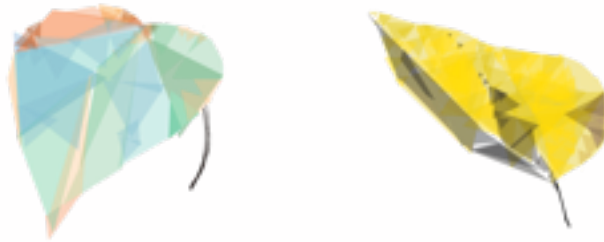


Databrary Advisory Board Spring Meeting

April 7, 2014
NYU



Meeting Agenda

- 1:00-1:15 Welcome and recent accomplishments
- 1:15-1:45 Roadmap
- 1:45-2:15 Policies
- 2:15-2:30 Break
- 2:30-3:30 Beta: Upload and search
- 3:30-4:00 Researcher asks and support
- 4:00-4:10 Break
- 4:10-5:00 Positioning Databrary for the future
- 5:00-5:30 Wrap up and discussion



Announcements

- ▶ Next Board Meeting: October 13, 2014 at NYU
- ▶ Wi-fi access: nyuguest
 - ◀ Guest ID: guest132
 - ◀ Password: keddlith
- ▶ Meeting will be recorded
- ▶ Push-to-talk microphones
- ▶ Reimbursement form and receipts to Lina Wictoren Roy
- ▶ Dinner following at Hundred Acres
 - ◀ 38 Macdougall Street, corner of Prince



Databrary Pls



Karen Adolph
NYU



Rick Gilmore
Penn State



David Millman
NYU

Agency partners

- ▶ Jim Griffin, NICHD, Science Officer
- ▶ Lisa Freund, NICHD, Program Officer
- ▶ Laura Namy, NSF, Director DLS



Special Guests

▶ Lynn Liben

- ◀ Distinguished Professor of Psychology, Penn State,
President of the Society for Research in Child Development

▶ Daphne Maurer

- ◀ Distinguished University Professor of Psychology, Neuroscience
& Behaviour, McMaster University,
President of the International Society on Infant Studies

▶ Amanda Woodward

- ◀ William S. Gray Professor of Psychology, University of Chicago,
President of the Cognitive Development Society



Special thanks to NYU

- ▶ David Ackerman: IT support, storage, hiring
- ▶ Nancy Daneau: Authorization
- ▶ Alison Dewhurst and Marti Dunne:
Participant release, IRB
- ▶ Pamela Morris: IHDSC, administrative support
- ▶ Eric Rasmussen and Mark Righter:
Authorization, university agreements, legal



Very special thanks

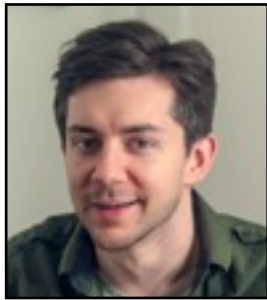
▶ Richard Louth



Databrary team



Dylan Simon



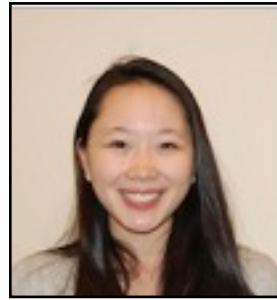
Mike Continues



Jon Coe



Jesse Lingeman



Vicky Foo



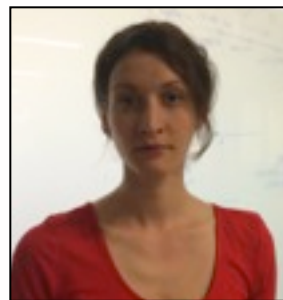
John Franchak



Lisa Steiger



Andrea Byrne



Lina Wictoren Roy

Board member updates

- ▶ Welcome to new advisors
 - ◀ Daniel Messinger, University of Miami, analytic tools
 - ◀ Brian Nosek, University of Virginia, COS
 - ◀ Tal Yarkoni, University of Texas, neurosynth.org
- ▶ Farewell to departing member
 - ◀ Sarah Morrow, PSU



Publications

- ▶ Submitted
 - ◀ International Journal of Digital Libraries
- ▶ In progress
 - ◀ WIRE's Cognitive Science (invited paper)
 - ◀ Best practices in behavioral coding
 - ◀ Call to arms
 - ◀ Top 10 reasons to share
 - ◀ Top 10 concerns about sharing



Presentations

- ▶ NFAIS/CENDI/Library of Congress (Nov. 12, 2013)
- ▶ CNI 2013 Fall Meeting (Dec. 9-10, 2013)
- ▶ Spatial DataMine Workshop (Feb. 7-9 2014)
- ▶ 12+ presentations to psychologists, pediatric physical/occupational therapists, movement scientists



Datavyu

- ▶ Released stable version of Datavyu 1.1
- ▶ Published user guide
- ▶ Best practices in behavioral coding forthcoming



Seeding the repository: Datasets acquired

- ▶ Adolph
- ▶ Tamis-LeMonda
- ▶ Karasik
- ▶ Gordon
- ▶ LoBue



Forthcoming datasets

- ▶ Messinger lab
- ▶ Gernsbacher lab
- ▶ Maher longitudinal study
- ▶ CHILDES videos
- ▶ Gesell archives
 - ◀ In partnership with MIAP, Howard Besser



Meeting Agenda

- 1:00-1:15 Welcome and recent accomplishments
- 1:15-1:45 Roadmap**
- 1:45-2:15 Policies
- 2:15-2:30 Break
- 2:30-3:30 Beta: Upload and search
- 3:30-4:00 Researcher asks and support
- 4:00-4:10 Break
- 4:10-5:00 Positioning Databrary for the future
- 5:00-5:30 Wrap up and discussion



Why Databrary exists



Why Databrary exists

- ▶ Challenge in developmental psychology
 - ◀ Most developmental researchers collect video data
 - ◀ Not exploiting the richness in video
 - ◀ Good data going to waste
 - ◀ Impeding transparency and the pace of discovery



Why Databrary exists

- ▶ Challenge in developmental psychology
 - ◀ Most developmental researchers collect video data
 - ◀ Not exploiting the richness in video
 - ◀ Good data going to waste
 - ◀ Impeding transparency and the pace of discovery
- ▶ Collecting and storing video is costly



Why Databrary exists

- ▶ Challenge in developmental psychology
 - ◀ Most developmental researchers collect video data
 - ◀ Not exploiting the richness in video
 - ◀ Good data going to waste
 - ◀ Impeding transparency and the pace of discovery
- ▶ Collecting and storing video is costly
- ▶ Open up videos to be repurposed for new questions



Key aims of the Databrary project

- ▶ Build Databrary repository for sharing video
- ▶ Create appropriate policy environment
- ▶ Provide data management tools
- ▶ Enhance Datavyu tool for scoring video
- ▶ Transform the culture of developmental science!



Roadmap



Videos to
Share



View & Use
Videos

Roadmap



Upload



Videos to
Share



View & Use
Videos

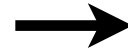
Roadmap



Upload



Videos to
Share



View & Use
Videos

- ▶ Data dump
 - ◀ Datasets are browsable but not searchable at session level

Roadmap



Upload



Videos to
Share



View & Use
Videos

- ▶ Data dump
 - ◀ Datasets are browsable but not searchable at session level
- ▶ Structured datasets and studies
 - ◀ Assign appropriate metadata to each session

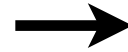
Roadmap



Upload



Videos to
Share



View & Use
Videos

- ▶ Videos are identifiable
 - ◀ So each video (session) must have a release level
 - ◀ Dataset dump not possible because metadata tags for each session are required

Roadmap

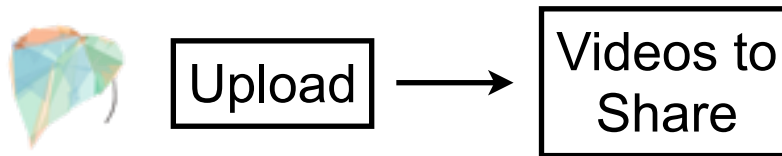


Upload

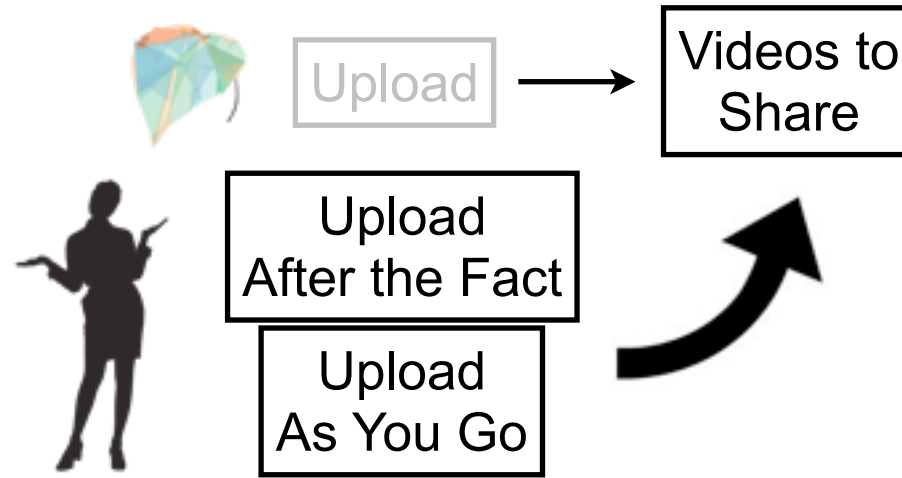


Videos to
Share

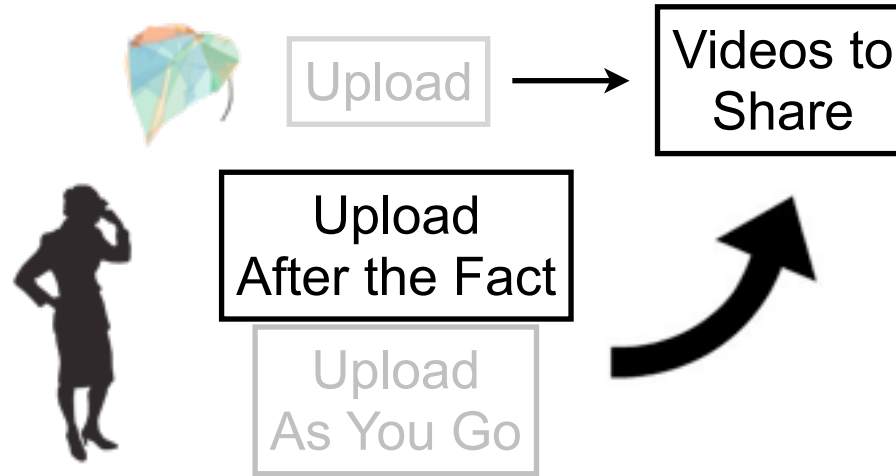
Roadmap



Roadmap



Roadmap



Incentives: Preservation; data easily available to collaborators; mensch...?
Disincentives: Time; effort; \$.



