Advances in Engineering Education

SPRING 2016

Transforming Education Research Through Open Video Data Sharing

RICK O. GILMORE
The Pennsylvania State University
University Park, PA

KAREN E. ADOLPH

DAVID S. MILLMAN

AND

ANDREW GORDON
New York University
New York, NY

ABSTRACT

Open data sharing promises to accelerate the pace of discovery in the developmental and learning sciences, but significant technical, policy, and cultural barriers have limited its adoption. As a result, most research on learning and development remains shrouded in a culture of isolation. Data sharing is the rare exception (Gilmore, 2016). Many researchers who study teaching and learning in classroom, laboratory, museum, and home contexts use video as a primary source of raw research data. Unlike other measures, video captures the complexity, richness, and diversity of behavior. Moreover, because video is self-documenting, it presents significant potential for reuse. However, the potential for reuse goes largely unrealized because videos are rarely shared. Research videos contain information about participants' identities making the materials challenging to share. The large size of video files, diversity of formats, and incompatible software tools pose technical challenges. The Databrary (databrary.org) digital library enables researchers who study learning and development to store, share, stream, and annotate videos. In this article, we describe how Databrary has overcome barriers to sharing research videos and associated data and metadata. Databrary has developed solutions for respecting participants' privacy; for storing, streaming, and sharing videos; and for managing videos and associated metadata. The Databrary experience suggests ways that videos and other identifiable data collected in the context of educational research might be shared. Open data sharing enabled by Databrary can serve as a catalyst for a truly multidisciplinary science of learning.

Key words: open data sharing, video, digital library

INTRODUCTION

Open data sharing has become a scientific imperative for research funding from the government (National Science Board, 2011) and foundations (Gates Foundation, n.d.). Open data sharing is common practice in many areas of biomedical (Kaye, 2009), physical (Young, 2010), biological (Reichman, Jones, & Schildhauer, 2011) and earth sciences (Kleiner, 2011), and it is an emerging priority in areas such as neuroscience (Poldrack & Gorgolewski, 2014; Poline et al., 2012). Open data sharing can help to translate insights from scientific research into applications that serve essential human needs. Open data sharing bolsters transparency and peer oversight (Miguel et al., 2014), encourages diversity of analysis and opinion, accelerates the education of new researchers, and stimulates the exploration of new topics not envisioned by the original investigators. Data sharing and reuse increase the impact of public and private investments in research and promise to lead to more effective public policy.

The value of open data sharing has been known for decades (Perneger, 2011), but a vigorous open science movement has gained momentum only in recent years. Open access journals now address general scientific topics (e.g., PeerJ, PLOS One, and GigaScience), specific domains (e.g., Advances in Engineering Education and BMJ Open), and even raw data (e.g., Nature Scientific Data). Data repositories from diverse areas of scholarly inquiry have begun to communicate with one another through groups such as the Research Data Alliance and work jointly to achieve common goals such as publishing author guidelines for transparent and open research (Nosek et al., 2015). New web-based platforms now enable open sharing of slides from talks (SlideShare), data, and materials (FigShare, Open Science Framework, and F1000 Research). The emergence of greater interest in open data sharing has coincided with the increasing popularity of web-based software sharing and version control systems such as GitHub, open source data analysis languages such as R and Python, and with the initiation of training programs such as Software Carpentry that help working scientists acquire state-of-the-art data management skills. These initiatives seek to speed research progress, improve research practice, and foster greater scientific transparency.

Unfortunately, despite efforts to make data sharing a norm in the social and educational sciences (AERA Code of Ethics, 2011; Nosek & Bar-Anon, 2011) most research on human learning and development remains shrouded in a culture of isolation (Adolph, Gilmore, Freeman, Sanderson, & Millman, 2012; Gilmore, 2016). Researchers share interpretations of distilled, not raw data, almost exclusively through publications and presentations. The path from raw data to research findings to conclusions cannot be traced or validated by others. Other researchers cannot pose new questions that build on the same raw materials.

This paper describes how the Databrary data library has overcome barriers to sharing *video*, a type of raw data collected by thousands of researchers who study learning and development in

classroom, museum, laboratory, and home contexts (Gilmore & Adolph, 2016). Video is a medium with unique potential for reuse by others, and this makes it critical to overcome barriers to open video data sharing. The Databrary project illustrates how open sharing of video data can improve scientific practice and advance research on learning and development.

