Continue" button is located below the form.

<http://databrary.org/register>



# Getting people in

- ▶ Promote via professional societies
  - ◀ ISIS 2014 Berlin, booth
  - ◀ SRCDD 2015 Philly, preconference & booth
  - ◀ AERA 2015, booth
  - ◀ CDS 2015 Columbus, preconference & booth
  - ◀ ISIS 2016 New Orleans, preconference & booth

# Getting people in

- ▶ Boilerplate language for data management plans
- ▶ Collaborating on grant submissions
  - ◀ Autism Intervention Trial (U.K.)
  - ◀ IU Cyberlearning
  - ◀ NSF 15-523

# Aim 2: Policies for sharing

- ▶ Consent to share
  - ◀ Builds on informed consent & video/photo release
  - ◀ Separate from consent, **after** the session
  - ◀ Template language on website
- ▶ Share only with authorized researchers
  - ◀ Institutional agreement to permit access, supervise use, co-signed by researcher

# Researchers promise to:

- ▶ Treat Databrary data ethically
- ▶ Respect participants' wishes
- ▶ Authorize people carefully

# Aim 3: Enhance Dataavyu.org

- ▶ Free, open source multi-platform video coding/tagging/annotation tool
- ▶ User guide
- ▶ Help forum
- ▶ Workshops for new users



	babyTalk (MATRIX)	MOMTalk (MATRIX)
00:00:15:470		
	1	00:00:10:831 00:00:11:000 What do you have?
	1 tea	00:00:14:916 00:00:15:220
		2 00:00:16:282 00:00:20:000 It's tea? No I want coffee. Give coffee.
		3 00:00:33:386 00:00:35:000 And a spoon.
00:00:37:695 00:00:37:695	2	00:00:37:695 00:01:28:001 Spoon
	2 here	00:00:38:049 00:00:38:077 00:00:37:697 00:00:38:000 Stir it. Stir it.
	here	00:00:38:077 00:00:38:077 00:00:38:077 00:00:38:077 Click to create observation
00:01:31:611 00:01:31:611		

# A video coding and data visualization tool

Cross-Platform

Fully Extensible

Open-Source



# Aim 4: Data management

- ▶ “After-the-fact”
  - ◀ Conventional route
  - ◀ Sharing awaits completion of study
  - ◀ Requires curation after study is completed or later
- ▶ Numerous disincentives for researchers
  - ◀ High cost in time, effort, \$
  - ◀ Reopening file drawer = high pain

# Aim 4: Data management

- ▶ “Upload-as-you-go”
  - ▶ New route
  - ▶ Self-curation, upload after each session
  - ▶ Sharing while study is in progress or with one button click
- ▶ Only incentives for researchers
  - ▶ Databrary as lab server & backup
  - ▶ Data management built in

# Upload-as-you-go design

- ▶ Familiar spreadsheet for data entry
- ▶ Enter data as collected or from paper
- ▶ Support standard session-level metadata
- ▶ Export metadata for further analysis, post-processing

# Illustration



Language, cognitive, and socio-emotional skills across the first 6 years in U.S. children from African-American, Dominican, Mexican, and Chinese backgrounds

## ▼ DESCRIPTION

This project is a longitudinal study of language, cognitive, social, emotional, and gender development from infants' birth through various critical transitions to schooling (preschool, kindergarten, 1st grade). Children came from diverse backgrounds (African-American, Dominican, Mexican, Chinese) and lived in the New York City metro area. Videos were recorded during various activities and tasks specific to each observation age (e.g., book reading, mother-child free play, novel toy, familiar toy at the 14-month visit) in children's homes or in the laboratory at 14, 24, 36, 52, 64, and 76 months of age.

Shared on October 2013  
Sessions 1253 sessions (1057 shared)  
Ages 9.8 mos - 7.6 yrs ( $M = 3.9$  yrs)

## ▼ HOW TO CITE

Tamis-LeMonda, C. (2013). Language, cognitive, and socio-emotional skills across the first 6 years in U.S. children from African-American, Dominican, Mexican, and Chinese backgrounds. *Databrary*. Retrieved February 7, 2015 from <http://databrary.org/volume/8>.



## ▼ INVESTIGATORS



Catherine Tamis-LeMonda  
New York University

<http://databrary.org/volume/8>

## ▼ MATERIALS

materials	release	group ID	file		
			name	classification	excerpt
Top-level materials		not grouped	no file		
		14 month	3 files		
		24 month	3 files		
		36 month	6 files		
		52 month	5 files		
		64 month	5 files		
		1st grade	5 files		

## ▼ DATA

export data

	session		ID	participant						task		group ID	context	
	test date	release		birthdate	gender	race	ethnicity	language	age	ID	summary		description	setting
	2006-07-07	🔓+	3814	2005-04-12	Male		Hispanic or Latino	English	14.8 mos		5 tasks	14 month	Home	NY
	2006-XX-XX	🔒	3811		Female		Hispanic or Latino	English	13.8 mos		5 tasks	14 month	Home	NY
	2006-07-01	🔓+	3808	2005-04-06	Female		Hispanic or Latino	English	14.8 mos		5 tasks	14 month	Home	NY
	2006-XX-XX	🔒	3806		Male	Black or African American		English	16.4 mos		5 tasks	14 month	Home	NY
	2006-06-05	🔓+	3804	2005-03-15	Female		Hispanic or Latino	English	14.7 mos		5 tasks	14 month	Home	NY
	2006-05-26	🔓+	3803	2005-03-07	Female		Hispanic or Latino	English	14.6 mos		5 tasks	14 month	Home	NY
	2006-04-18	🔓+	3801	2005-02-15	Female		Hispanic or Latino	English	14.0 mos		5 tasks	14 month	Home	NY