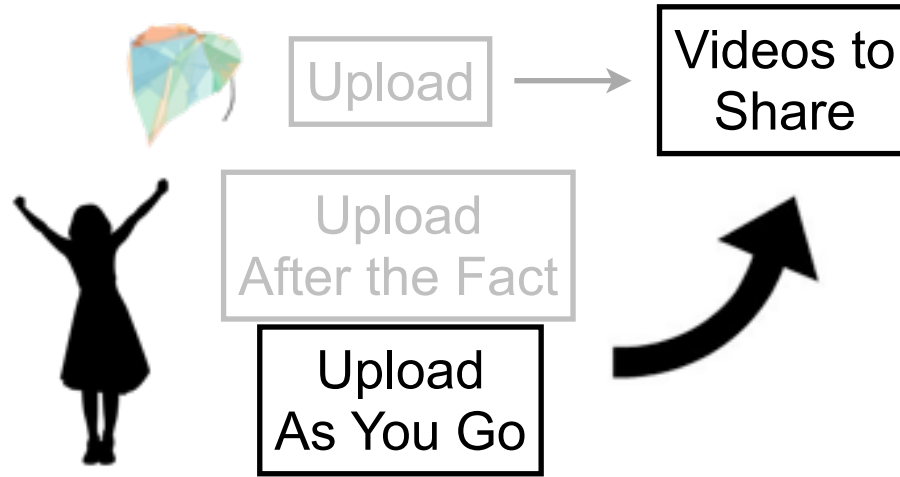
Upload
After the Fact

Upload
As You Go

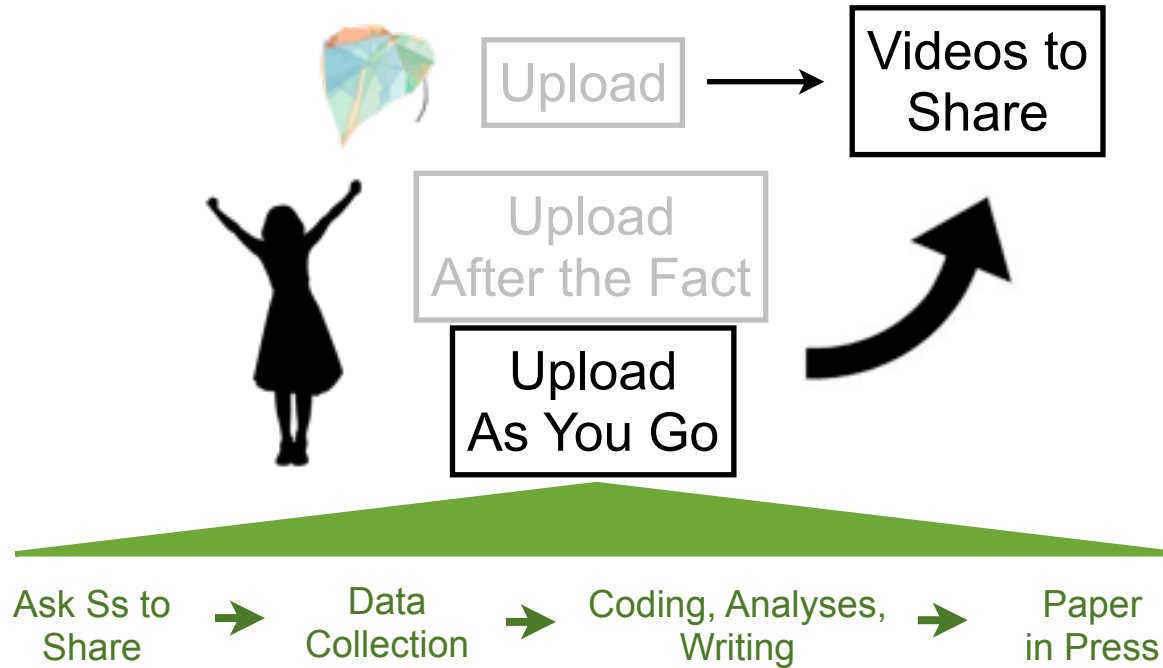


Share

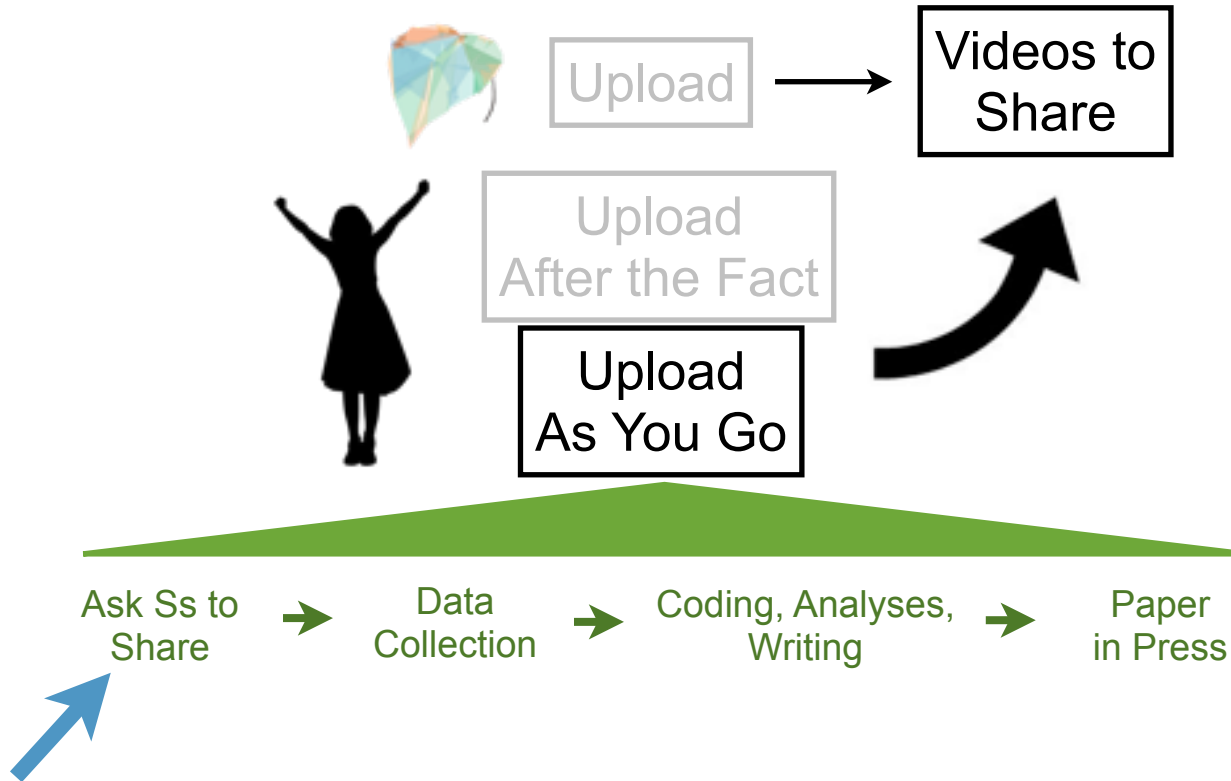
Roadmap



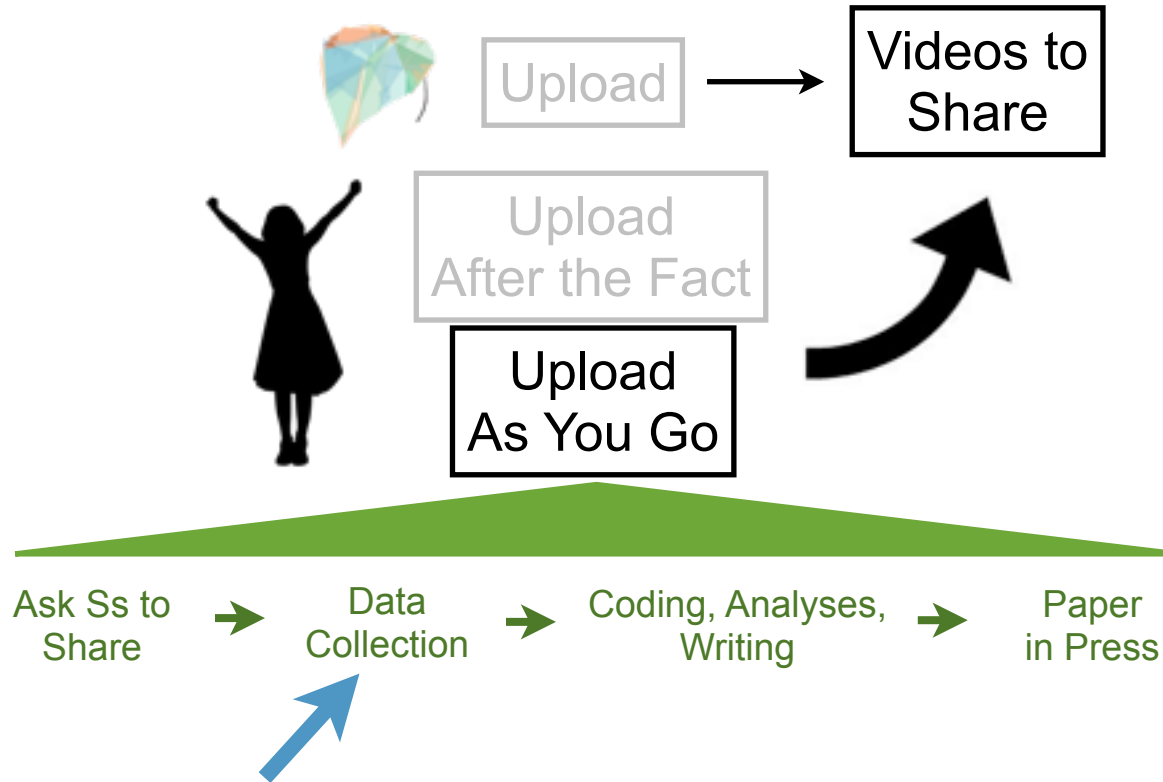
Roadmap



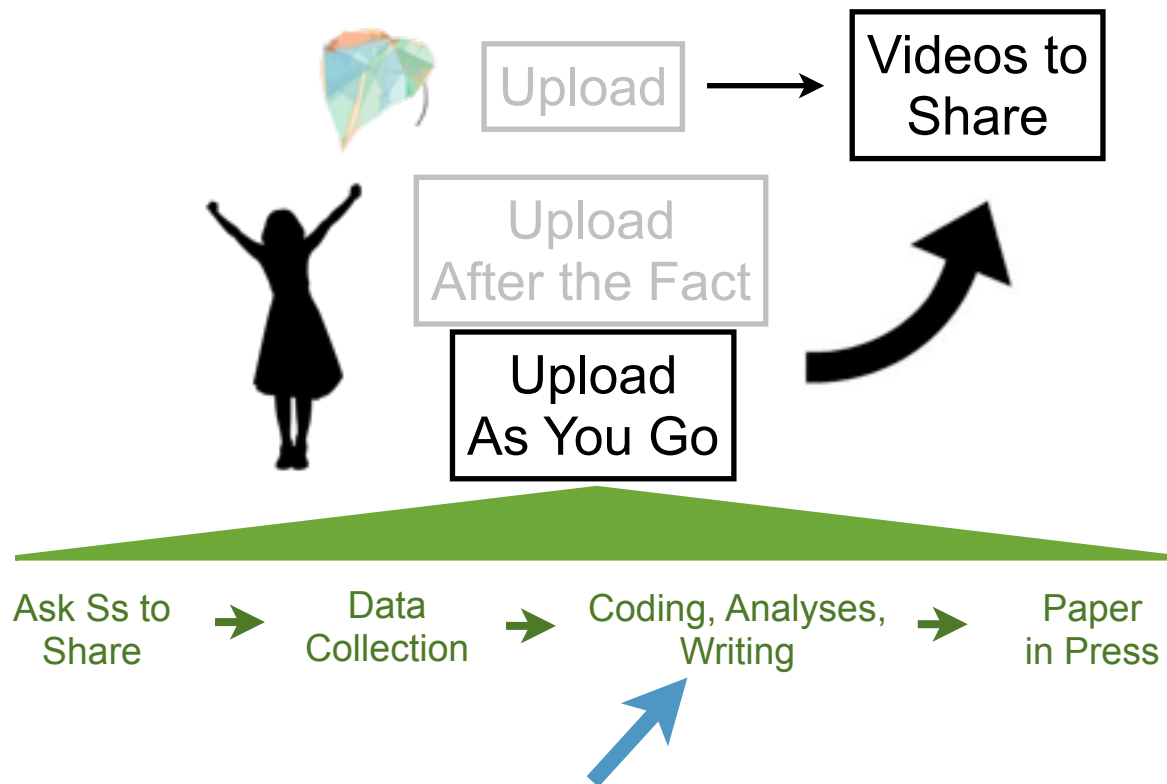
Roadmap



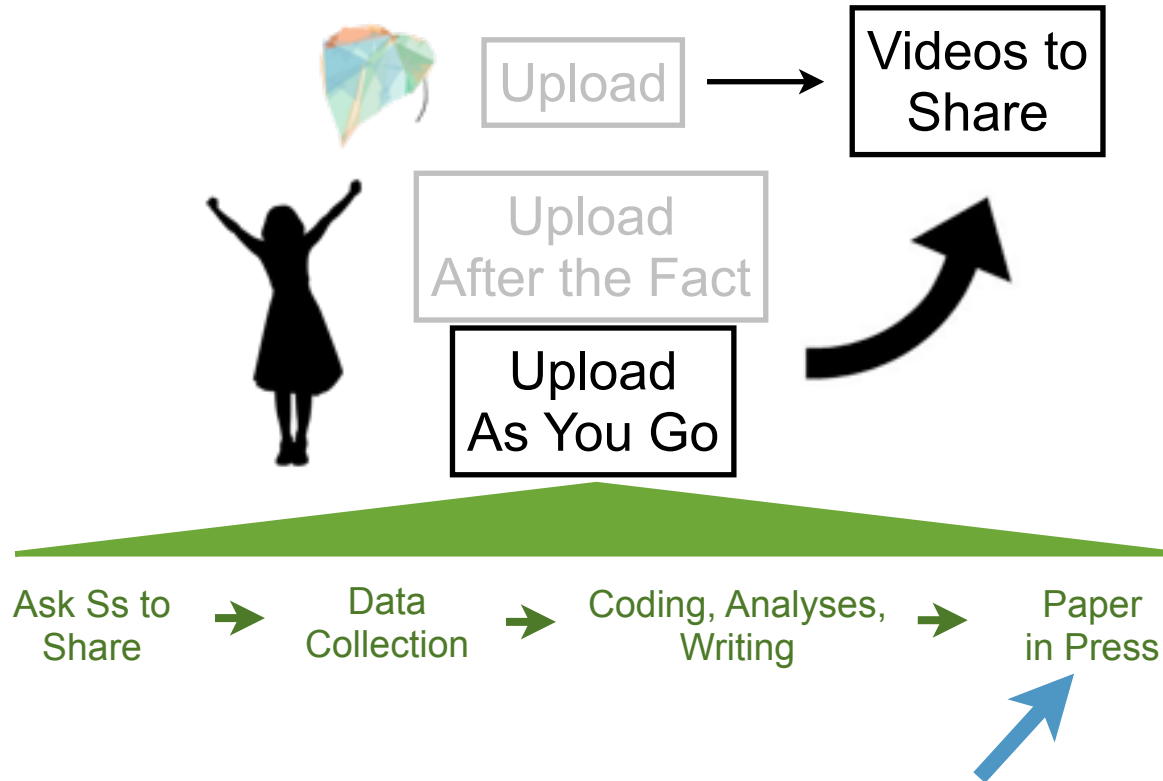
Roadmap



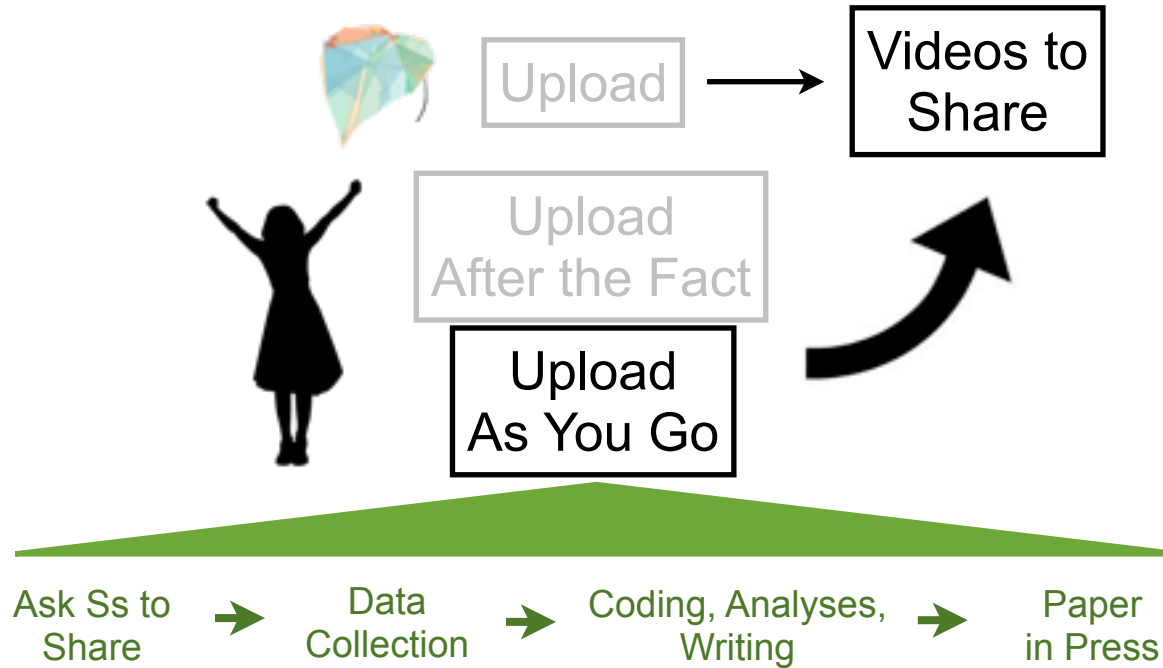
Roadmap



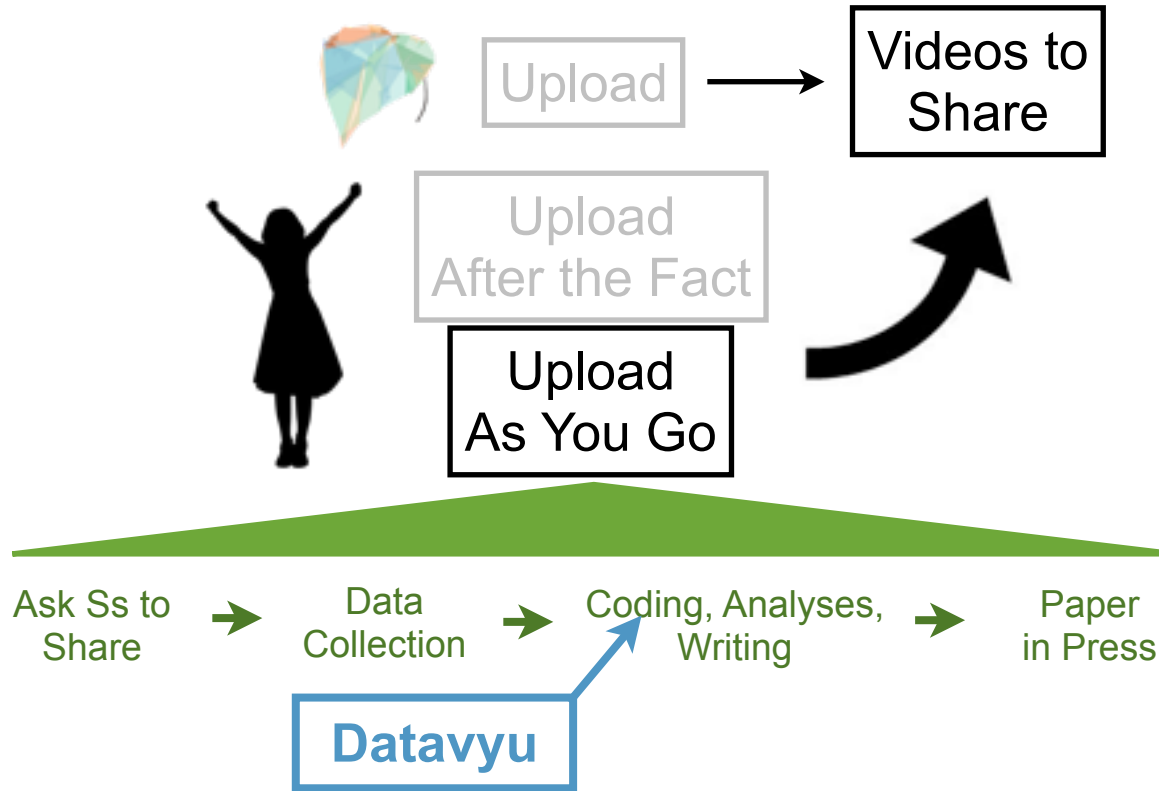
Roadmap

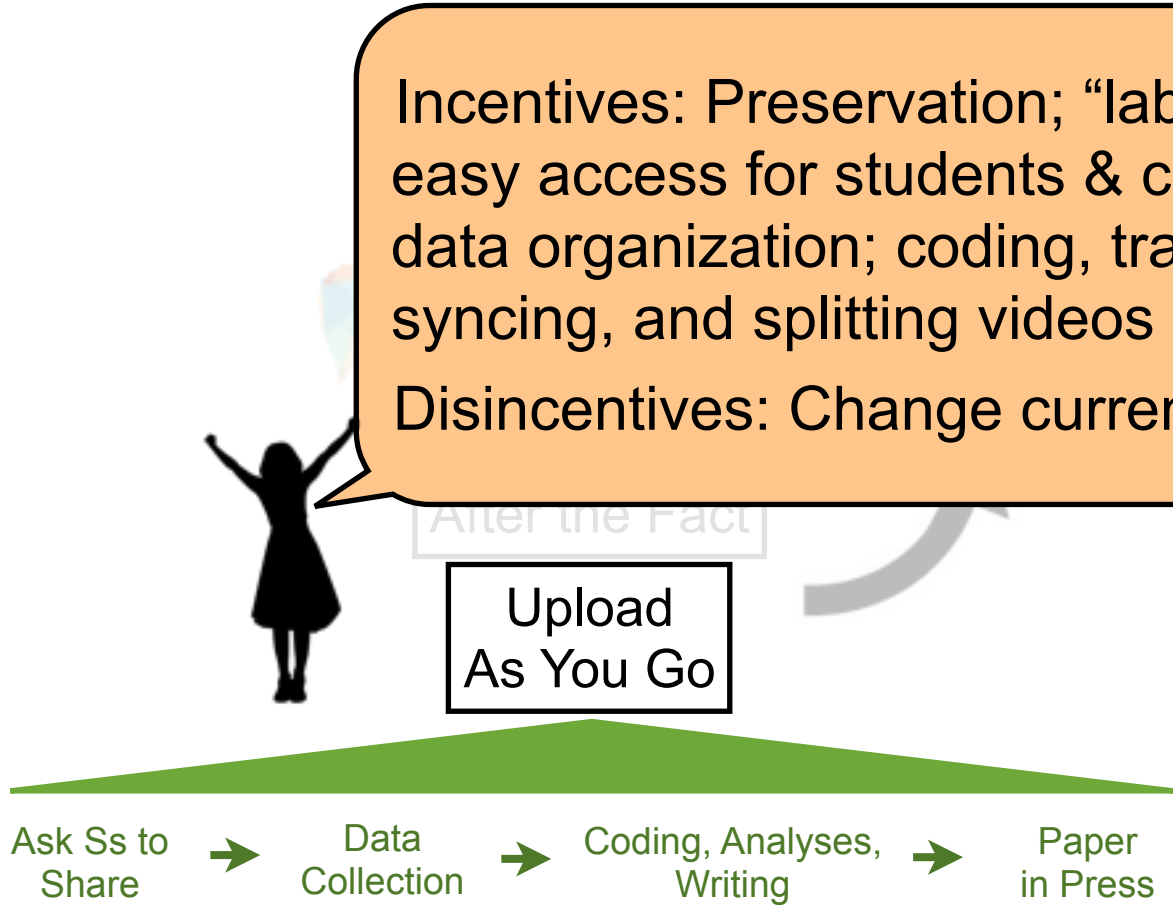


Roadmap