THE PROMISE AND CHALLENGE OF VIDEO

More than 1 billion users worldwide upload more than 300 hours of video to YouTube every minute. Google, the owner of YouTube, has built an immense data infrastructure to upload, store, convert, tag, and stream video. The current scale of video collection in research contexts is much smaller of course and more focused in purpose, but it is large and growing. For example, the Measures of Effective Teaching (MET) Project funded by the Gates Foundation generated more than 20,000 videotaped lessons from more than 3,000 K-12 teachers in 6 urban school districts, covering core subjects such as mathematics and language arts from multiple camera angles. The data, constituting tens of terabytes of storage, are hosted at the University of Michigan and streamed to registered viewers across the country. Beyond research on classroom practices, video recordings are becoming critical for teacher observation, evaluation and certification through tools such as Edthena and the National Board for Professional Teaching Standards' ATLAS initiative

Video data are unique because video captures the richness and complexity of human behavior—the interactions between people and their environments. Video captures what people say and what they do. It captures when, where, and how they look, gesture, and move, and how their looking, gesturing, and moving correspond to what others are doing. Because of these characteristics, video—and before that, film—has a long history in the study of learning and development (Goldman et al., 2014; Curtis, 2011). Educators, teacher trainers, and evaluators use video to record what teachers do to help improve teaching practices or to evaluate performance (Bakken & Pierroux, 2015; Blomberg et al., 2013; Masats & Dooly, 2011; Baecher, Kung, Jewkes, & Rosalia, 2013). Researchers use video in home and laboratory contexts to study how infants, children, and adults behave in natural or experimenter-imposed tasks (Karasik, Tamis-LeMonda, & Adolph, 2014). Researchers record videos of students in classrooms (Alibali & Nathan, 2012) to understand what teachers do and how students respond. Because video closely mimics the multisensory experiences of live human observers, recordings collected by one person for a particular purpose may be readily understood by another person and reused for a different purpose. Moreover, the success of YouTube and other video-based social media demonstrates that web-based video storage and streaming systems are now sufficiently well developed to satisfy large-scale demand.

Thus, the question for researchers and policymakers is how to capitalize on the unique potential of large-scale research video to improve teaching and learning. The answer requires overcoming technical, ethical, practical, and cultural challenges to sharing research video.

File sizes and diverse formats present special technical challenges for sharing. Video files are large (one hour of HD video can consume 10+ GB of storage) and come in varied formats (from cell phones to high-speed video). Many studies require multiple camera views to capture desired behaviors. Thus, sharing video requires substantial storage capacity and significant computational resources for transcoding videos into common formats that can be preserved over the long term.

Technical challenges involved in searching the contents of videos present barriers to sharing. Videos contain rich and diverse information that requires significant effort by human observers to extract. Researchers make use of videos by watching them and, using paper and pencil or more automated computerized coding software, translating observations into ideas and numbers. In many cases, researchers assign codes to particular portions of videos. In principle, these codes make the contents of videos searchable by others. However, researchers focus on different questions from varied theoretical perspectives and lack consensus on conceptual ontologies. So, in practice, most coded data are not easily shared. Although human-centered video coding capitalizes on the unique abilities of trained observers to capture important dimensions of behavior, machine learning and computer vision tools may provide new avenues for tagging the contents of videos for educational and developmental research (Amso, Haas, Tenenbaum, Markant, & Sheinkopf, 2014; Yu & Smith, 2013; Fathi, Hodgins, & Rehg, 2012; Google Research, 2014; Raudies & Gilmore, 2014).

Open video sharing must overcome *ethical challenges* linked to sharing personally identifiable data. Although policies exist for sharing de-identified data, video contains easily identifiable data—faces, voices, names, interiors of homes and classrooms, and so on. Removing identifiable information from video severely diminishes its value for reuse and puts additional burdens on researchers. So, open video sharing requires new policies that protect the privacy of research participants while preserving the integrity of raw video for reuse by others.

Open video sharing presents practical challenges of data management. Developmental and educational research is inundated by data, most of which is inaccessible to other researchers (Gilmore, 2016). Research video creates a data explosion: A typical lab studying infant or child development collects 5-10 hours of video/week (Gilmore & Adolph, 2012; 2016). Researchers lack time to find, label, clean, organize, and copy their files into formats that can be used and understood by others (Ascoli, 2006a). Study designs vary widely, and no two labs manage data in the same way. Idiosyncratic terminology, record-keeping, and data management practices are the norm. Few researchers document workflows or data provenance. Although video requires minimal metadata to be useful to

others, video files must be electronically linked to what relevant metadata exist including information about whether participants have given permission to share.

Perhaps the most important challenge is cultural. Community practices must change. Most researchers in the education, learning, and developmental sciences do not reuse their own videos or videos collected by other researchers; they neither recognize nor endorse the value of open sharing. Contributing data is anathema and justifications against sharing are many. Researchers cite intellectual property and privacy issues, the lack of data sharing requirements from funding agencies, and fears about the misuse, misinterpretation, or professional harm that might come from sharing (Ascoli, 2006b; Ferguson, 2014). Data sharing diverts energy and resources from scholarly activities that are more heavily and frequently rewarded. These barriers must be overcome to make video data sharing a scientific norm.

DATABRARY.ORG

The Databrary project has built a digital data library (http://databrary.org) specialized for open sharing of research videos. Databrary has overcome critical barriers to sharing video, including solutions for respecting participants' privacy; for storing, streaming, and sharing video; and for managing video datasets and associated metadata (Gordon, Millman, Steiger, Adolph, & Gilmore, 2015). Databrary's technology and policies lay the groundwork for securely sharing research videos on teaching and learning. In just over two years of operation, Databrary has collected thousands of individual videos—4,400+ hours of recordings—featuring more than 4,200 infant, child, and adult participants. More than 520 authorized researchers and affiliates representing more than 200 institutions across the globe have been granted access to the library. Video data are big data, and the interest in recording and sharing video for research, education, and policy purposes continues to grow.

