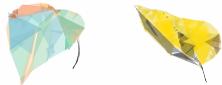


# Databrary Advisory Board Fall Meeting

October 17, 2014  
NYU



## Housekeeping

- ▶ Mark your calendars for next board meeting  
**May 4, 2015 in NYC**
- ▶ Push microphone button to talk
- ▶ Remote advisors: Push space bar to unmute
  - ◀ Jump in anytime!
- ▶ Wifi
  - ◀ Login: nyuguest
  - ◀ Password:

## Agenda

- 1:00-1:15 Welcome
- 1:15-2:30 State of the project, outreach, & policies
- 2:30-2:45 Break
- 2:45-3:45 Databrary 1.0 demo
- 3:45-4:00 Break
- 4:00-4:45 Future directions
- 4:45-5:00 Wrap up & discussion

## Welcome

- ▶ New Databrary staff
  - ◀ Drew Gordon, Information engineer
  - ◀ Gladys Chan, UI designer
  - ◀ Omran Majumder, Quality assurance & curation
- ▶ New board member: Alan Mendelsohn, NYU
- ▶ Invited guests
  - ◀ Diane Brentari, Susan Goldin-Meadow, Lynn Liben, Daphne Maurer, Cybele Raver, Louise Wilkinson, Amanda Woodward

## Many thanks to NYU

- ▶ Nancy Daneau
  - ◀ Director of Office of Sponsored Programs
- ▶ Alison Dewhurst & Marti Dunne
  - ◀ Human Research Compliance Director
  - ◀ Associate Vice Provost for Research Compliance
- ▶ Eric Rasmussen & Mark Righter
  - ◀ Associate General Counsel

## Many thanks

- ▶ Lisa Freund
  - ◀ Branch Chief, Child Development & Behavior, NICHD
- ▶ James Griffin
  - ◀ Deputy Chief, Child Development & Behavior, NICHD
- ▶ Laura Namy
  - ◀ Director, Development & Learning Sciences, NSF

## Why Databrary exists

- ▶ Challenge in developmental psychology
  - ▷ Most developmental researchers collect video data
  - ▷ But don't exploit the richness in video
  - ▷ Good data going to waste
  - ▷ Impedes transparency & slows the pace of discovery
- ▶ Collecting & storing video are costly
- ▶ Databrary allows videos to be reused for new purposes & to address new questions

7

## Strategic focus

- ▶ Video
- ▶ Domain specific repository
- ▶ Developmental science community
  - ▷ Video = primary form of developmental data
  - ▷ Many use cases, including reuse, for video data
  - ▷ Coherent community is big incentive for new user
- ▶ However, tools are broadly applicable beyond video & developmental science community

8

## Transforming the culture of developmental science

Five aims:

- ▶ Enhance Datavyu tool for scoring video ✓
- ▶ Create appropriate policy environment ✓
- ▶ Build Databrary repository for sharing video ✓
- ▶ Provide data management tools
- ▶ Create a community of researchers committed to open data sharing

9

## Fighting the clock

- ▶ NSF \$ now in carry-over
- ▶ NIH \$ ends 3-2018
- ▶ New proposals before funding ends
- ▶ Plan for long-term sustainability
- ▶ Critical to demonstrate community of researchers sharing & using Databrary data

10

## Reminder: 2 ways to acquire data

- ▶ “After-the-fact”
  - ▷ Conventional route for putting data into a repository
  - ▷ Sharing awaits completion of study
  - ▷ Requires curation after study is completed
- ▶ “Upload-as-you-go”
  - ▷ New route for putting data into Databrary
  - ▷ Researcher self-curates and uploads after each data collection session...while study is in progress
  - ▷ Sharing can occur while study is in progress

11

## After-the-fact sharing

- ▶ Incentives
  - ▷ Long-term preservation
  - ▷ Additional citations (to data)
- ▶ Numerous disincentives for researchers
  - ▷ High cost: Time, effort, money
  - ▷ Reopening the file drawer: High pain
- ▶ Disincentives for Databrary are manageable
  - ▷ Hired information scientist Drew Gordon to help with after-the-fact curation