Roadmap





Incentives: Preservation; “lab server;”
easy access for students & collaborators;
data organization; coding, transcoding
syncing, and splitting videos

Disincentives: Change current practices

Upload
As You Go

Ask Ss to
Share



Data
Collection

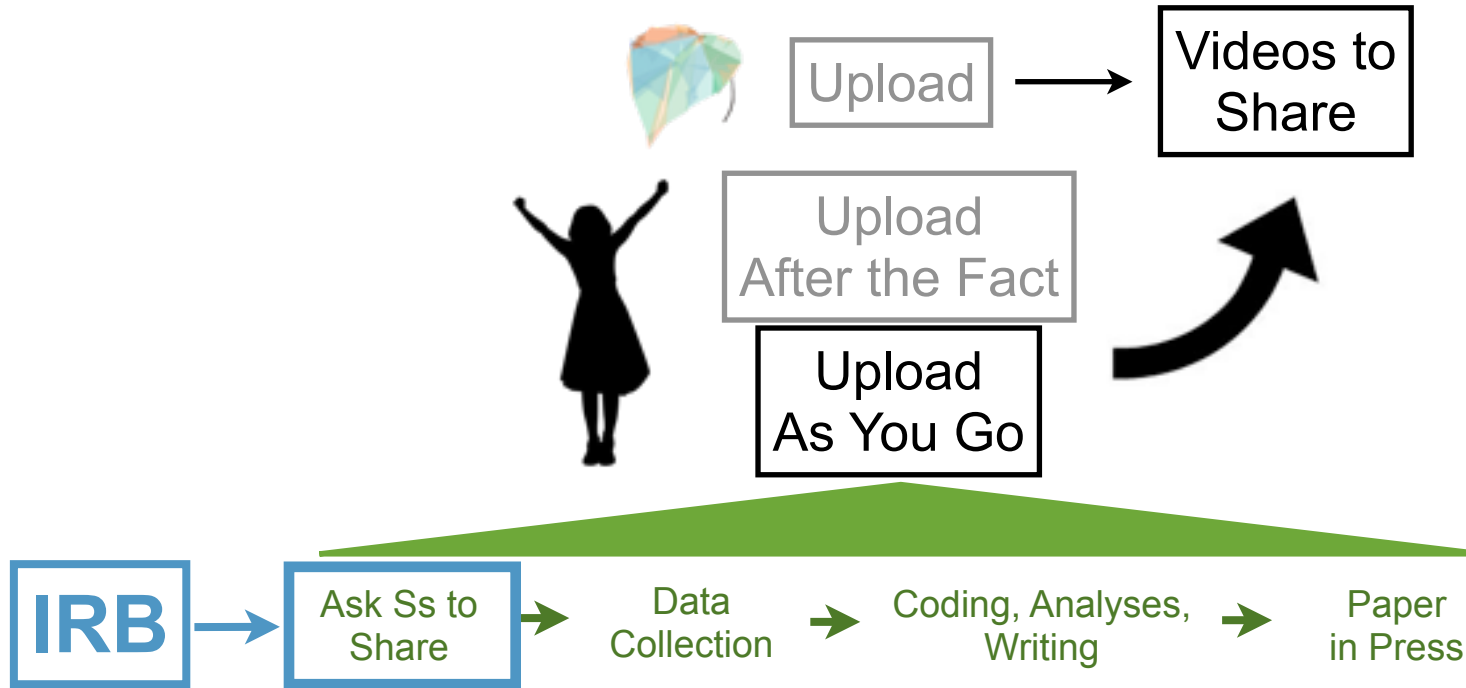


Coding, Analyses,
Writing



Paper
in Press

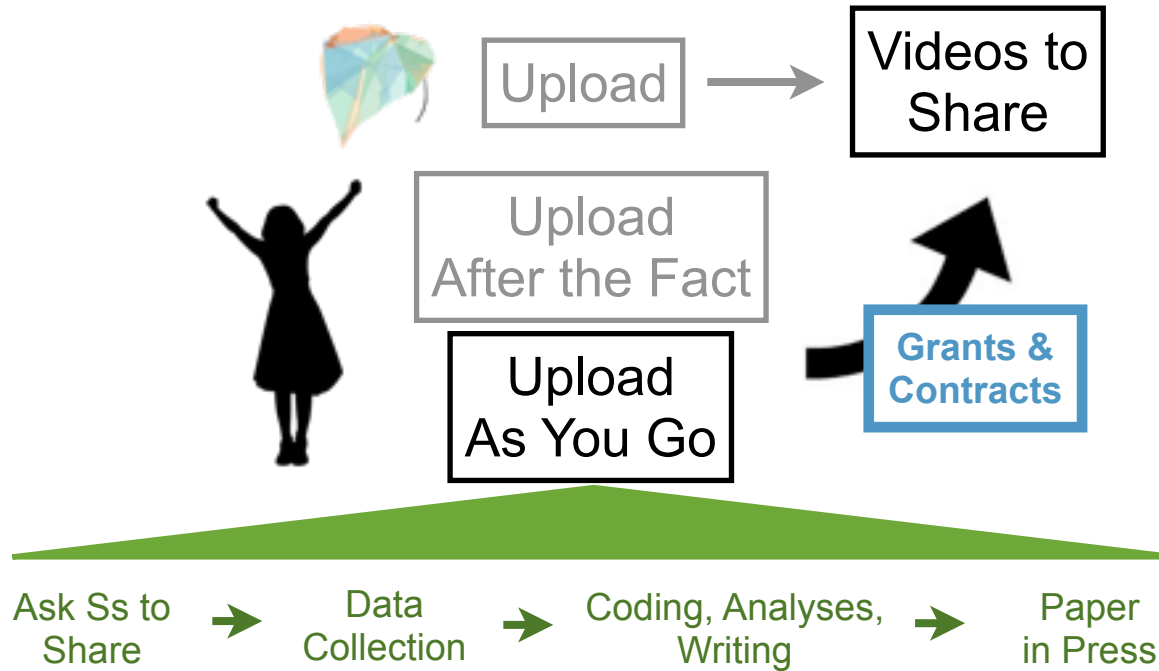
Roadmap



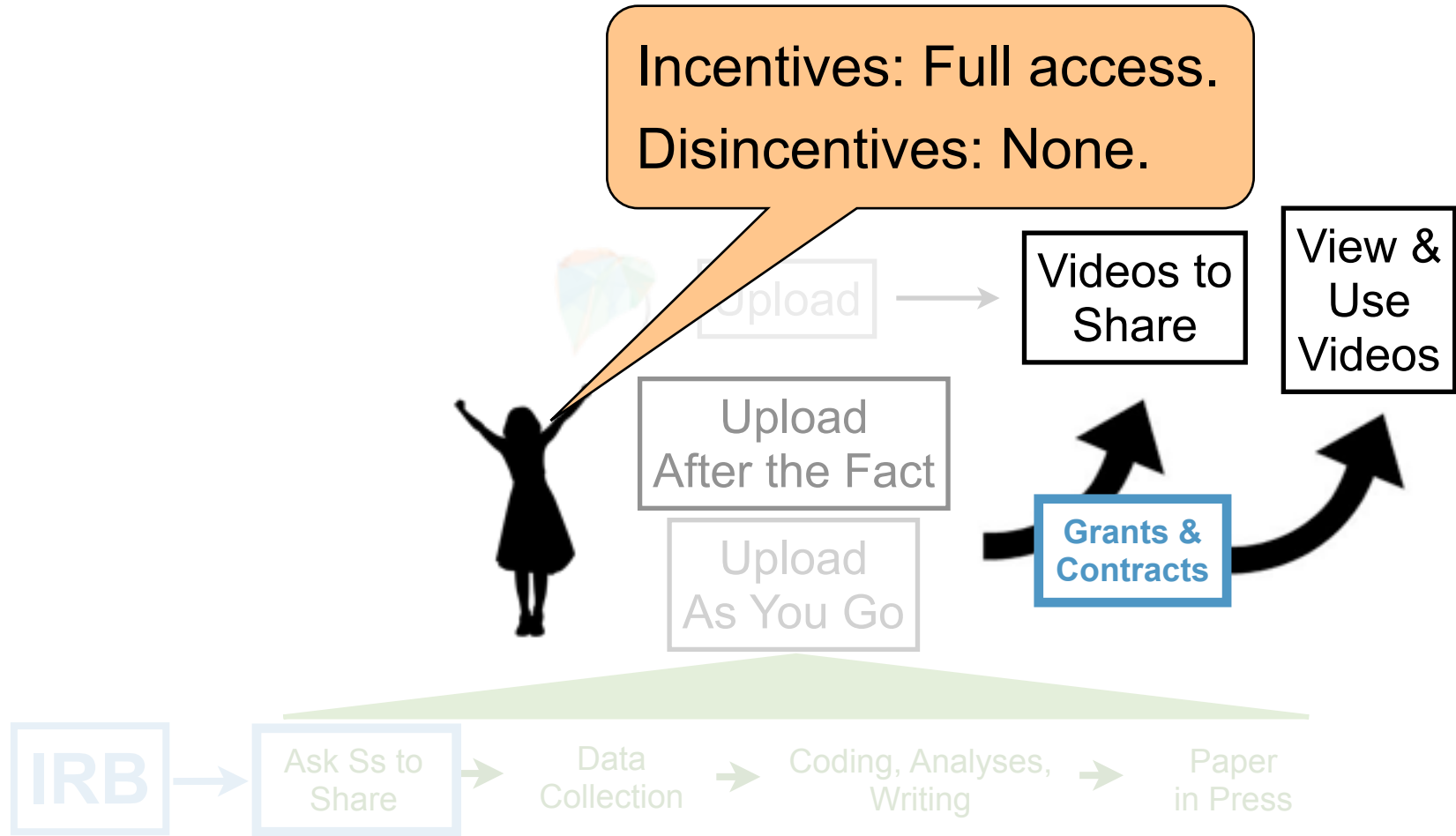
Incentives: Use of excerpts.
Disincentives: None!