The Databrary project (databrary.org) arose to meet the challenges of sharing research video and to deliver on the promise of open data sharing in educational and developmental science. With funding from NSF (BCS-1238599) and NIH (NICHD U01-HD-076595), Databrary has focused on building a data library specialized for video, creating data management tools, crafting new policies that enable video sharing, and fostering a community of researchers who embrace video sharing. Databrary also developed a free, open-source video coding tool, Datavyu. The project received funding in 2012-2013, began a private beta testing phase in the spring of 2014, and opened for public use in October of 2014.

System Design

The Databrary system enables large numbers of video and related files to be uploaded, converted, organized, stored, streamed, and tagged. Databrary is a free, open source web application. The data

are preserved indefinitely in a secure storage facility at New York University. Databrary can house video and audio files along with associated materials, coding spreadsheets, and metadata. Video and audio data are transcoded into standard and HTML5-compatible formats. This ensures that video data can be streamed and downloaded by any operating system that supports a modern browser. Copies of original video files are also stored. Databrary stores other data in their original formats (e.g., .doc, .docx, .xls, .xlsx, .txt, .csv, .pdf, .jpg, .png).

The system's data model embodies flexibility. Researchers organize their materials by acquisition date and time into structures called *sessions*. A session corresponds to a unique recording episode featuring specific participants. It contains one or more videos and other file types and may be linked to user-defined metadata about the participants, tasks or measures, and locations. A group of sessions is called a *volume*. Databrary contributors may combine sessions or segments with illustrative video excerpts and photos, coding manuals and coding spreadsheets, questionnaires and other sources of data, IRB documents, computer code, sample displays, statistical analyses, and links to published journal articles. The volume represents one view of a set of data and materials that may correspond to a particular published product or a set of published products. Figures 1 and 2 show screenshots from an illustrative volume on the site. Figure 3 shows how individual sessions and their associated metadata are depicted.

Databrary does not enforce strict ontologies for tagging volumes, sessions, or the contents of videos. Video data are so rich and complex that in many domains, researchers have not settled on standard definitions for particular behaviors and may have little current need for standardized tasks, procedures, or terminology. Indeed, standardized ontologies are not necessary for many use cases. The system allows users to upload coding manuals with rich, text- or image-based definitions of codes. Databrary also empowers users to add keyword tags and to select terms that have been suggested by others without being confined to the suggestions. Moreover, Databrary encourages user communities within Databrary to converge on shared vocabulary based on the most common keyword tags, and to construct and enforce common procedures and tasks wherever this makes sense. Figure 4 shows a screenshot from the Databrary splash page that shows a recent activity feed and a set of clickable tags ranked by popularity of current use that select sets of volumes linked to the tagged words.

A future challenge is to enhance the capacity to search for tagged segments inside of videos. Volume-level search functionality exists in the current software, with more extensive capabilities on the near horizon. A related challenge involves importing files from the most popular desktop video coding tools used by investigators to annotate recordings (Gilmore & Adolph, 2016). Databrary already allows coding files from the Datavyu tool to be uploaded and shared. Future enhancements will allow the codes and definitions contained in these files to be imported, indexed, and added to