12

## After-the-fact curation in progress

- ▶ Adolph data, including 100+ home videos
- ▶ Tamis-Lemonda NSF 8-year longitudinal
- ▶ Karasik Tajik data, 150+ home videos
- ▶ Maher 20-year longitudinal math learning & Robert B Davis math learning archives
- ▶ Gesell films, many ages, tasks
- ▶ Neiderhiser twins, longitudinal

13

## Excerpts: Easy after-the-fact case

- ▶ Teaching excerpts = most common current use case for developmental videos
  - ◀ Used for teaching developmental courses
- ▶ Incentives
  - ◀ People teach your work
  - ◀ Excerpts linked with appropriate papers
- ▶ Collection of community-contributed real-data teaching materials
  - ◀ Not publisher compiled fake-data video clips

14

## Excerpts for sharing already exist

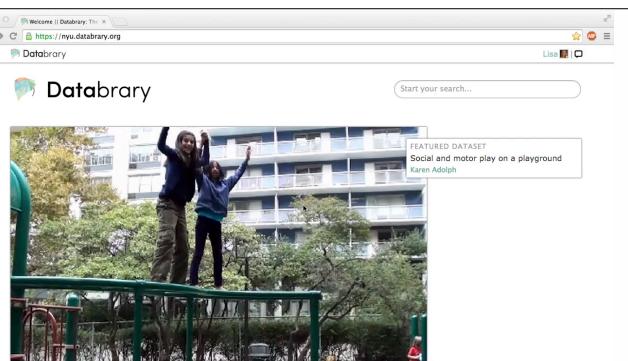
- ▶ Most developmental researchers show videos in colloquia/conferences to illustrate their work
  - ◀ Displays
  - ◀ Children doing the task
- ▶ So, either don't require permission or already have permission
- ▶ Easy for researchers to find the excerpts
  - ◀ Minimal curation needed

15

## Super easy to create excerpt “studies”

- ▶ Disincentives minimal
  - ◀ < 3 minutes to do
- ▶ Add curated, contributor, & user defined tags
  - ◀ Retains idiosyncratic original name for contributor but downloads with standard convention linked to paper

16



17

A screenshot of a researcher profile page on Databrary. The profile belongs to Letitia Naigles. At the top, there is a photo of Letitia Naigles, a woman with short brown hair, smiling. Below the photo, her name "Letitia Naigles" is displayed. There are sections for "CONTACT" (Affiliation: University of Connecticut, Email: letitia.naigles@uconn.edu) and "DATA". The "DATA" section contains a card with the title "Children use syntax to learn verb meanings" by Letitia Naigles. The card includes a brief description: "Verb learning is clearly a function of observation of real-world contingencies; however, it is argued that such observational information is insufficient to account fully for vocabulary acquisition. This paper provides..." Below the "DATA" section is a "AUTHORIZATIONS" section which says "Authorized by University of Connecticut".

18

## Upload-as-you-go sharing

- ▶ Incentives for researchers are considerable
  - ◀ Long-term preservation, short-term back-up
  - ◀ Lab server to enable sharing among students & collaborators
  - ◀ Data organization & management
  - ◀ Transcoding, syncing, & splitting videos
  - ◀ More data management features to be added
- ▶ Disincentives are minimal or nonexistent
  - ◀ Potentially new practices introduced into workflow

19

## Upload-as-you-go sharing

- ▶ Recent & current focus of development effort
- ▶ Working system ready for release!
  - ◀ Demo later today
  - ◀ Release timeline
  - ◀ Beta testing, feedback plan
  - ◀ Additional features

20

## Databrary ~ “walled garden”

- ▶ Shared data inside
- ▶ Researchers require authorization to access shared data
  - ◀ Institution will vouch for you
  - ◀ “PI” status
  - ◀ IRB, CITI, ethical training