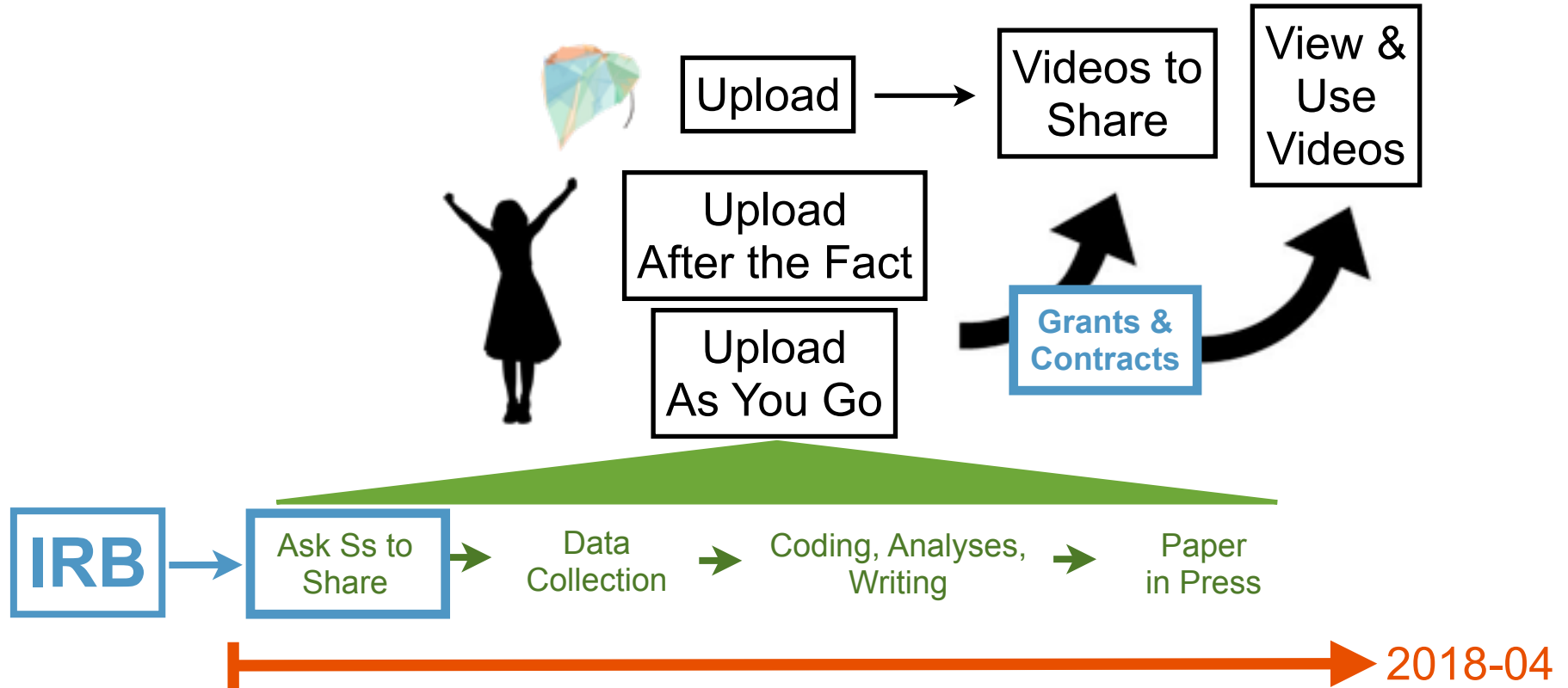
Roadmap



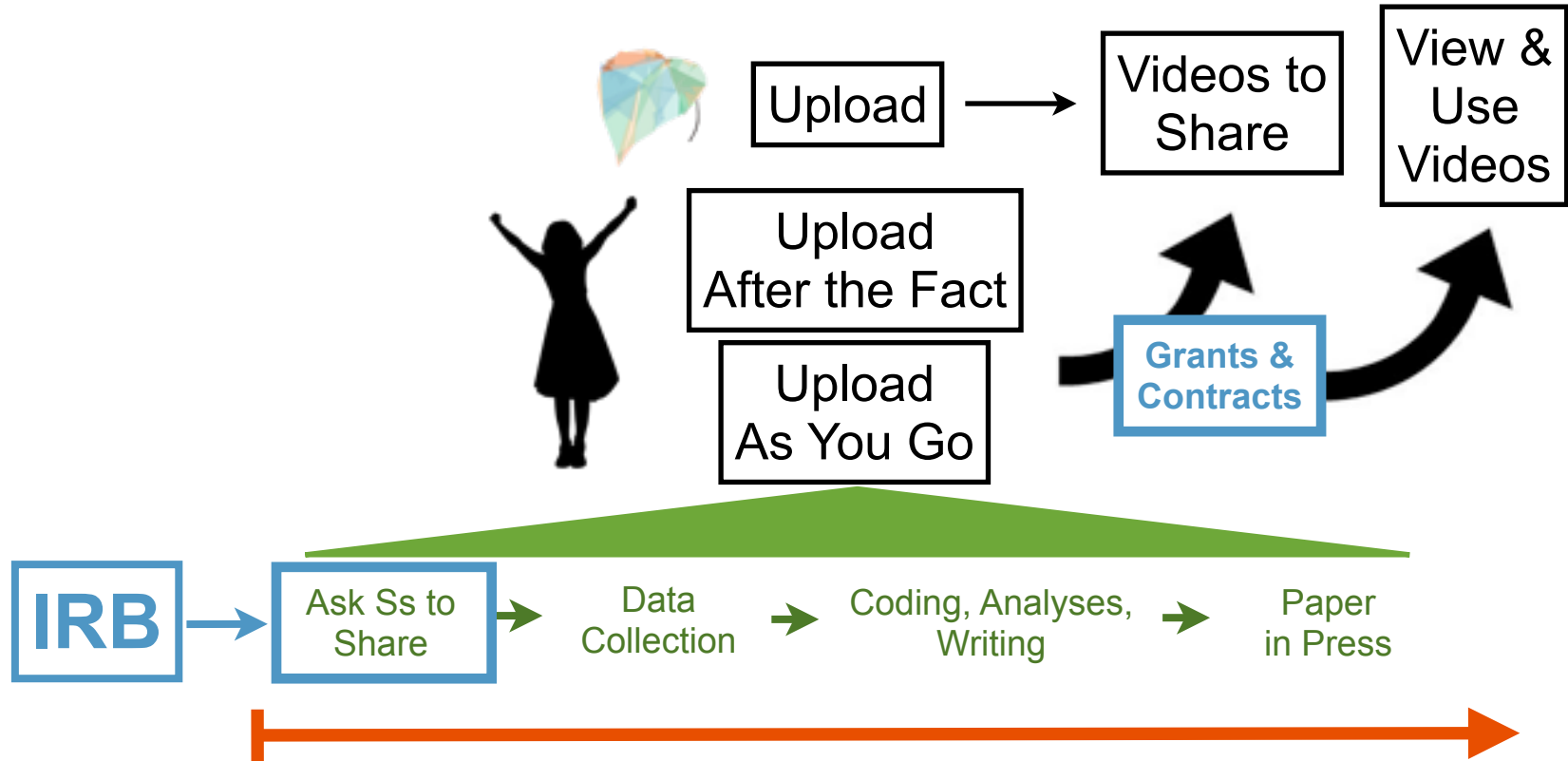
Incentives: Full access.
Disincentives: None.