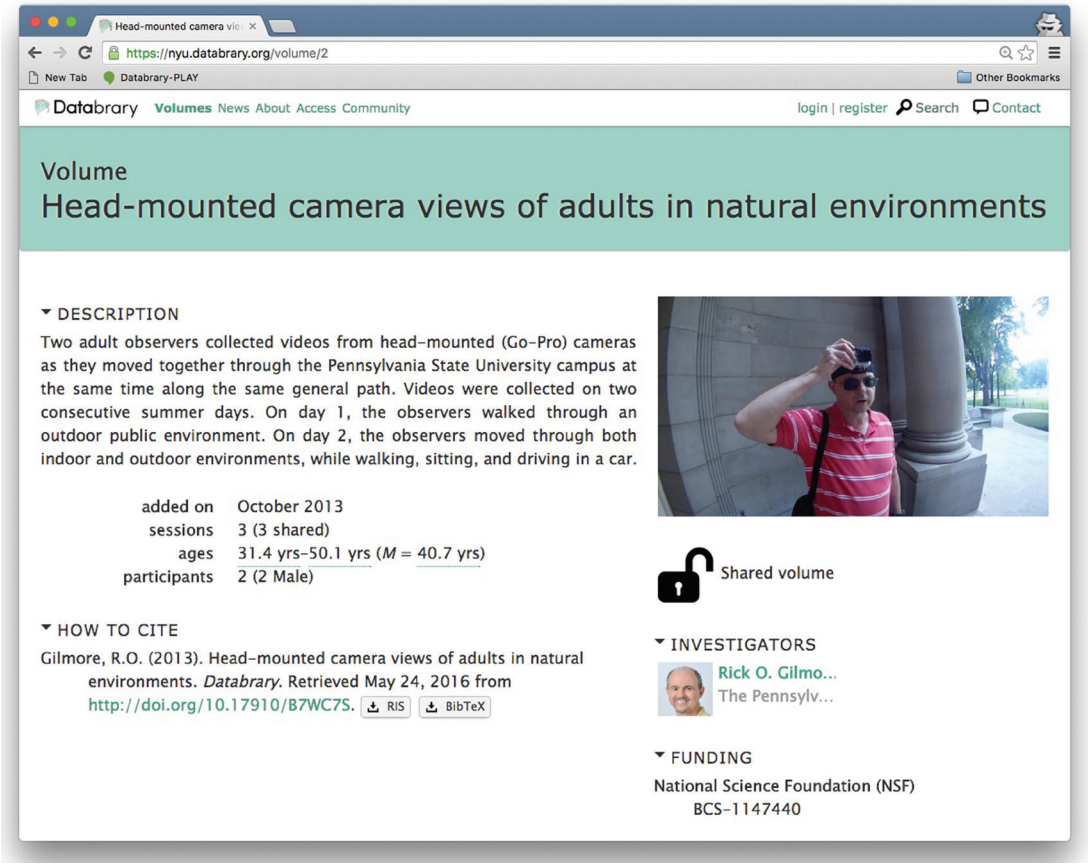


Figure 1. Example of a Databrary volume. Data and materials from a research project are organized into a volume that contains a description of the study, information about the participants, a standard citation, and a list of funding sources.

Databrary's search engine. This will allow users to search for specific tags, read text-based descriptions of the codes, and visualize temporal relationships independent of the desktop software deployed in a particular project. These features should reduce the barriers