<http://databrary.org/volume/8>



# Session timeline

Session (2011-10-03) Download all files as zip

5:26.565

0:00 1:12:01

participant 1

participant 2

context Outdoor, NY

- Monkey bar front
- Monkey bar side
- Tree climbing park side
- Tree climbing park front
- Walkway side

Monkey bar front

FILE FORMAT: MPEG-4 video (mp4)

POSITION: 0:00.000 - 22:08.342 (22:08.342)

<https://nyu.databrary.org/volume/9/slot/6113?segment=-&asset=9607&select=>

# Aim 5: Databrary

The screenshot shows the Databrary website interface. At the top, there are navigation links for 'Datasets', 'User Guide', and 'Community'. On the right, the user 'Rick O. Gilmore' is logged in, with a search icon and a notification icon. The main header features the Databrary logo and a green banner with the text: 'Databrary is an open data library for developmental science. Share video, audio, and related metadata. Discover more, faster.' Below the banner are three white buttons: 'Your Profile' (with a gear icon and subtext 'Manage your data and authorizations.'), 'Browse Data' (with a magnifying glass icon), and 'Browse People' (with a person icon). The main content area is split into two columns. The left column features a 'FEATURED DATASET' section with a video thumbnail of two children on a playground, titled 'Social and motor play on a playground' by Karen Adolph. Below this is a 'RECENT ACTIVITY' section showing a 'New authorized investigator' David Kosson from Rosalind Franklin University of Medicine and Science, dated Feb 5, 2015. The right column contains a 'TOP TAGS' section with a list of tags in rounded rectangular buttons: infants, perception, eye-tracking, numerical cognition, number, apertures, judgments, face perception, language development, affordances, visual attention, falling, cognitive development, online comprehension, and electrophysiology.

Datasets User Guide Community Rick O. Gilmore Search

**Databrary**

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

**Your Profile**  
Manage your data and authorizations.

**Browse Data**

**Browse People**

**FEATURED DATASET**  
Social and motor play on a playground  
Karen Adolph

**RECENT ACTIVITY**  
New authorized investigator Feb 5, 2015  
David Kosson  
Rosalind Franklin University of Medicine and Science

**TOP TAGS**  
infants  
perception  
eye-tracking  
numerical cognition  
number  
apertures  
judgments  
face perception  
language development  
affordances  
visual attention  
falling  
cognitive development  
online comprehension  
electrophysiology

# Databrary.org

- ▶ Beta release early 2014, public release October 2014
- ▶ 71 authorized researchers, 35 contributors
- ▶ 40 institutions from U.S., Canada, South America, and Europe

# Databrary.org

- ▶ 5700 videos; 1600 hrs; 1100 participants
  - ◀ Raw data, excerpts, displays
- ▶ 2,200 other files

# Representative NSF-funded projects

- ▶ Tamis-Lemonda (0721383)
  - ◀ <https://databrary.org/volume/8>
- ▶ LoBue (1247590)
  - ◀ <https://databrary.org/volume/30>
- ▶ Messinger (0808767, 1052736)
  - ◀ <https://databrary.org/volume/15>
- ▶ Karasik (1349044)
  - ◀ <https://databrary.org/volume/11>
- ▶ Barr (1023772)
  - ◀ <https://databrary.org/volume/40>
- ▶ Gilmore (1147440)
  - ◀ <https://databrary.org/volume/23>



# 5 potential *new* avenues

- ▶ Expand target community
- ▶ Create policies for wider video sharing
- ▶ Build web-based Datavyu
- ▶ Enhance Databrary
- ▶ Plan for sustainability



# Avenue 1: Expand community

- ▶ New groups of researchers
  - ◀ Education/teachers, computer science, cognitive/vision science, linguistics, anthropology, ethology/animal behavior
- ▶ Regional workshops
- ▶ Partnerships with professional societies
- ▶ Links to other repositories (imaging, etc)

# Avenue 2: Create policies for wider video sharing

- ▶ Permissions for classrooms & other groups
- ▶ Authorize teachers, clinicians
- ▶ Target valuable existing data
- ▶ Play leadership role in policies for dealing with identifiable data

# Avenue 3: Web-based Datavyu

- ▶ Eliminate video downloads
- ▶ Simplify data management & collaboration
- ▶ Enable frame-by-frame coding/tagging, editing; hyperlinked codebooks
- ▶ Support multiple, synchronized video views
- ▶ Spatial coding

# Avenue 3: Web-based Datavyu

- ▶ Allow group-sourced coding/tagging
- ▶ Reduce hardware/operating system dependencies
- ▶ Support import/export with existing coding tools

# Avenue 4: Enhance Databrary

- ▶ Search, filtering, downloads
- ▶ Search inside videos by time-locked tag
- ▶ View multiple video/data streams
  - ◀ Non-video data streams (physiology, imaging)
- ▶ User metrics
- ▶ Enable collaboration and dialogue

# Avenue 5: Sustainability

- ▶ Pilot and test models
  - ◀ When/how to charge for access
  - ◀ Institutional subscriptions