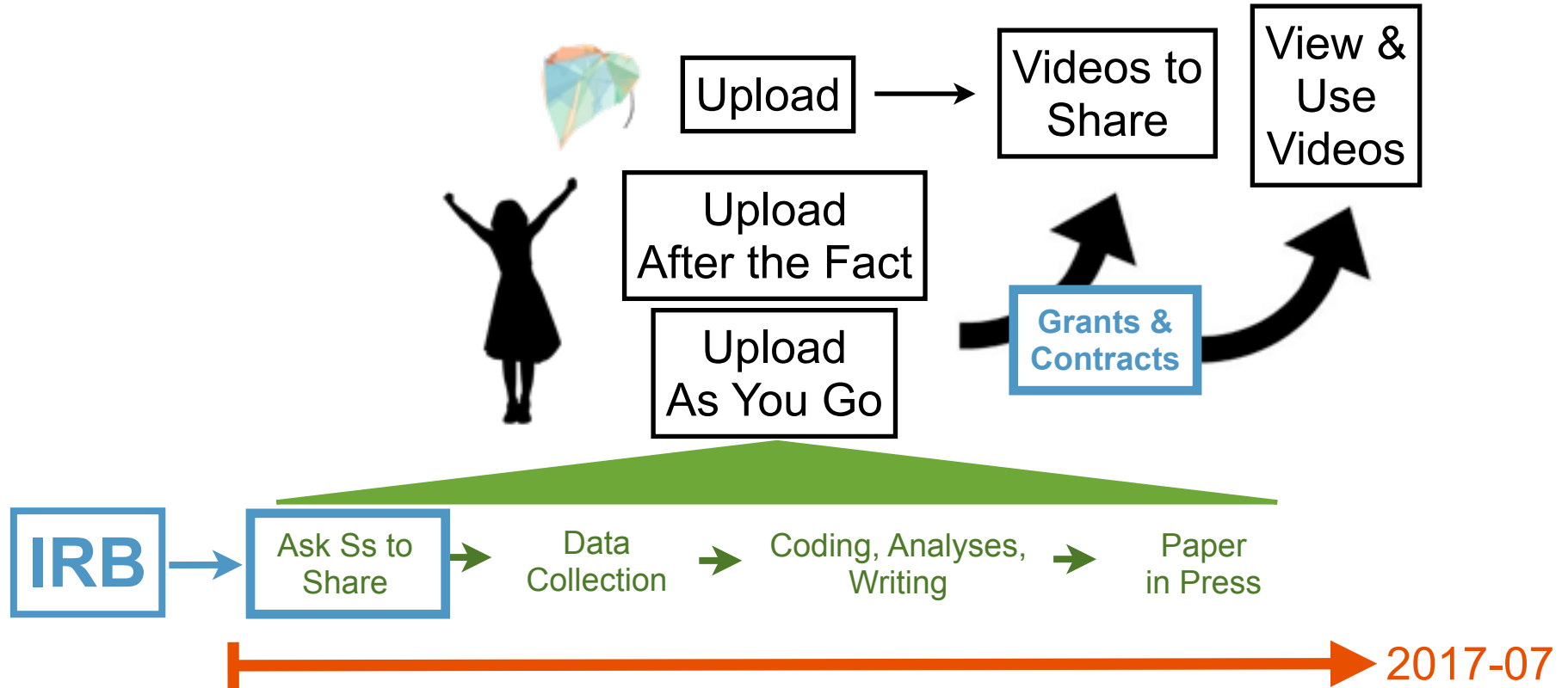
Timeline



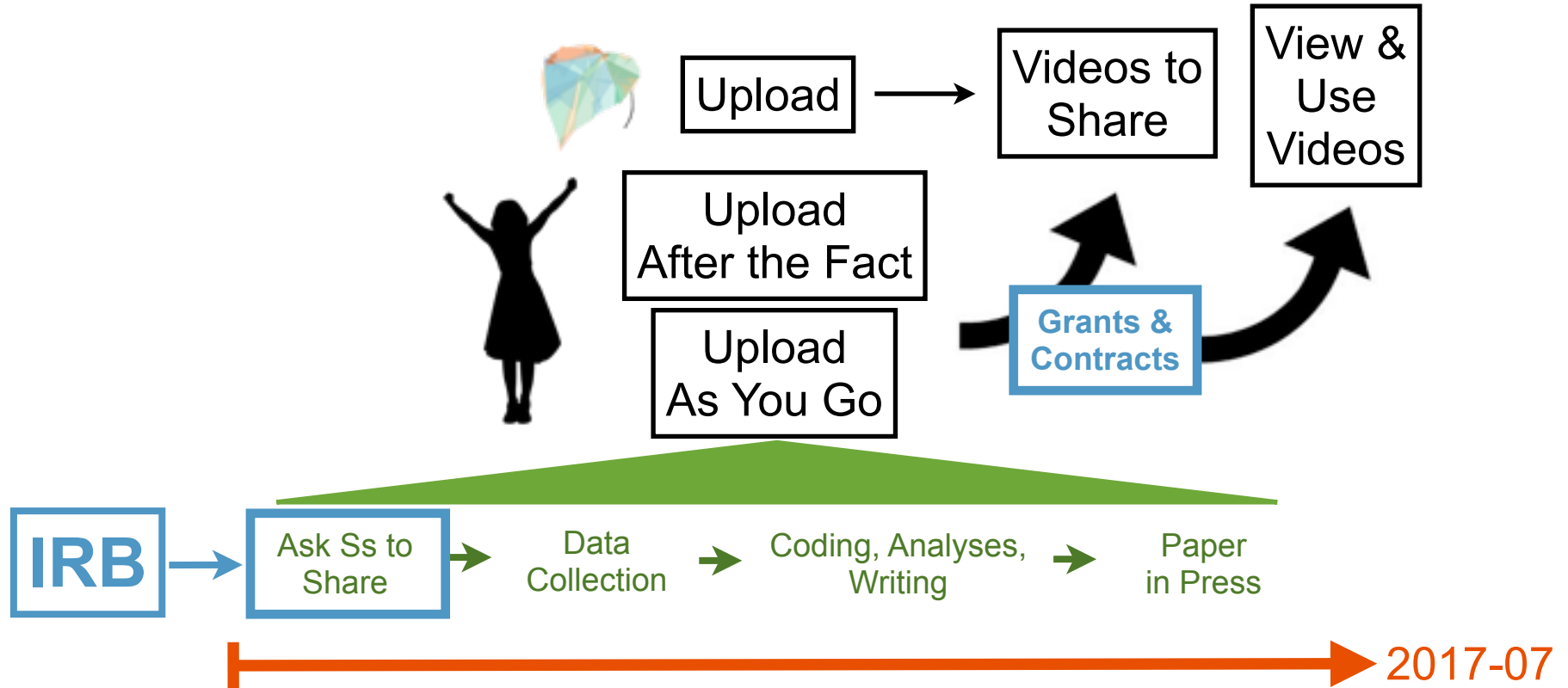
Timeline



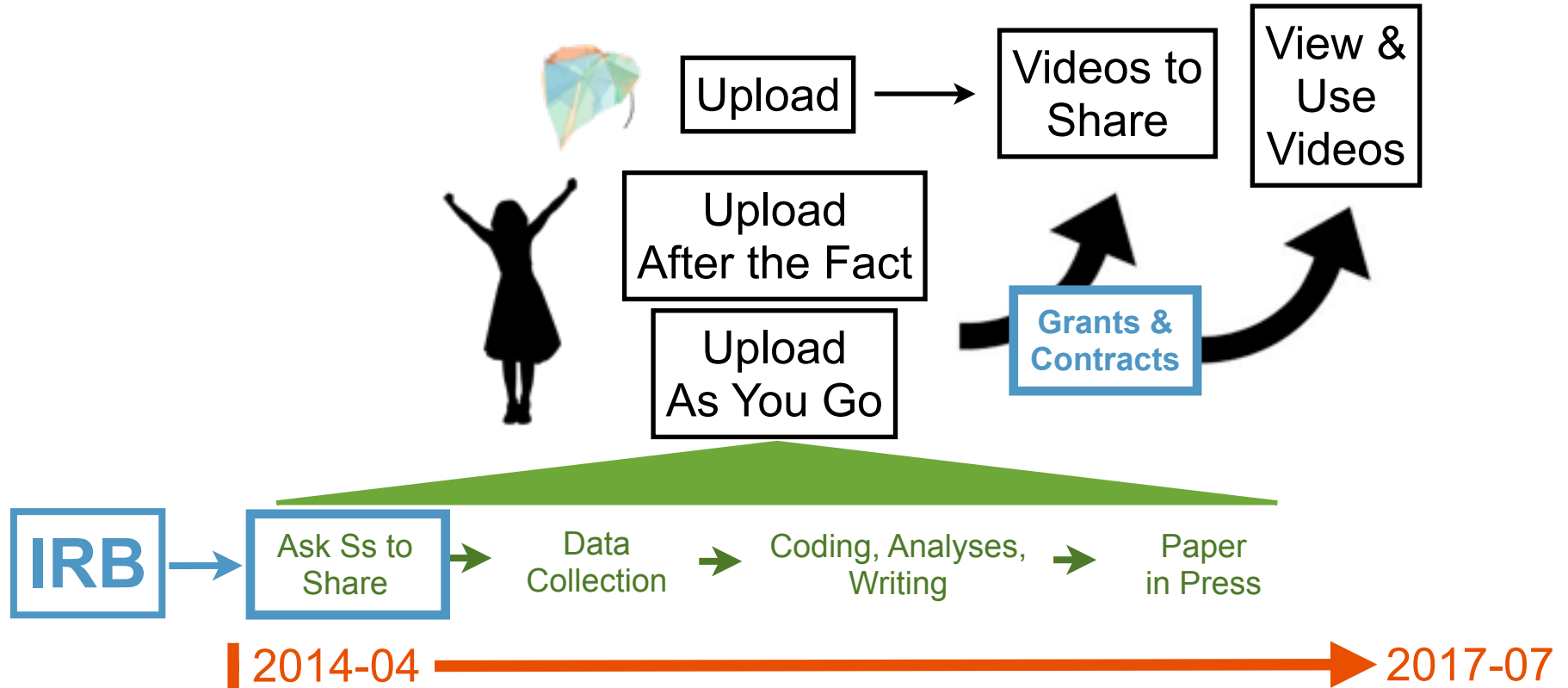
Timeline



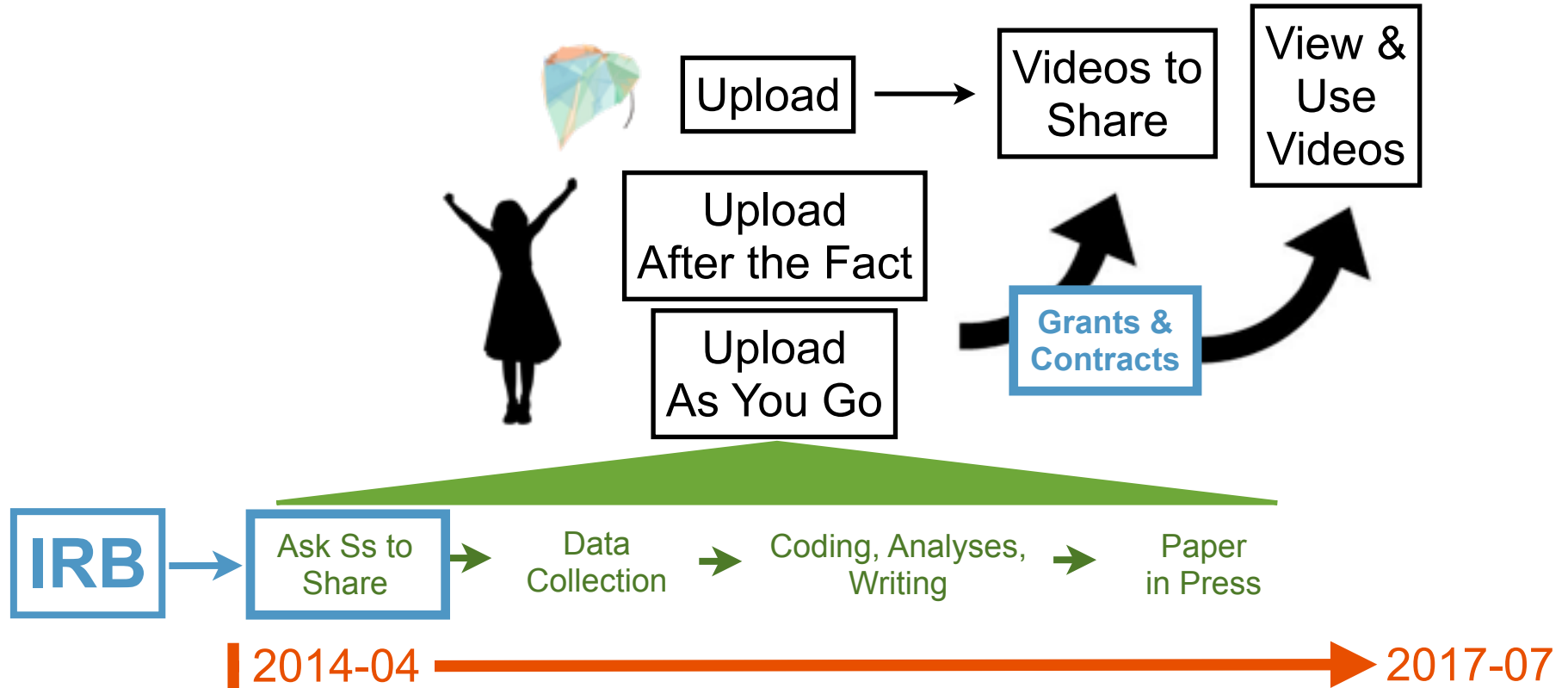
Timeline



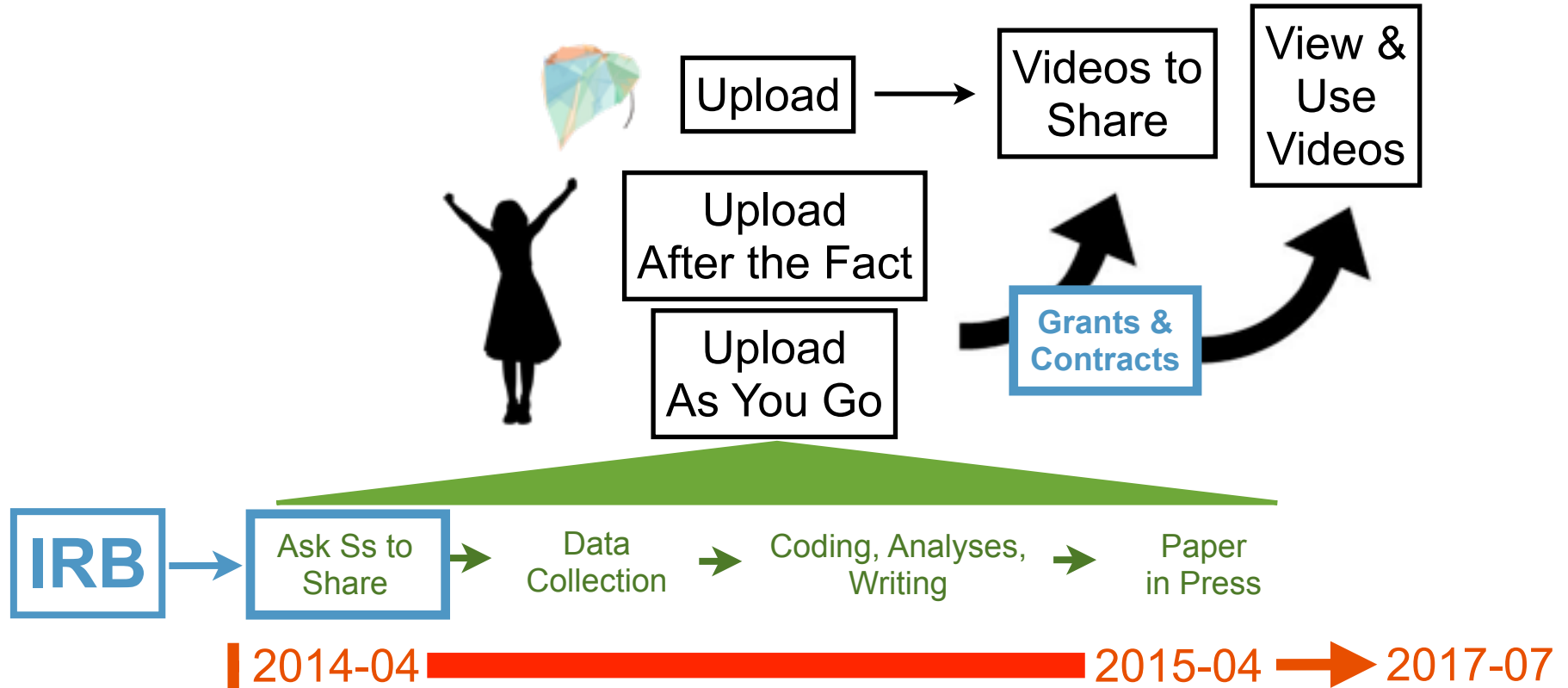
Timeline



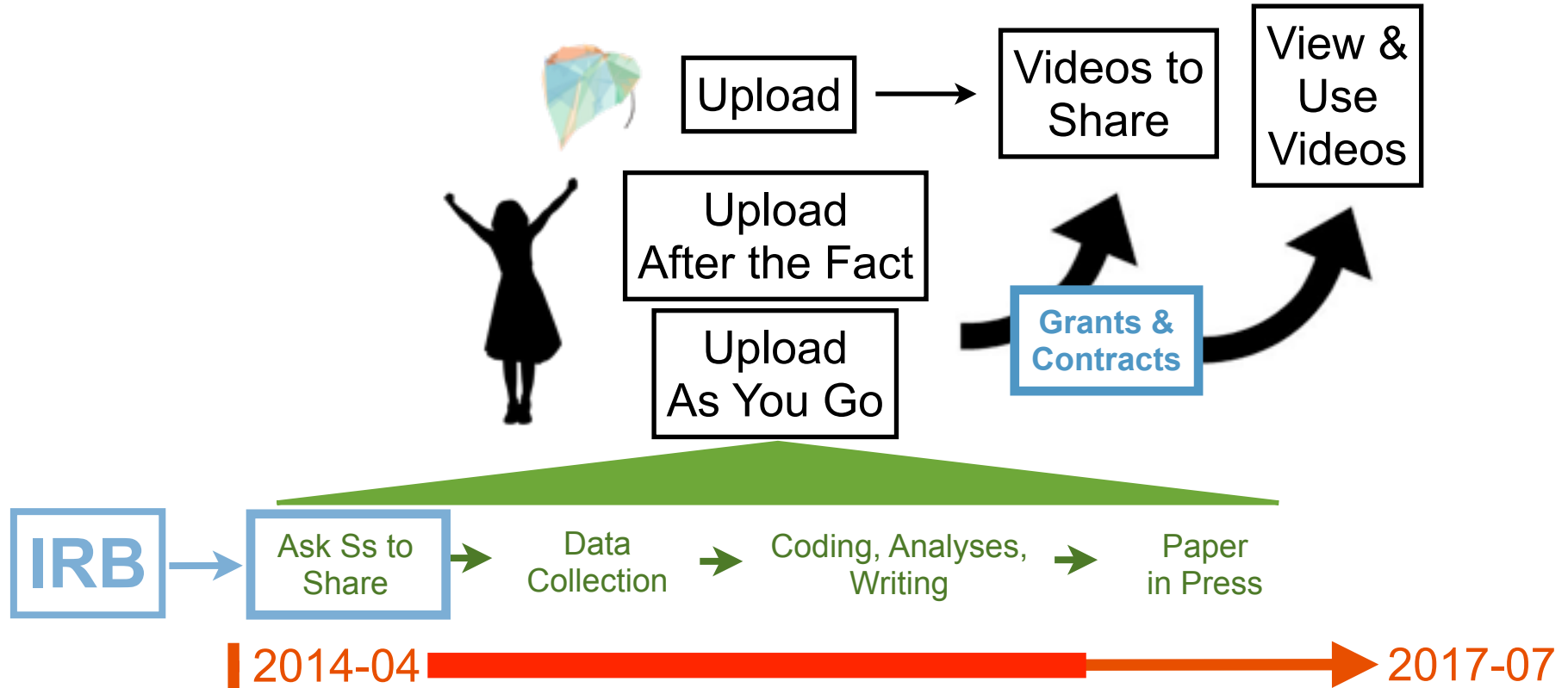
Timeline



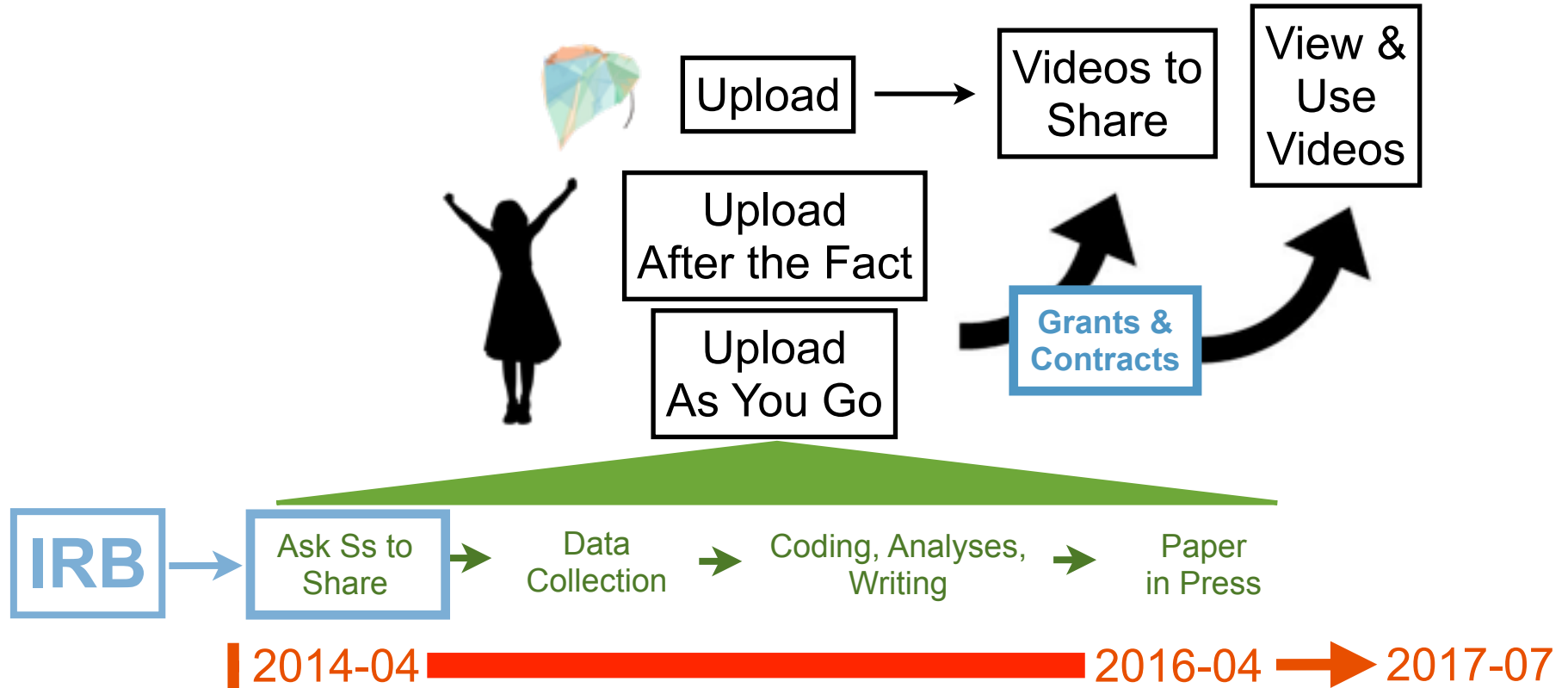
Timeline



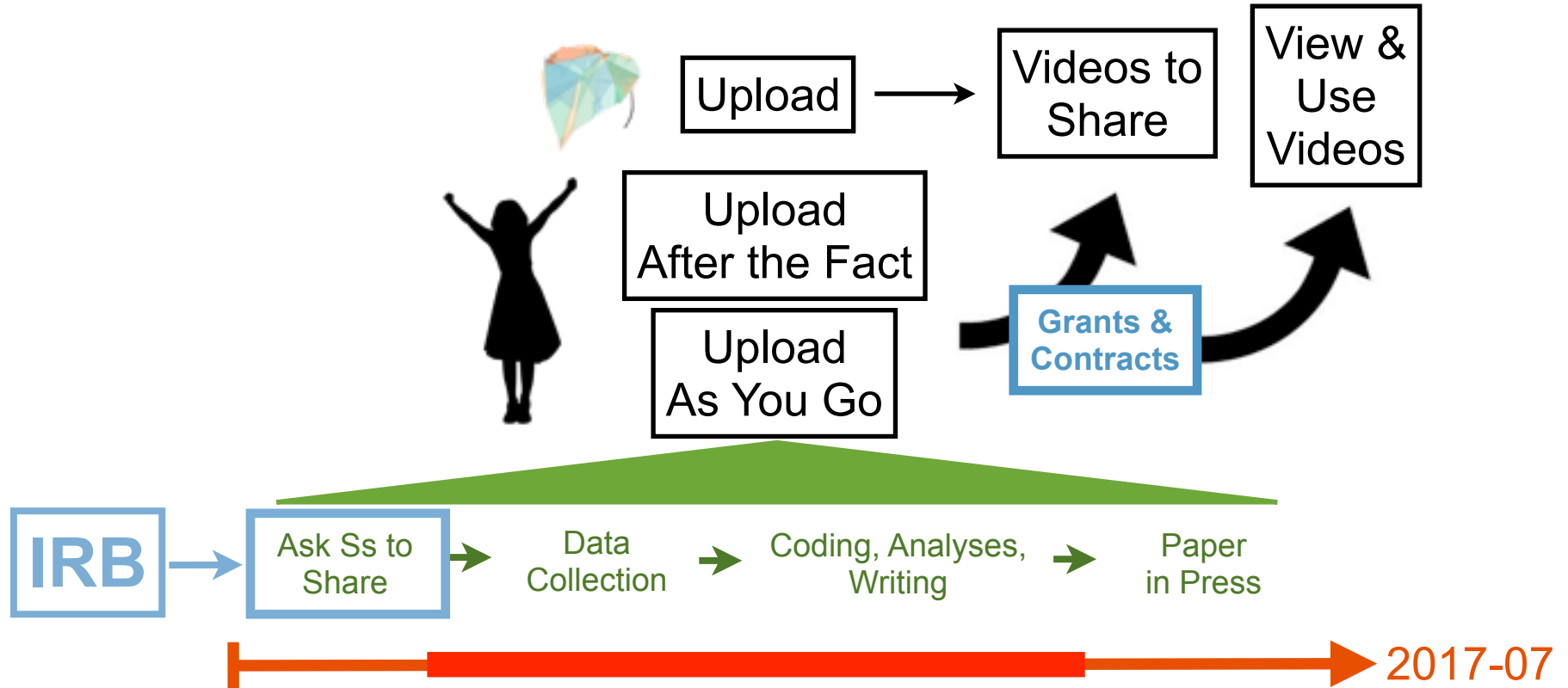
Timeline



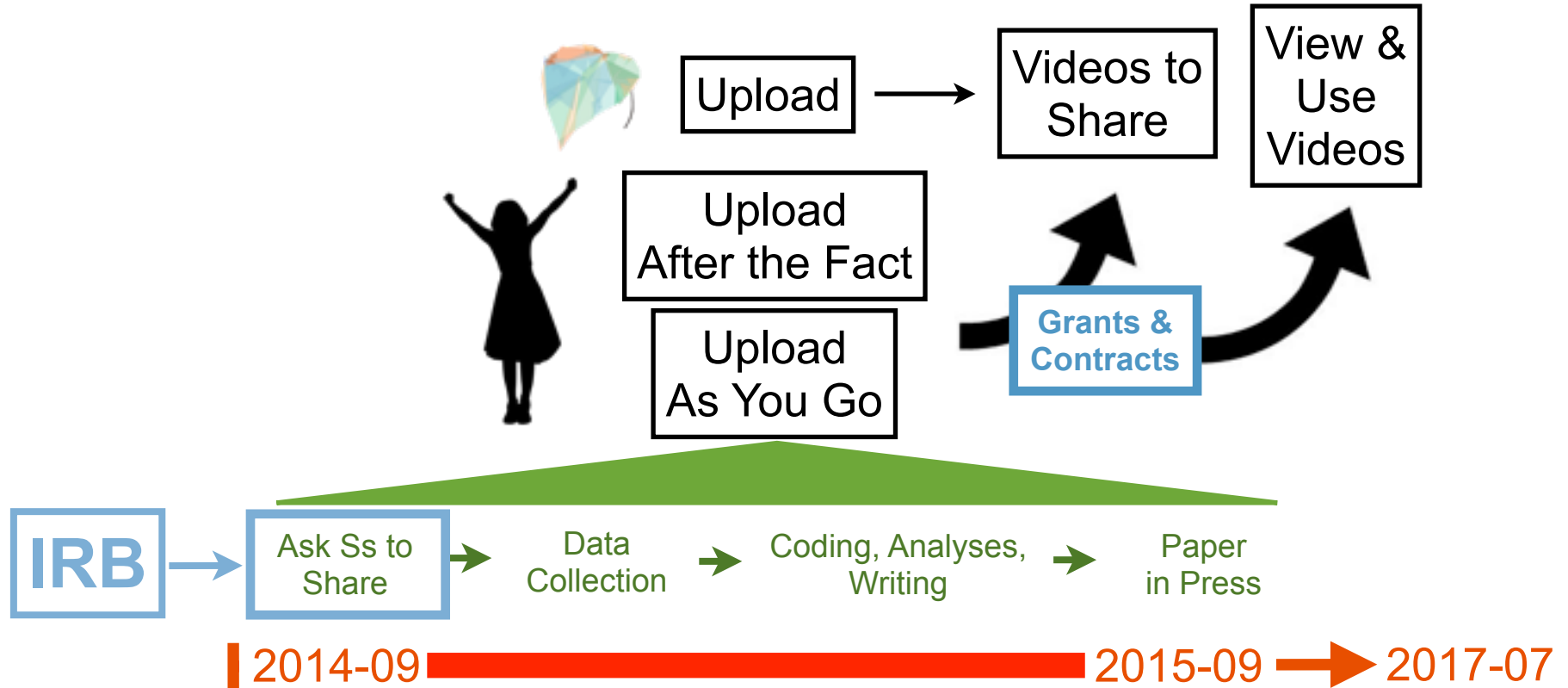
Timeline



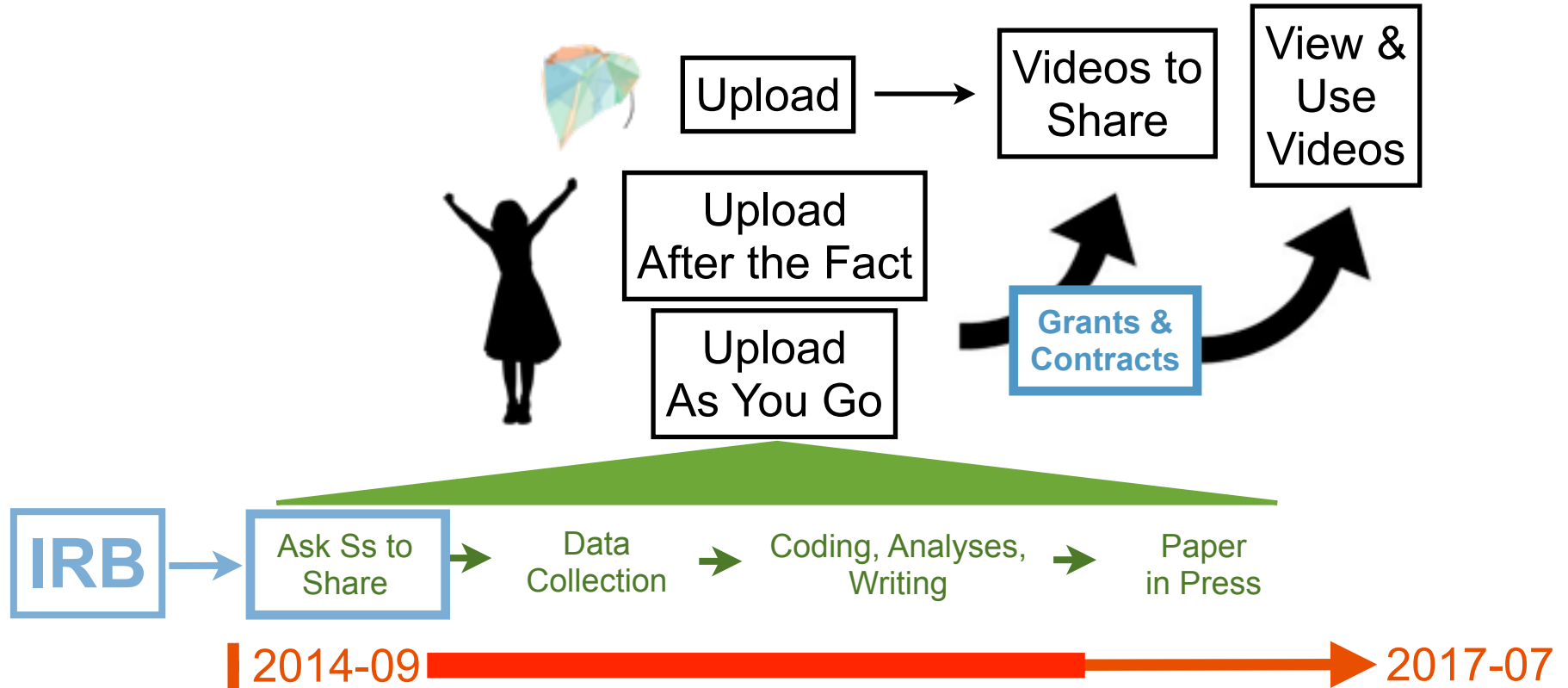
Timeline



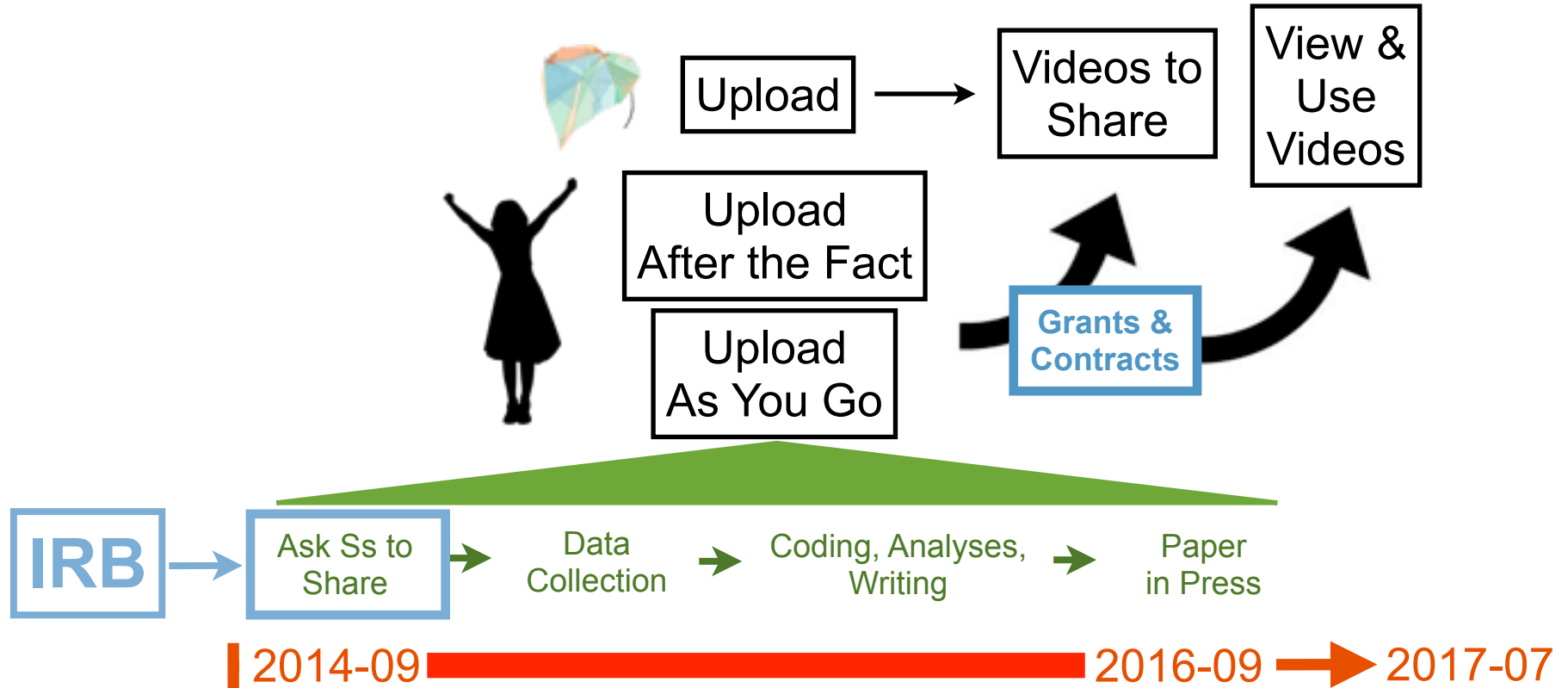
Timeline



Timeline



Timeline



Upload as you go (Labnanny)

- ▶ We need to get the video data into Databrary
- ▶ We can't rely on researchers uploading data after their paper is submitted
- ▶ We need to make it easy to upload data while researchers are collecting it



What we need from advisory board

- ▶ Advocate for data sharing and Databrary
- ▶ Lead by example
- ▶ Identify and share already-collected data
- ▶ Request participants to share their data
- ▶ Become authorized Databrary Investigators
- ▶ Explore the beta site and give us feedback
- ▶ Provide insights for designing upload and search



Meeting Agenda

- 1:00-1:15 Welcome and recent accomplishments
- 1:15-1:45 Roadmap
- 1:45-2:15 Policies**
- 2:15-2:30 Break
- 2:30-3:30 Beta: Upload and search, Datavyu 2.0
- 3:30-4:00 Researcher asks and support
- 4:00-4:10 Break
- 4:10-5:00 Positioning Databrary for the future
- 5:00-5:30 Wrap up and discussion



Policy framework

- ▶ Special problem with video: Identifiable data
- ▶ Policies that enable open sharing of identifiable data
 - ◀ Informed consent to share videos
 - ◀ Ensuring adherence to a common set of practices and ethical principles



Databrary release

- ▶ Template release form
- ▶ Informed consent: can share identifiable data if you tell participants (all depicted individuals) and they agree
- ▶ Standardization across contributors
 - ◀ Developed to correspond directly with release levels



Release levels

- ▶ Available restrictions on identifiable data:
 - ◀ **Did not ask:** If undocumented; functions as private
 - ◀ **Private:** Restricted to data owners and editors
 - ◀ **Shared:** Restricted to authorized Databrary investigators
 - ◀ **Shared + excerpts:** Restricted like shared, but excerpts may be used for informational, scientific, and educational purposes
 - ◀ **Public:** Available to the public



What data are being released?