that can exist in the interpretation and reuse of some forms of qualitative data. We envision a parallel set of export functions that permit full interoperability among coding tools. The priority will be to create interoperability with tools using open, not proprietary file formats. Databrary also recognizes the need to develop open standards and interfaces that enable Databrary to link to and synchronize with outside sources that specialize in other data types.

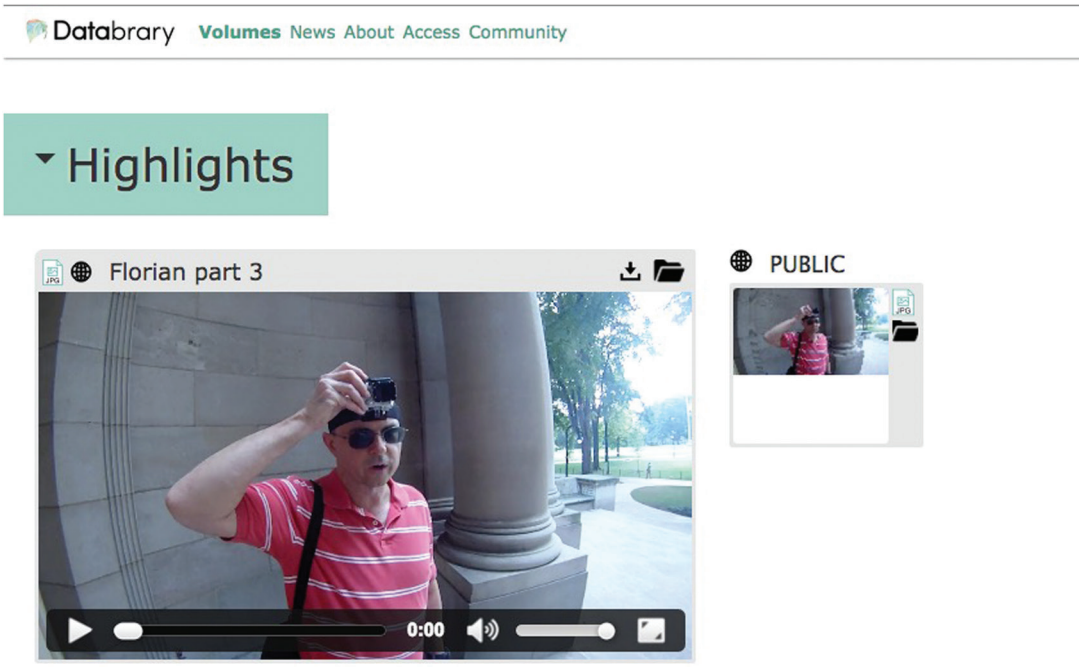
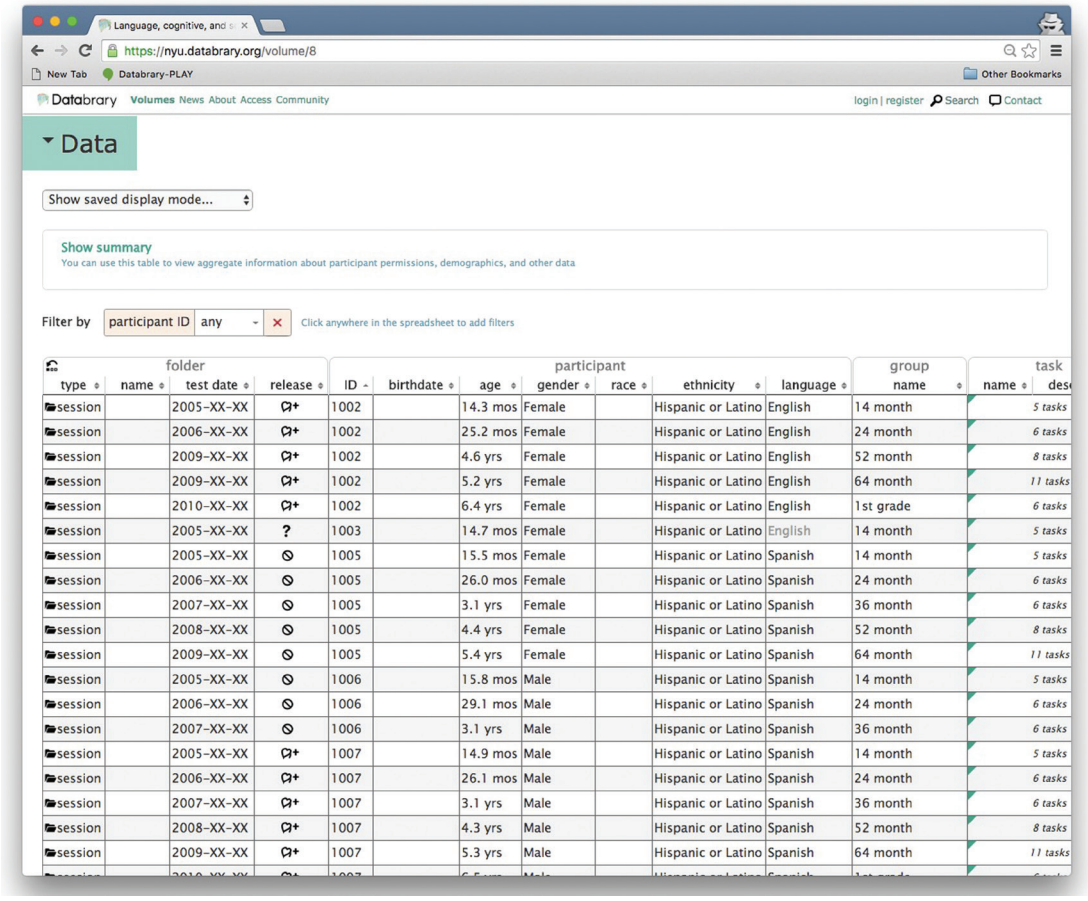