21

## In the Databrary garden



22

## In the Databrary garden



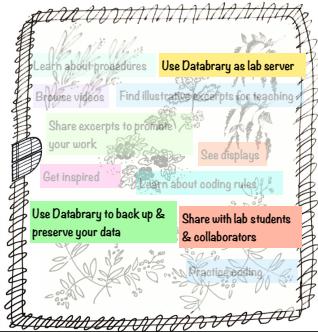
23

## In the Databrary garden



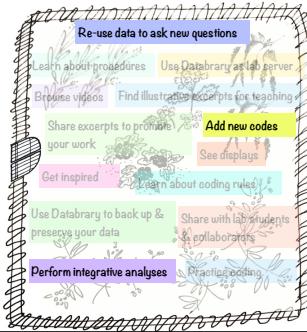
24

## In the Databrary garden



25

## In the Databrary garden



26

## In the Databrary garden



27

## Why restrict access?

- ▶ Developmental researchers more comfortable
  - ◀ Sharing video data
  - ◀ Asking participants to share video data
  - ...if data are shared with other researchers like themselves
- ▶ Not hand-picking users, but want to assure contributors that users are legitimate PIs at a legitimate institution
  - ◀ Extends the “zone of trust”

28

Questions?

29

## Since last board meeting

- ▶ Released the Databrary beta
- ▶ About to release Databrary 1.0
- ▶ Filling the garden with data
- ▶ Authorizing researchers
- ▶ And...we've learned a lot

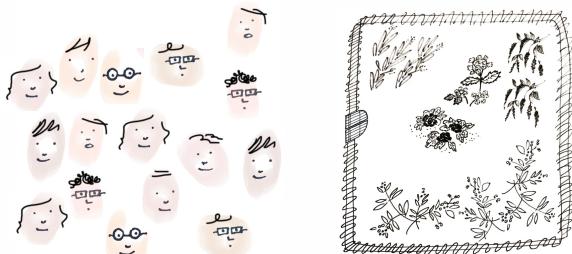
30

## Lessons learned: Guiding philosophy

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

31

## Lessons learned: Community building



32

## Lessons learned: Community building

- ▶ We used to worry about how to get *data* in

33

## Lessons learned: Community building

- ▶ We used to worry about how to get *data* in
- ▶ Then we realized, the first step is to get *researchers* in
  - ◀ Only then can we help them to take the next step



34

## Getting them in the door

- ▶ Made registration an easy 2-minute process
- ▶ We handle the authorization burden
- ▶ Adapted our outreach approach to emphasize registration first, other steps next

35

## Community building: Personal outreach

- ▶ Reaching out to researchers one by one
  - ◀ Personal email requests from Karen & Rick
  - ◀ Personal follow-up emails from Lisa
  - ◀ Staff visits to labs/universities
  - ◀ Private training workshops at NYU
  - ◀ Targeting researchers at authorized schools

36

## Pitching Databrary at invited colloquia

- ▶ Colloquia
  - ▷ ICIS 2014, Berlin
  - ▷ Max Planck Institute for Human Development, Berlin
  - ▷ Center for Brain & Cognitive Development, Birkbeck
  - ▷ Children's Hospital of Philadelphia
- ▶ Would love it if advisory board did likewise
  - ▷ We will give you template slides to use

37

## Questions?

38

## Transforming the culture of developmental science

Five aims:

- ▶ Enhance Datavyu tool for scoring video ✓
- ▶ Create appropriate policy environment ✓
- ▶ Build Databrary repository for sharing video ✓
- ▶ Provide data management tools
- ▶ Create a community of researchers committed to open data sharing

39

## Presence in...

- ▶ Policies for data access and sharing space
  - ▷ Digital Learning Data as a Public Good workshop
- ▶ Information science space
  - ▷ NFAIS webinar



40

## Presence in...

- ▶ Open science, research transparency space
  - ▷ Data on the Mind Research Transparency Forum, Berkeley Initiative for Transparency in the Social Sciences
  - ▷ Submitted paper for BITSS Annual Meeting, December 2014



41

## Social media presence

- ▶ Campaigns to increase visibility
  - ▷ Facebook
  - ▷ Twitter
  - ▷ Blog posts



42

## Community building: Societies

- ▶ Programmatic shift to promote Databrary in prominent developmental science venues
  - ▷ Exhibitor booths at key conferences
  - ▷ Preconference workshops
  - ▷ Regional workshops
  - ▷ Society list-serves
- ▶ Get people registered & using the site
  - ▷ Face to face interaction facilitates adoption