- ▶ Videos from a session (or part of a session)
 - ◀ Can decline sharing for sessions or segments of sessions that contains sensitive information
- ▶ Session- and participant-level metadata that are identifiable
 - ◀ Birthdates, faces, names, interior of homes, classrooms, disabilities, self-reported health info, etc.
- ▶ Codes of behaviors



Databrary release: What is unique?

- ▶ Completely separates consent to participate from release to share
 - ◀ Consent to participate (in the study) before
 - ◀ Release to share after—it's clear what was recorded



Our template works

- ▶ Multiple iterations of release form and procedure for talking with participants
 - ◀ Tested on our own participants!
- ▶ Adaptable and flexible
 - ◀ Only local IRB must approve
 - ◀ Can collect decisions of all depicted individuals on one page
- ▶ Used in many contexts
 - ◀ International, oral/written, diverse ethnic groups, children with disabilities



Video example

databrary.org/user-guide/getting-started/release-script/example-videos.html

Video example



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 everything, so long as participants are asked
- ▶ We need as many researchers as possible to incorporate Databrary releases



Progress on releases

- ▶ 9 labs are currently requesting participant release
- ▶ 10 labs have pending IRB protocol amendments
- ▶ 17 other labs in process of adding Databrary to IRB



Discussion re: excerpts

- ▶ Most common (pre-Databrary) use case
 - ◀ Clips for talks, teaching
 - ◀ Demonstration of specific procedure or method
 - ◀ Relations between behaviors and codes
- ▶ Why ask for excerpts separately?
 - ◀ Effectively public
 - ◀ Needs to be clear to participants



Databrary and excerpts

- ▶ Should Databrary make excerpts publicly available on the site?
- ▶ Who decides which excerpts are public on Databrary?
 - ◀ Only the data owner or any PI on Databrary?



Authorization

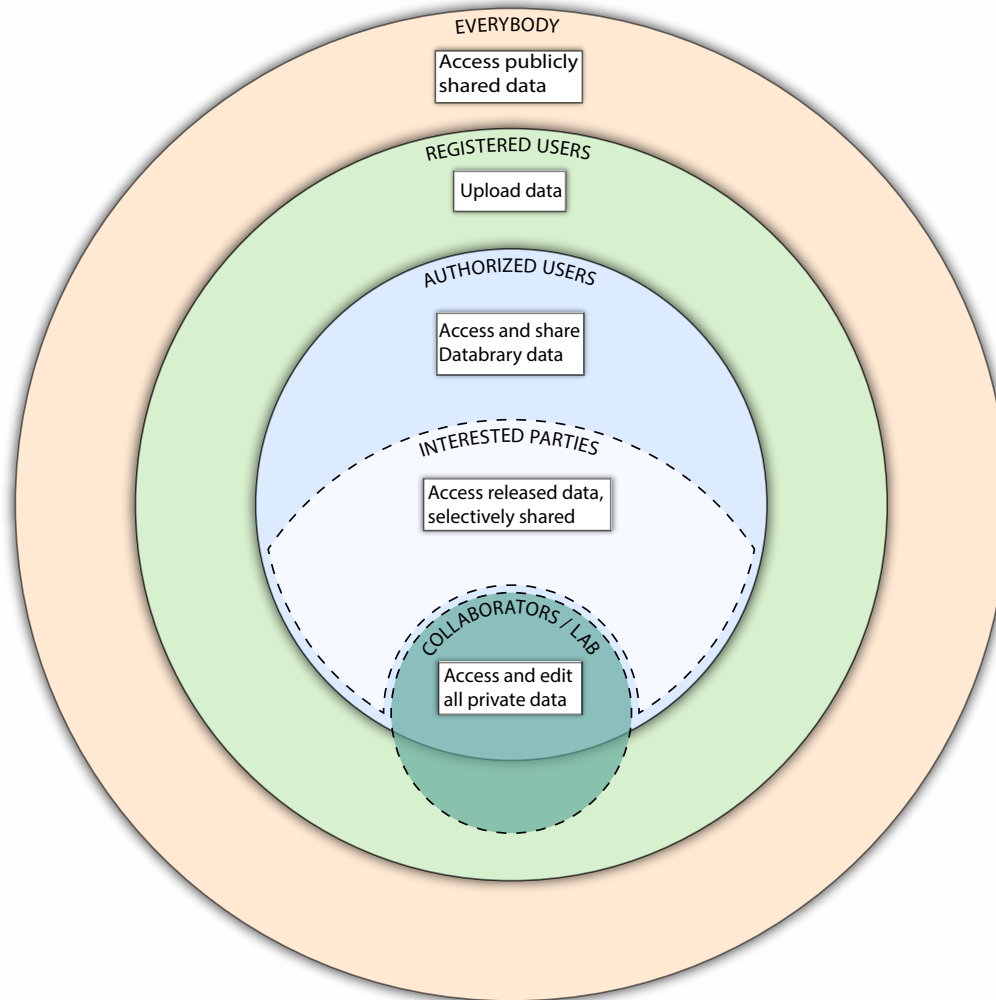
- ▶ Access to shared videos
- ▶ Contribute and/or use data
 - ◀ No requirement to contribute
- ▶ Institutional sign-off by Authorizing Official
 - ◀ Grants and contracts
 - ◀ No IRB needed



Access

- ▶ PIs can authorize and manage affiliates
- ▶ Can also add lab members and collaborators
 - ◀ Regardless of authorization
- ▶ Four pre-selected options
 - ◀ Lab and collaborators only
 - ◀ Databrary data only
 - ◀ Lab data and Databrary data
 - ◀ Proxy





Terms of the agreement

- ▶ Researcher upholds same ethical standards as with their own data
- ▶ Responsibility of researcher to get whatever approval is necessary when use constitutes human subjects research
- ▶ Institution must verify that Investigator is eligible to be a Databrary Principal Investigator



Progress on authorization

- ▶ 7 schools signed:
 - ◀ NYU, PSU, Rochester, Indiana, McMaster, UVA, Rutgers
- ▶ Points of clarification
 - ◀ Does not go to the IRB office for signing
 - ◀ Further, IRB approval not required for authorization
 - ◀ Only authorizing official can bind institution or enter agreement



Feedback

- ▶ Agreement written with NYU and PSU
- ▶ Piloted agreement on advisory board
- ▶ We want as much feedback as possible
 - ◀ Where are researchers getting hung up?
 - ◀ Where are institutions getting hung up?



Registration demo



Next step for authorization policy

- ▶ Shift to institutional agreements
 - ◀ Once schools are comfortable with Investigator Agreement
 - ◀ No substantive changes to re-frame as an institutional agreement
- ▶ Allows universities to manage their authorizations on Databrary



Beta release

- ▶ Sending email tomorrow (4/8) with private beta link
 - ◀ Register and request authorization
- ▶ Explore the site and tell us what you think!
- ▶ We will authorize you temporarily through the beta if needed
- ▶ Please don't download Tajik videos



Outreach plan for Databrary 1.0

- ▶ Community outreach
 - ◀ Invites to ICIS, CDS, SRCD members
 - ◀ Conference exhibits at major academic society meetings
 - ◀ Workshops on sharing identifiable data
- ▶ Timesavers as incentives
 - ◀ Developing 'boilerplate' Databrary language for grant proposals and reports
 - ◀ Data management plans (partner with DMPtool)
 - ◀ Resource sharing plans



What we need from advisory board

- ▶ Advocate for data sharing and Databrary
- ▶ Lead by example
- ▶ Identify and share already-collected data
- ▶ Request participants to share their data
- ▶ Become authorized Databrary Investigators
- ▶ Explore the beta site and give us feedback
- ▶ Provide insights for designing upload and search



Break

Meeting Agenda

- 1:00-1:15 Welcome and recent accomplishments
- 1:15-1:45 Roadmap
- 1:45-2:15 Policies
- 2:15-2:30 Break
- 2:30-3:30 Beta: Upload and search**
- 3:30-4:00 Researcher asks and support
- 4:00-4:10 Break
- 4:10-5:00 Positioning Databrary for the future
- 5:00-5:30 Wrap up and discussion



Datavyu current and future

- ▶ Released stable Datavyu 1.1 desktop version and user guide
- ▶ Addressing remaining bugs as they trickle in, but otherwise complete
- ▶ Next steps (to begin 2015+): Bring Datavyu into Databrary web framework
 - ◀ Target simple coding tasks, on-line
 - ◀ Later possibility for off-line version



Databrary site



October 2013 board meeting

- ▶ Lots of videos
 - ◀ Piled up into studies
 - ◀ Minimal organization
- ▶ “Static” views
 - ◀ Single presentation option
 - ◀ Limited flexibility



Development progress

- ▶ Revised technical requirements
 - ◀ Flexible discovery, browsing options
 - ◀ Responsive, intuitive editing, uploading
 - ◀ High-level visualizations, summaries
 - ◀ API, scripting interface
- ▶ Built new, completely dynamic JS interface
- ▶ Front-end uses API



Uploading after the fact

- ▶ Learned a lot from manually curating and organizing existing and new datasets
- ▶ Broadened types of possible, standard metadata
- ▶ Developed better ways to group and relate videos, metadata, studies, datasets



Lessons from the data

- ▶ Always exceptions, irregularities in real data
 - ◀ Missing, multiple videos
 - ◀ Don't fit cleanly in defined groups
 - ◀ Excluded and later re-purposed data
 - ◀ Always something *non-rectangular*
- ▶ Different types of data (classroom, public, longitudinal) have different needs



Updated model

- ▶ Better understanding of constraints, required flexibility
- ▶ Need responsive, iterative, “data-driven” approach
 - ◀ Real data are always more nuanced



Walkthrough of live site

Next steps

- ▶ Improve discovery within and across datasets
 - ◀ Searching: by keyword, text, descriptions, names, etc.
 - ◀ Filtering: limiting results by age, gender, numeric values, etc.
 - ◀ Sorting: order to present results, for scanning visually
- ▶ View summarized or aggregated data (counts, distributions, visualizations, etc.)
- ▶ Many existing user stories and use cases



User stories: Teaching and talks

- ▶ Video clips for teaching
- ▶ Illustrate an idea
- ▶ Show the range of behaviors and exceptions
- ▶ Show an excerpt in a talk



User stories: Pre-research

- ▶ Browse the work in my field
- ▶ Decide whether a study is worth doing
- ▶ Preliminary data for grant proposal
- ▶ Ideas and inspiration
- ▶ Replicate, expand on, or review previous work based on the procedure or coding manual



User stories: Research

- ▶ Repurpose videos for new uses
- ▶ Replicate existing work by recoding videos
- ▶ Grow sample size
- ▶ Include participants from other contexts and populations
- ▶ Conduct integrative analyses
- ▶ Complete grant progress report



Next steps

- ▶ Primary focus *was* on browsing, searching, using (downloading, coding, commenting, tagging)
- ▶ New focus: Upload as you go
- ▶ Build browser interface that also makes sense for entering and editing data



User stories: 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 analysis
- ▶ Customize presentation of dataset (title, excerpts, display)



Design process for upload as you go

- ▶ Focus group including developers, UI/UX experts, researchers
- ▶ Refine user stories
- ▶ Identify requirements and priorities
- ▶ Iterative process, test and improve interface



Inferring structure as you go

- ▶ Initial questions about study design are useful
- ▶ But, things are always changing midcourse
- ▶ Allow users to iterate easily by finding smart ways to expand structure as they go

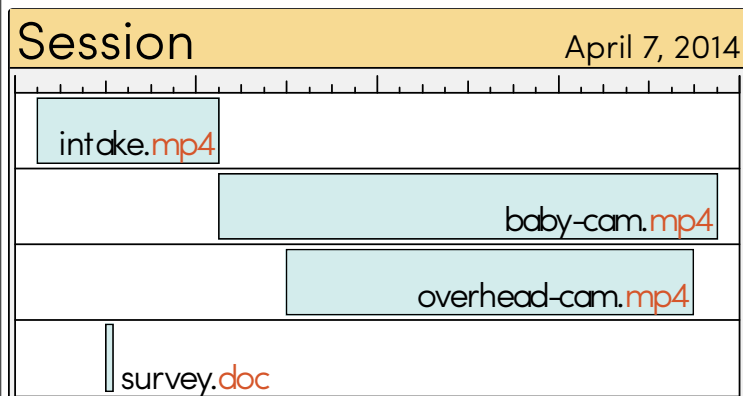


Meeting Agenda

- 1:00-1:15 Welcome and recent accomplishments
- 1:15-1:45 Roadmap
- 1:45-2:15 Policies
- 2:15-2:30 Break
- 2:30-3:30 Beta: Upload and search
- 3:30-4:00 Researcher asks and support**
- 4:00-4:10 Break
- 4:10-5:00 Positioning Databrary for the future
- 5:00-5:30 Wrap up and discussion



Session is the basic unit



- ▶ Covered by a *release level*
- ▶ Data collected at same time
- ▶ Contains timeline of raw data files
- ▶ Annotated with metadata



Session metadata

- ▶ Release level
- ▶ *Date*
- ▶ Pilot
- ▶ Excluded
- ▶ Participants
- ▶ Conditions
- ▶ Tasks
- ▶ Groups
- ▶ Location



Session release level

- ▶ Not asked
- ▶ Private
- ▶ Shared
- ▶ Shared + excerpts
- ▶ Public



Session date

- ▶ Technically considered identifiable information
- ▶ Could be unknown (Gesell)



Pilot

Indicates that the methods used in the session were not finalized or standard



Excluded

Indicates that session was not usable for the target study

▶ Reason

- ◀ Did not meet inclusion criteria
- ◀ Procedural/experimenter error
- ◀ Withdrew/fussy/tired
- ◀ Outlier
- ◀ ...