Figure 2. Volume highlights and data. Each Databrary volume depicts video or image highlights, study-wide materials, and data. If the participant has given permission, highlights may be shown for educational or research purposes.

Policies for Safe and Secure Video Sharing

Policies for openly sharing identifiable data in ways that securely preserve participant privacy are essential for sharing research video. Databrary does not attempt to de-identify videos. Instead, Databrary maximizes the potential for video reuse by keeping recordings in their original unaltered form. To make unaltered raw videos available to others for reuse, Databrary has developed a two-pronged access model that (a) restricts access to authorized researchers, and (b) enables access to identifiable data only with the explicit permission of the participants.

To gain access to Databrary, the researcher must register on the site. Applicants for "authorized researcher" status agree to uphold Databrary's ethical principles and to follow accepted practices concerning the responsible use of sensitive data. Each applicant's institution must co-sign an access agreement. Full privileges are granted only to those applicants with independent researcher status at their institutions. Others may be granted privileges if they are affiliated with a researcher who agrees to sponsor their application and to supervise their use. Ethics board or IRB approval is not required to gain access to Databrary because many use cases do not involve



folder				participant							group	task	
type	name	test date	release	ID	birthdate	age	gender	race	ethnicity	language	name	name	des
session		2005-XX-XX	CC+	1002		14.3 mos	Female		Hispanic or Latino	English	14 month	5 tasks	
session		2006-XX-XX	CC+	1002		25.2 mos	Female		Hispanic or Latino	English	24 month	6 tasks	
session		2009-XX-XX	CC+	1002		4.6 yrs	Female		Hispanic or Latino	English	52 month	8 tasks	
session		2009-XX-XX	CC+	1002		5.2 yrs	Female		Hispanic or Latino	English	64 month	11 tasks	
session		2010-XX-XX	CC+	1002		6.4 yrs	Female		Hispanic or Latino	English	1st grade	6 tasks	
session		2005-XX-XX	?	1003		14.7 mos	Female		Hispanic or Latino	English	14 month	5 tasks	
session		2005-XX-XX	⊘	1005		15.5 mos	Female		Hispanic or Latino	Spanish	14 month	5 tasks	
session		2006-XX-XX	⊘	1005		26.0 mos	Female		Hispanic or Latino	Spanish	24 month	6 tasks	
session		2007-XX-XX	⊘	1005		3.1 yrs	Female		Hispanic or Latino	Spanish	36 month	6 tasks	
session		2008-XX-XX	⊘	1005		4.4 yrs	Female		Hispanic or Latino	Spanish	52 month	8 tasks	
session		2009-XX-XX	⊘	1005		5.4 yrs	Female		Hispanic or Latino	Spanish	64 month	11 tasks	
session		2005-XX-XX	⊘	1006		15.8 mos	Male		Hispanic or Latino	Spanish	14 month	5 tasks	
session		2006-XX-XX	⊘	1006		29.1 mos	Male		Hispanic or Latino	Spanish	24 month	6 tasks	
session		2007-XX-XX	⊘	1006		3.1 yrs	Male		Hispanic or Latino	Spanish	36 month	6 tasks	
session		2005-XX-XX	CC+	1007		14.9 mos	Male		Hispanic or Latino	Spanish	14 month	5 tasks	
session		2006-XX-XX	CC+	1007		26.1 mos	Male		Hispanic or Latino	Spanish	24 month	6 tasks	
session		2007-XX-XX	CC+	1007		3.1 yrs	Male		Hispanic or Latino	Spanish	36 month	6 tasks	
session		2008-XX-XX	CC+	1007		4.3 yrs	Male		Hispanic or Latino	Spanish	52 month	8 tasks	
session		2009-XX-XX	CC+	1007		5.3 yrs	Male		Hispanic or Latino	Spanish	64 month	11 tasks	

Figure 3. Spreadsheet interface to raw video data and session metadata. Each volume presents a spreadsheet-like interface to depict and enable browsing of session-level metadata. The session-level data fields can be customized by data contributors to suit a specific study. Importantly, the spreadsheet indicates what level of release has been granted to a given session. These release levels govern what materials may be shared and with whom.

research. IRB approval *is* required to contribute data and for research uses. Once authorized, a researcher has full access to shared data on the site, and may browse, tag, download for later viewing, and conduct non-research or pre-research activities. These policies are spelled out fully in an online user guide.

Unique among data repositories, the Databrary access agreement authorizes both data use and contribution. However, researchers agree to store on Databrary only materials for which

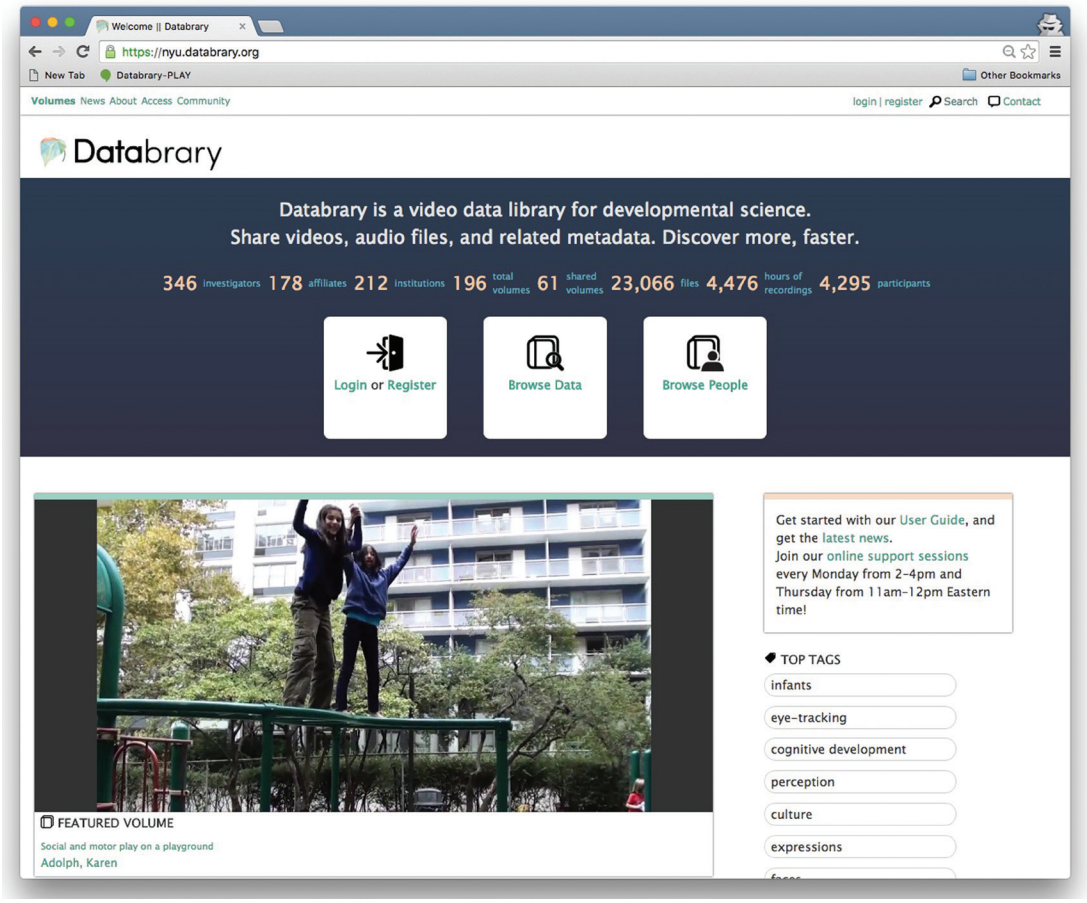


Figure 4. Databrary home page. The Databrary splash or home page depicts a featured data set (volume), recent activity feed, links to browse data or people, and a list of clickable tags that link to studies containing these specific tag words.