43

## Society outreach plans

- ▶ ICIS July 2014
- ▶ SRCD March 2015
  - ▷ Preconference workshop, invited address
- ▶ AERA April 2015
- ▶ CDS October 2015
- ▶ ICIS May 2016
  - ▷ Preconference workshop, presidential address
- ▶ Regional workshops Fall 2015, Spring 2016

44

## Supporting our existing community

- ▶ Lab-wide release for staff recorded in videos
- ▶ Boilerplate language for grants:
  - ▷ Data Management Plan
  - ▷ Resource Sharing Plan
- ▶ Letters of support for grants

45

## Stronger ties to education community

- ▶ Focusing on reaching out to educational community
  - ▷ Robert B. Davis archival math data & Rutgers-Kenilworth longitudinal study (many thanks to Carolyn Maher)
- ▶ Learn about unique needs & new use cases of education research community
- ▶ Advice?

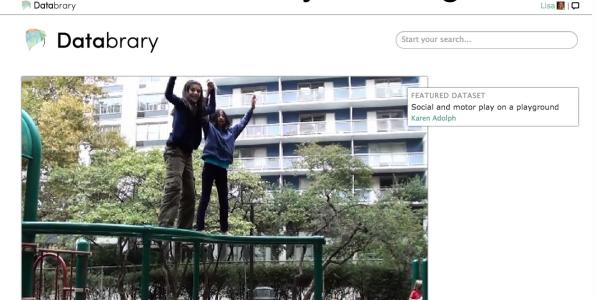
46

## Supporting & serving sub-domains

- ▶ Learning about & serving special needs of sub-domains
  - ▷ Gesture: Susan Goldin-Meadow, Center for Gesture, Sign & Language, University of Chicago
  - ▷ Autism, Clinical: Tony Charman, Autism intervention clinical trial, Kings College London
  - ▷ Temperament and emotion: Kristin Buss & Nathan Fox, PSU, University of Maryland

47

## Community building



48



49

## Contributions

- ▶ Labs requesting release: 13
- ▶ Teaching excerpt studies: 36
  - ▷ From 23 people, 5 board members
- ▶ Full datasets: 8
  - ▷ New: Child Affective Facial Expressions (CAFE)
  - ▷ All are “after-the-fact” archival datasets
  - ▷ Others in the queue

50

## Recapping our policy framework

- ▶ Special problem with video: *Identifiable* data
- ▶ Requires policies that enable open sharing
  - ▷ Informed consent for sharing identifiable data
  - ▷ “Wall” of the garden, access restrictions that protect sensitive shared data

51

## Access model

- ▶ Unique from other data sharing agreements
- ▶ Authorization grants researchers user **and** contributor privileges to the entire library
  - ▷ Including for non-research purposes
- ▶ Agreement requires institutional sign-off
  - ▷ Databrary took on burden of working with the institution
  - ▷ Easy for any additional researchers at the institution to get authorized

52

## Authorization progress

- ▶ Authorizing institutions: 26
  - ▷ In the queue: 40
- ▶ Authorized investigators: 51
- ▶ Registered, authorization pending: 53

53

## Our approach was working... but

- ▶ Institutions weren't always clear about the process or structure of agreement
  - ▷ Are investigators entering into agreement themselves? No
  - ▷ Who co-signs the agreement? Only the authorizing official
  - ▷ Does the IRB need to review or approve agreement? No
  - ▷ Is IRB approval required for authorization & access? No

54

## Clarifying questions about agreement

- ▶ Questions created burden for researchers
- ▶ Became a pain point to solve
- ▶ Started by clarifying requirements, responsibilities, and roles

55

## Refining our agreement

- ▶ Restructuring agreement, *not* access model
- ▶ Pared down “Access Agreement”
- ▶ Will point to the “Access” page on our website
  - ◀ Stipulates expectations of researchers & institutions *by role*
  - ◀ More plain English, less legalese, fewer redundancies
  - ◀ ‘Appendix’ for institutions to add additional authorized investigators
- ▶ Emphasizes three key ethical principles

56

## Three key ethical principles

1. Ensure that participants' wishes about sharing their data are respected
2. Treat Databrary data with the same high standard of care that you treat data collected in your own laboratory
3. Take care in granting & managing access for affiliates & take responsibility for their use of Databrary