Participant

Represents an individual depicted or represented in this session

- ▶ **Subject ID**
- ▶ ***Birthdate***
- ▶ **Gender**
- ▶ **Race (NIH)**
- ▶ **Ethnicity (NIH)**
- ▶ *Disability/typically dev*
- ▶ Language
- ▶ Country (of origin)
- ▶ SES?
- ▶ ...



Task

Represents a particular task, activity, or phase that occurred during all or a portion of the session

- ▶ **Name**
- ▶ Summary
- ▶ **Description**



Condition

Represents an experimenter-determined manipulation (within or between sessions)

- ▶ **Name**
- ▶ Summary
- ▶ Description
- ▶ Other variables...



Group

Represents a grouping of sessions determined by an aspect of the data (participant ability, age, experience, measurements used/available)

- ▶ **Name**
- ▶ Summary
- ▶ Description
- ▶ Other variables...



Location

Represents a particular setting or other variable aspect of where/when/how session data were collected

- ▶ **Setting**

- ◀ lab, home, classroom, outdoors, clinic, museum ...

- ▶ Language (of study materials)

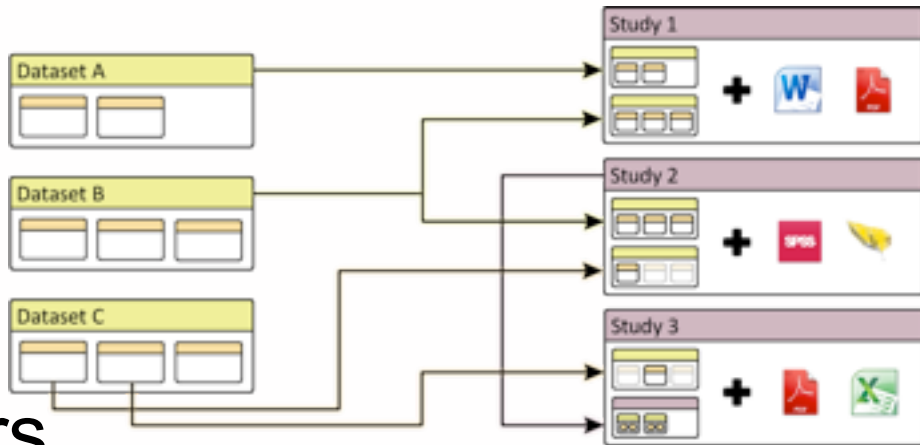
- ▶ **State**

- ▶ **Country**

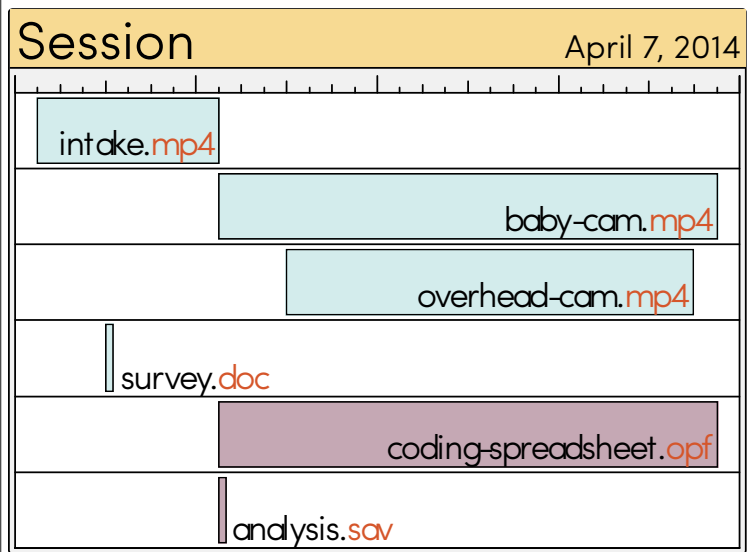


Datasets and studies (provenance)

- ▶ Datasets: raw data collected in sessions
- ▶ Studies: generated data, analyses, papers
- ▶ Studies include sessions from datasets and “layer” files over them
- ▶ Allow re-purposing, re-presenting data



Session is the basic unit



- ▶ Covered by a *release level*
- ▶ Data collected at same time
- ▶ Contains timeline of raw data files
- ▶ Annotated with metadata
- ▶ Analysis files layered on



Datasets and Studies

- ▶ One to one mapping likely to be rare
 - ◀ Excluded Ss, pilot Ss
 - ◀ In interest of transparency and repurposing, should be no disincentive for sharing
- ▶ Excerpts only, displays only
 - ◀ Studies, no datasets
- ▶ Repurposing data
 - ◀ Studies and datasets



Simple after-the-fact cases

- ▶ Easy to upload studies with:
 - ◀ Publication, abstract
 - ◀ Excerpts only
 - ◀ Stimuli, displays



What's on the table for standardizing?

- ▶ Metadata variables describing sessions, participants, studies, datasets
- ▶ Not interfering with the science
- ▶ Not standardizing tasks
 - ◀ Subgroups can do this (emo-group)
- ▶ Not standardizing codes
 - ◀ Datavyu is agnostic



Discussion: How much to standardize

- ▶ Continuum between enforced structure and flexibility
- ▶ Currently, types of metadata (participant, task, location, etc.) are fixed
 - ◀ Is this enough? Too much? Are there others?
 - ◀ Should we let users define their own?
- ▶ Variables within metadata (birthdate, language, description) are user-extensible
 - ◀ Should we restrict these? Add more?



Benefits and costs of standardization

- ▶ More powerful searching
 - ◀ Video or session level searches
- ▶ Aggregating across datasets
- ▶ But possible change in current practices
- ▶ More work for PIs to document measures



Potential solutions for flexibility

- ▶ Encourage use of existing tags (pop-up suggestions)
- ▶ Our curators clean up after the fact (merge conceptually identical tags)
- ▶ Leave the mess as-is
 - ◀ Semantic search (google: “crawl” finds “crawling”, “crawlers”, etc.)



Already doing stuff as you go

- ▶ Organizing, typing, writing data anyway after each data collection session
 - ◀ How is this done now?
 - ◀ How can we make this non-invasive/simpler than current practices?
- ▶ Uploading, transcoding, splitting videos anyway
- ▶ Exporting to SPSS/Excel/Datavyu anyway



Things that change later

- ▶ Exclusions, groups, entire study organization can change later
- ▶ Needs to be easy to do this in Databrary, drag around sessions, include/exclude, add to study, make new group
- ▶ Easy to create, download excerpts
- ▶ Last step: add publication citation, done!



Taking advantage of your Databrary lab

- ▶ Organization within labs requires less standardization
 - ◀ Each can have their own standards
- ▶ Search within your own data
- ▶ Mine your data for subsequent analyses
- ▶ Easy to disseminate examples
- ▶ Need effective 'overview' and 'monitoring' tools



What we need from advisory board

- ▶ Advocate for data sharing and Databrary
- ▶ Lead by example
- ▶ Identify and share already-collected data
- ▶ Request participants to share their data
- ▶ Become authorized Databrary Investigators
- ▶ Explore the beta site and give us feedback
- ▶ Provide insights for designing upload and search



Break

Meeting Agenda

- 1:00-1:15 Welcome and recent accomplishments
- 1:15-1:45 Roadmap
- 1:45-2:15 Policies
- 2:15-2:30 Break
- 2:30-3:30 Beta: Upload and search
- 3:30-4:00 Researcher asks and support
- 4:00-4:10 Break
- 4:10-5:00 Positioning Databrary for the future**
- 5:00-5:30 Wrap up and discussion



Quick SWOT analysis

Strengths	Weaknesses
Opportunities	Threats



Strengths

- ▶ Solution for sharing identifiable data
- ▶ Growing awareness, visibility within content domain
- ▶ Flexible data model, poised for growth
- ▶ University, agency support
- ▶ Free, open-source video coding tool



Weaknesses

- ▶ Data curation hard, slow
- ▶ Fixing the past vs. fixing the future
- ▶ Sustainability model undetermined
- ▶ Small curation, development staff
- ▶ Datavyu user base small
 - ◀ How to compete/cooperate with commercial entities?



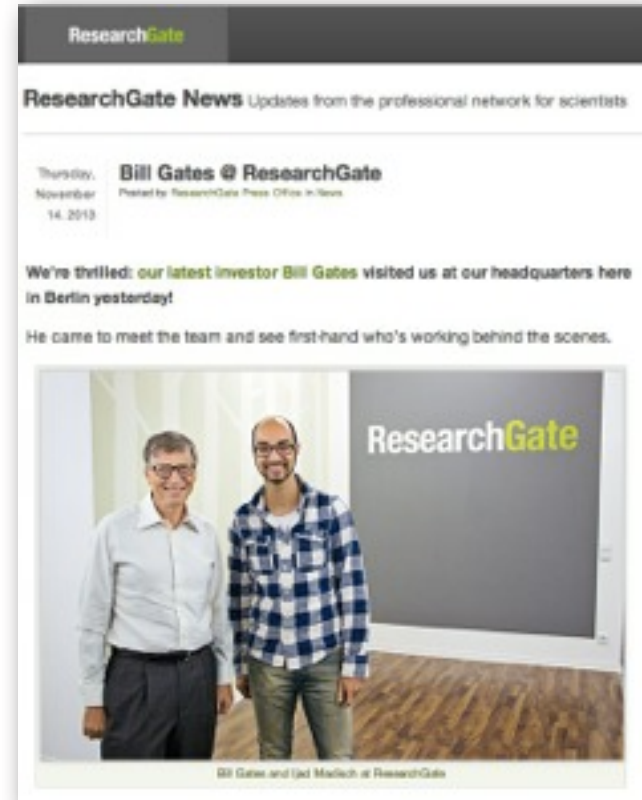
Your thoughts?

- ▶ Strengths we neglected or overstated?
- ▶ Weaknesses we neglected or overstated?

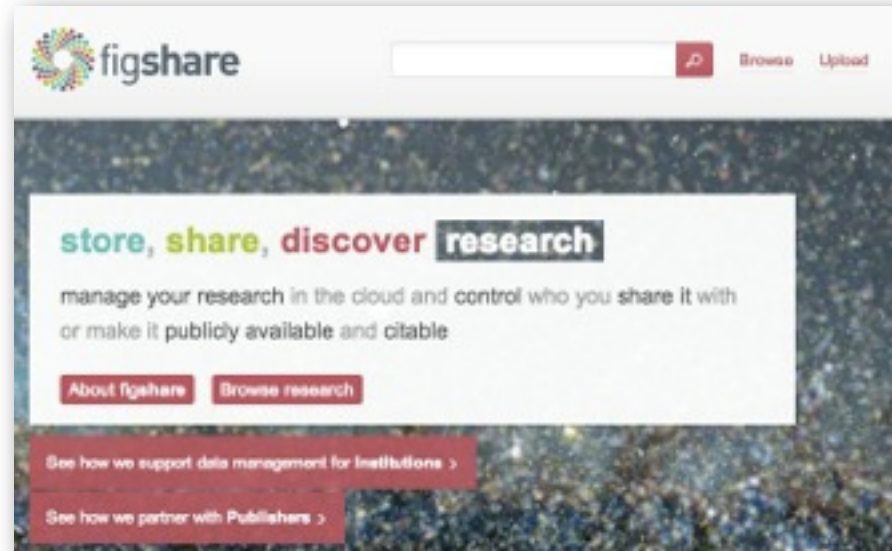


Threats

- ▶ Open science space moving quickly
- ▶ Attracting venture capital



Data-sharing



GitHub



box



Your stuff, anywhere

[Sign up](#)

or [Sign in](#)



Google Drive



Open Science Framework_{BETA}



ICPSR

A partner in social science research

The
**Dataverse
Network**TM
Project

A graphic element to the right of the text, consisting of three interconnected circles of varying sizes, with the largest circle at the bottom.

Academic Torrents 🔍 Browse ⬆️ Upload ⓘ About Login **Signup**

Welcome to Academic Torrents!
Currently making 6.93TB of research data available.

PENNSTATE
 SCHOLARSphere

SEARCH

Browse By What is ScholarSphere?



Threats

- ▶ Databrary software not yet released
- ▶ How can Databrary compete?
- ▶ Why share with Databrary vs. alternatives?
- ▶ What features will attract, keep users?
- ▶ What features will ensure sustainability, change the culture of developmental science?



Your turn

- ▶ More threats?
- ▶ Different spin on them?



Opportunities

- ▶ Potential for partnering
 - ◀ TalkBank/CHILDES
 - ◀ Other sharing organizations
 - Research Data Alliance (RDA)
 - Center for Open Science (COS)
 - Stanford Meta-Research Center
 - ◀ Societies
 - Report on developmental society meeting



Opportunities

- ▶ Potential for partnering
 - ◀ Journals
 - Store supplemental data?
 - Store extended materials for all submissions?
 - Which fields?
 - Developmental science only
 - Psychology
 - Neuroscience



Opportunities

- ▶ Potential for partnering
 - ◀ Conferences (both society-affiliated and not)
 - Store, share talk recordings?
 - Store supplemental data?
 - ◀ Sustainability component
 - Member or attendee fees-for-service?



Opportunities

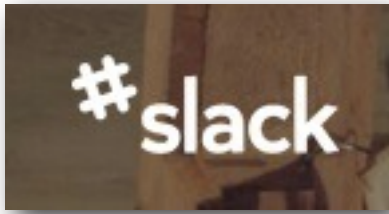
- ▶ Emphasize competitive advantages
 - ◀ Domain-focus, visibility
 - Your colleagues are here
 - ◀ Policies to share identifiable research data, especially recordings
 - ◀ University & Federal support
 - ◀ Stability and long-term sustainability
 - ◀ Sharing is our business model



Your thoughts?

- ▶ Partnerships
- ▶ Journals
- ▶ Conferences
- ▶ Sharpen our message
- ▶ Other ideas?





Opportunities

- ▶ Integration with external services
 - ◀ Cloud storage/data sharing
- ▶ Offer new services via APIs
 - ◀ Lab management
 - ◀ Project/task management
 - ◀ Data visualization
 - ◀ Social media, communications
 - ◀ Electronic CV, profile management



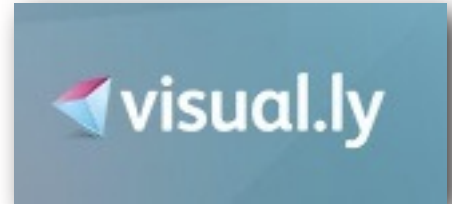
Your thoughts?

- ▶ Do you use any of these services?
- ▶ Is providing data sharing services enough?
- ▶ Does integration with other tools help or hinder our goals?



Opportunities

- ▶ Beyond storing, tagging video
 - ◀ Visualization
 - ◀ Additional analysis streams



Video **A**nnotation **T**ool from **I**rvine, **C**alifornia



Meeting Agenda

- 1:00-1:15 Welcome and recent accomplishments
- 1:15-1:45 Roadmap
- 1:45-2:15 Policies
- 2:15-2:30 Break
- 2:30-3:30 Beta: Upload and search
- 3:30-4:00 Researcher asks and support
- 4:00-4:10 Break
- 4:10-5:00 Positioning Databrary for the future
- 5:00-5:30 Wrap up and discussion



Thank you