they have ethics board or IRB approval. Data may be stored on Databrary for the contributing researcher's use regardless of whether the records are shared with others or not. When a researcher chooses to share, Databrary makes the data openly available to the community of authorized researchers.

In addition to restricting access to authorized researchers, Databrary has extended the principle of informed consent to participate in research to encompass permission to share data with other researchers. To formalize the process of acquiring permission, Databrary has developed a Participant Release Template with standard language recommended for use with study participants. This language helps participants to understand what is involved in sharing video data, with whom the

data will be shared, and the potential risks of releasing video and other identifiable data to other researchers.

Databrary's access model arose from several inspirations. Many researchers who collect video, audio, or photographs currently seek permission from participants or parents to use images for talks, websites, publications, or teaching. Participants grant or deny permission through a separate release form. Databrary deliberately built the library's sharing release language and policies upon this existing, well-established foundation because it is familiar to researchers and ethics boards. The Gates Foundation-funded Methods of Effective Teaching (MET) Project served as another inspiration. Databrary acquired the MET data contribution and access agreements from our colleagues at the University of Michigan, and the team adapted the agreements, making them both broader and more general in scope. Finally, Databrary took inspiration from other big data projects that specifically seek participant permission to share potentially identifiable data: The Human Connectome Project, the Personal Genome Project, and the Open Humans Project.

Managing Data for Sharing

When researchers *do* share, standard practice involves organizing data after a project is finished, perhaps when a paper goes to press. This "preparing for sharing" after the fact presents a difficult, time-consuming, and unrewarding chore for investigators. It makes curating and ingesting datasets challenging for repositories, as well. Databrary has created a different route to curation that reduces these barriers.

Databrary has developed a data management system that empowers researchers to upload and organize data as it is collected. Immediate uploading reduces the workload on investigators, minimizes the risk of data loss and corruption, and accelerates the speed with which materials become openly available. The system employs familiar, easy-to-use spreadsheet and timeline-based interfaces (Figures 2–3, and 5) that allow users to upload videos, add metadata about tasks, settings, and participants, link related files, and assign appropriate permission levels for sharing. To encourage immediate uploading, Databrary provides a complete set of controls so that researchers can restrict access to their own labs or to other users of their choosing. Datasets can be openly shared with the broader research community at a later point when data collection and ancillary materials are complete, whenever the contributor is comfortable sharing, or when journals or funders require it.

Building a Community

Data sharing works only when the scientific community embraces it. From the beginning, Databrary has sought to cultivate a community of researchers who support data sharing and commit to