57

## Access

The screenshot shows the Databrary Access page. At the top, there's a logo and a link to 'enter databrary beta'. Below that is a navigation bar with 'About', 'Access' (which is highlighted in green), and 'Community'. The main content area has a section titled 'ACCESS' with a sub-section for 'Rights and Responsibilities'. It includes a note about researchers and their institutions, a link to the 'User Guide', and a link to the staging site: [staging.databrary.org/resources](https://staging.databrary.org/resources). There are also links for 'Affiliates', 'Rights Granted to Databrary', and 'User Guide'. A note at the bottom says 'Words in *italics* are defined in our glossary.'

58

Databrary's three guiding ethical principles embody the commitment that is expected of all community members to build a community that follows the highest ethical standards.

As a member of the Databrary community you promise to:

1. Treat Databrary data with the same high standard of care that you treat data collected in your own laboratory.
2. ensure that participants' wishes about sharing their data are respected.
3. authorized investigators must take care in granting and managing access for Affiliates and take responsibility for their use of Databrary.

### The Databrary Approach

Databrary strives to enable widespread video data sharing while protecting confidentiality and privacy. The Databrary approach consists of several components:

#### Asking Permission

If a video or other research data contain identifiable information, researchers must ask participants (or the participant's parent or guardian) for their permission to share this information with other researchers. Only if the participant agrees may videos or identifiable information be shared with other researchers. Databrary provides a template sharing release, similar to a photo release, that researchers may adapt for use in their own laboratories. Participants may choose to not to share at all, share only with other researchers, or allow a short excerpt or clip to be shared with the public for educational or scientific purposes, like a classroom lecture or research presentation.

- Collecting Shareable Data
  - Adding the Databrary Release to Your IRB Protocol
  - Obtaining Participant Permissions
  - Sample Participant Release Forms
  - Video Example of Obtaining Databrary Release
  - Data Release Levels
  - Grandfathering Already Collected Data
  - Preparing to Contribute Data
    - Organizing Your Data
    - Examples of Suggested Fields
    - Sharing Data
    - Boilerplate Language for Funding Proposals
- Policies
  - Participant Release Template
  - Databrary Access Agreement
  - Data Management Plan Template
  - Best Practices for Data Security
  - Data Release Levels
  - Data Sharing Manifesto
  - Glossary

59

## Questions?

60

## Agenda

1:00-1:15 Welcome  
1:15-2:30 State of the project, outreach, & policies  
**2:30-2:45 Break**  
2:45-3:45 Databrary 1.0 demo  
3:45-4:00 Break  
4:00-4:45 Future directions  
4:45-5:00 Wrap up & discussion

61

## Agenda

1:00-1:15 Welcome  
1:15-2:30 State of the project, outreach, & policies  
2:30-2:45 Break  
**2:45-3:45 Databrary 1.0 demo**  
3:45-4:00 Break  
4:00-4:45 Future directions  
4:45-5:00 Wrap up & discussion

62

## Upload-as-you-go goals

- ▶ Enter data once, directly into Databrary
- ▶ Efficient, familiar interfaces
- ▶ Standard session-level metadata
- ▶ Support existing workflows
- ▶ Continuous feedback and improvement

63

## 1.0 preview

64

## Session-level metadata

- ▶ Participants: birthdate, gender, race, ethnicity, language, disability...
- ▶ Experimental conditions
- ▶ Groups (longitudinal visit, ability, etc.)
- ▶ Task (activities, parts, experiments)
- ▶ Context: setting, location, language
- ▶ Pilot indicator
- ▶ Exclusion: reason

65

## Incorporating feedback

- ▶ Iterative enhancements, improve upload-as-you-go
- ▶ Support current research workflows
- ▶ Community features: topical excerpt collections, tagging, commenting

66

## Goals for the next release

- ▶ Import existing codes, tags, annotations
- ▶ Data reuse: studies, views
- ▶ Search: improve discoverability, sorting, & filtering of data

67

## Future possibilities

- ▶ Online video coding
- ▶ Audio visualization
- ▶ Automate transcription & frame tagging

68

## Questions?

69

## Agenda

- 1:00-1:15 Welcome  
1:15-2:30 State of the project, outreach, & policies  
2:30-2:45 Break  
2:45-3:45 Databrary 1.0 demo  
3:45-4:00 Break  
4:00-4:45 Future directions  
4:45-5:00 Wrap up & discussion

70

## Agenda

- 1:00-1:15 Welcome  
1:15-2:30 State of the project, outreach, & policies  
2:30-2:45 Break  
2:45-3:45 Databrary 1.0 demo  
3:45-4:00 Break  
4:00-4:45 Future directions  
4:45-5:00 Wrap up & discussion

71

## The pathway forward

- ▶ Don't reinvent wheels
- ▶ Solve researchers' problems
- ▶ Build on existing practices, but improve them



72

## The path forward

- ▶ Establish partnerships
- ▶ Focus on developmental audience
- ▶ Build flexible tools applicable to many
- ▶ Attract funders



73

## Five opportunities

1. Automate tagging, annotation
2. Automate data ingest, upload
3. Enhance mining of Databrary's "nuggets"
4. Support other data streams, formats
5. Build strategic partnerships

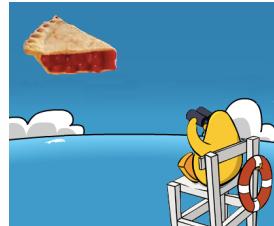
74

### 1. Automate tagging, annotation

- ▶ Video data are rich
  - ◀ But, expensive to tag, annotate
  - ◀ Personnel, time, etc.
  - ◀ As quantity grows, exceeds human capabilities to code
- ▶ Make video more valuable for searching, browsing

75

### So is this just...



76

### Opportunities for automatic tagging

- ▶ Automatic speech transcription
- ▶ Automatic face detection
- ▶ Automatic object recognition
- ▶ Detect physiology from video

77

### Challenges

- ▶ Identify useful software services
- ▶ Connect them to Databrary
- ▶ Maintain security

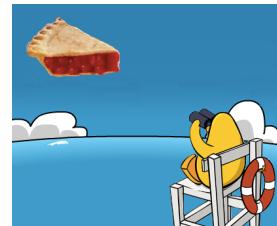
78

## 2. Automate data ingest, upload

- ▶ Connect data streams directly to Databrary
- ▶ Avoid manual upload of videos & data entry of metadata

79

So is this just...



80

## Opportunities for automatic ingest

- ▶ Web cams
- ▶ Wearable cams
- ▶ Smart phone video
- ▶ Wearable devices
- ▶ Smart phone apps
- ▶ Tablets, touch screens
- ▶ Console video games



81

## Challenges

- ▶ Which devices are a priority?
- ▶ Connect to Databrary
- ▶ Maintain security

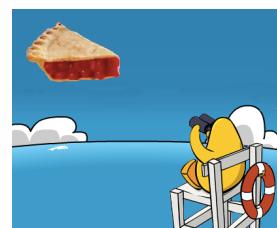
82

## 3. Enhance mining of Databrary's "nuggets"

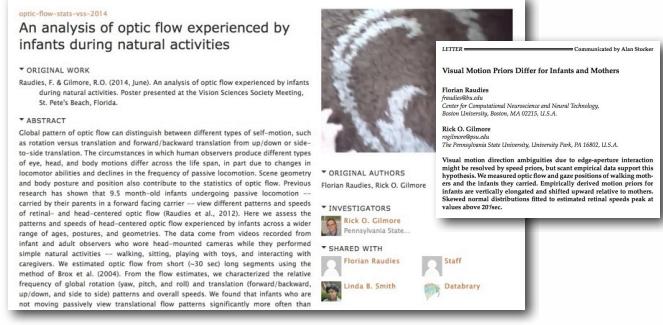
- ▶ Find new value for existing data
- ▶ Connections between, across datasets
- ▶ Access from stats, data tools over the web
- ▶ Data mining (Databrary & published lit)
- ▶ Access Databrary from R, Matlab, SPSS
- ▶ Visualization
- ▶ Machine learning

83

So is this just...



84



85

## Experts in this area on our board



86

## Challenges

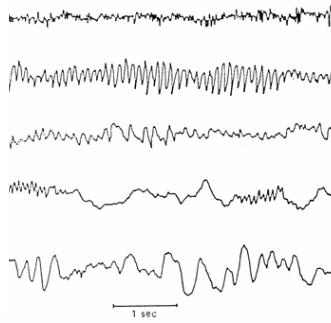
- ▶ Which links are the highest priority?
- ▶ Feasibility
- ▶ Partners with mutual interests, compatible time tables

87

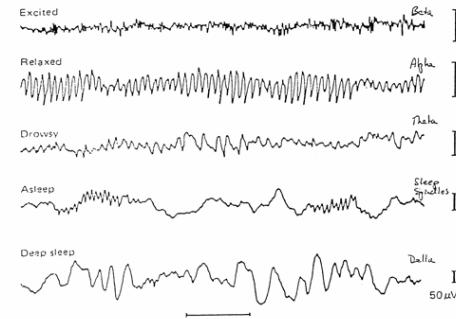
## 4. Support other data streams, formats

- ▶ Developmental scientists collect
  - ◀ EEG/ERP, NIRS, or other physiological data
  - ◀ Desk-top eye tracking
- ▶ EEG, eyetracking
  - ◀ Temporally dense like video
  - ◀ Collected with video
- ▶ Databrary can store non-video files
  - ◀ But, no preview, conversion, visualization, indexing

88



89



90

## Challenge: Data provenance

- ▶ Metadata make streams interpretable & useable
  - ▷ Timelocked displays/video or audio
- ▶ Sharing within labs & among collaborators

91



92

## 5. Build strategic partnerships

- ▶ Interoperability with existing tools
- ▶ Interoperability with other services/archives
- ▶ Connections with researchers via societies and journals

93

## Interoperability with existing tools

- ▶ Datavyu -> CHILDES CLAN ✓
- ▶ CHILDES CLAN -> Datavyu (in prep)



94

## Interoperability with existing tools

- ▶ Popular video coding, annotation tools
  - ▷ Elan, Transana
  - ▷ Noldus Interact, Mangold Observer
- ▶ Specialized file formats, etc.

95

## Interoperability

- ▶ ICPSR
- ▶ Center for Open Science (COS)/Open Science Framework (OSF)
- ▶ Dataverse
- ▶ Research Data Alliance (RDA)

96

## Connections with societies & journals

- ▶ Store (unlimited) supplemental materials
- ▶ Store materials from conferences, workshops (talks, posters)
- ▶ Identifiability issues reduced or absent

97

## Connections with societies & journals

- ▶ Developmental societies
  - ◀ Society for Research in Child Development (SRCD)
  - ◀ International Society on Infant Studies (ISIS)
  - ◀ Cognitive Development Society (CDS)
- ▶ Other societies
  - ◀ Vision Sciences Society
  - ◀ American Speech-Language-Hearing Association (ASHA)
  - ◀ Clinicians

98

## Challenges

- ▶ Courting takes time
- ▶ Mutual interest essential

99



Cake at the Like

100

## Your thoughts?



101

## Your thoughts?

- ▶ Other directions?
- ▶ Are these exciting, useful, fundable?
- ▶ Other partners?
- ▶ Priorities?

102

## Developmental science can lead the way

- ▶ Origin of complex behavior a core scientific question
- ▶ Development never stops
- ▶ Learning, plasticity, aging, intervention, therapy, recovery == development
- ▶ Developmental science marginalized



103



104

## Developmental science can lead the way

- ▶ Databrary infrastructure & tools are broadly applicable to all of behavioral science
- ▶ On the vanguard of big data & data science
  - ▷ Video media
  - ▷ Identifiable data

105



106



About Resources Community



Databrary is an [open data library](#) for developmental science. Share video, audio, and related metadata. Discover more, faster. >>>

107

## Agenda

- 1:00-1:15 Welcome
- 1:15-2:30 State of the project, outreach, & policies
- 2:30-2:45 Break
- 2:45-3:45 Databrary 1.0 demo
- 3:45-4:00 Break
- 4:00-4:45 Future directions
- 4:45-5:00 Wrap up & discussion

108