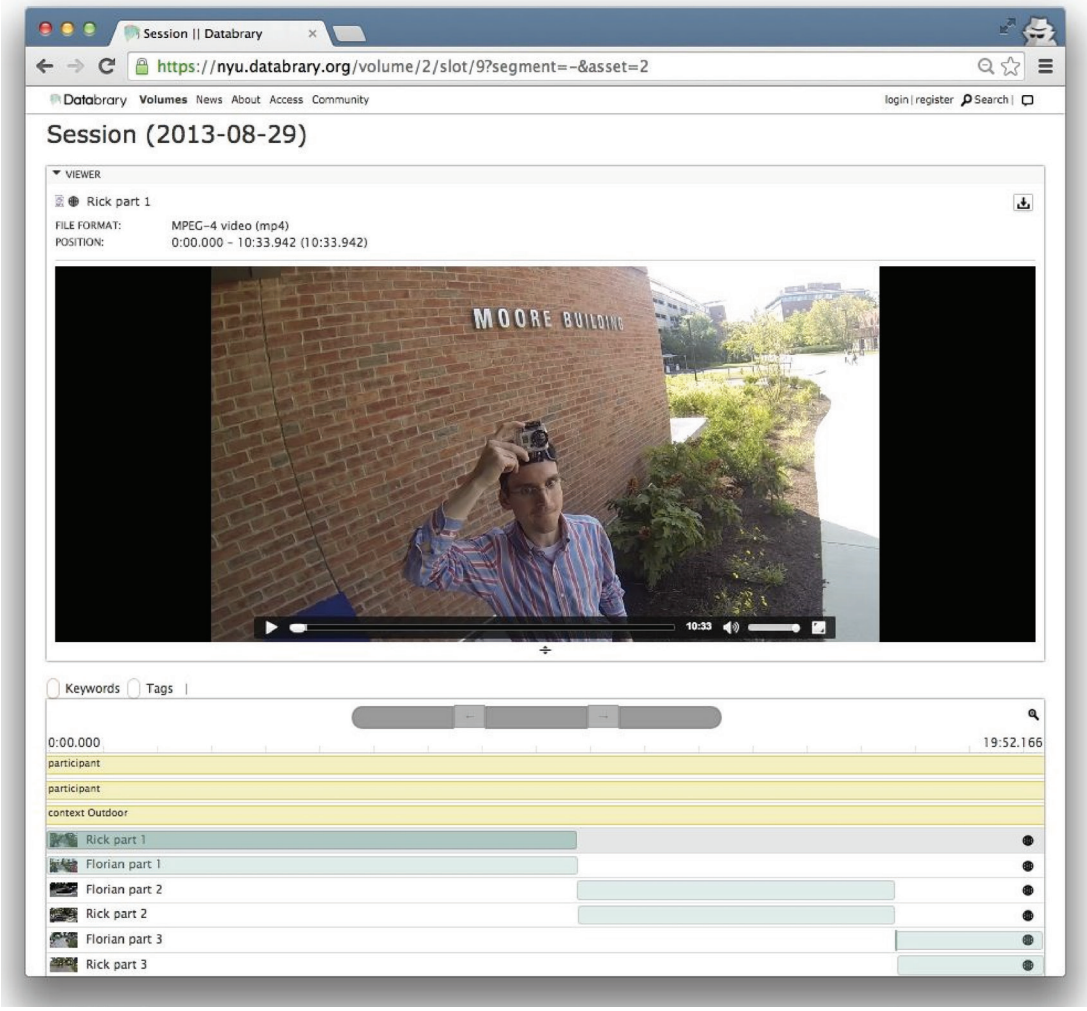


Figure 5. Timeline representation of a session. Each session is represented using a timeline that depicts the temporal relations among multiple, temporally overlapping data streams. The timeline interface permits viewers to add text-based keywords and tags and to view videos within the browser.

making sharing and reuse part of their normal practices. Databrary's community-building efforts involve many interacting components. These include active engagement with professional associations, conference-based exhibits and training workshops, regional workshops, communications with research ethics and administration staff, talks and presentations to diverse audiences, and one-on-one consultations with individual researchers and research teams. These activities are time consuming and labor intensive, but they are critical to changing community attitudes toward data

sharing in the educational and learning sciences. Databrary's rapid growth—the number of authorized investigators has doubled in the last six months—suggests that community engagement, while difficult, pays dividends. Looking ahead, it will be critical to engage funders, journals, and professional organizations in the effort to forge community consensus about the importance, feasibility, and potential of open video data sharing.

As it stands now, Databrary serves the community of researchers who study infant, child, and adolescent development, and the library is beginning to serve the needs of researchers who study K-12 education. However, the software is free and open source and could readily be adapted for use by investigators interested in studying and sharing findings from research into best practices in engineering education or other fields.

CONCLUSION

Imagine a time in the near future when researchers interested in studying how best to train future scientists and engineers can mine an integrated, synchronized, interoperable, open, and widely shared dataset. The components include video from multiple camera angles, eye tracking, motion, and physiological measurements, and information about historical and real-time student performance. Imagine that this classroom-level data can be linked with grade, school, neighborhood, community, region, and state-level data about education practice, curriculum, and policy. Then, imagine training a cadre of experts with skills in the data science of learning and education who are sensitive to privacy, confidentiality, and ethical issues related to research involving identifiable information. We empower these learning scientists to extract from the data meaningful insights about how educational practice and policy might be improved. In short, imagine a science of teaching and learning that can be personally tailored to individuals in ways analogous to the impact of big data on medicine. The barriers to realizing this vision are similar to those that confront the vision of personalized medicine—the development of technologies that enable data to be collected, synchronized, tagged, curated, stored, shared, linked, and aggregated; policies and practices that ensure security and individual privacy; and the cultivation of professional expertise needed to turn raw data into actionable insights.

As Gesell once noted, cameras can record behavior in ways that make it "...as tangible as tissue" (Scott, 2011). The Databrary team contends that video has a central role to play in efforts to make tangible the anatomy of successful teaching and learning. In fact, we argue that video can be the core around which other measures of teaching and learning cluster. This requires reducing barriers to sharing video and fostering new community values around data sharing that make it

indispensible and routine. The Databrary project has built technology and policies that overcome many of the barriers to widespread sharing within the developmental sciences community. Databrary suggests ways that video and other identifiable data collected in the context of education research might also be shared. Technologies and policies for providing secure access to videos for broader use cases need to be developed, tools that allow desktop coding software files to be seamlessly converted to and from one another will have to be perfected, and ways of synchronizing and linking disparate data streams will have to be created. Equally important, communities of scholars dedicated to collecting, sharing, and mining education-related videos will have to be cultivated. But, we believe that the widespread sharing of high value, high impact data of the sort that video can provide promises to achieve this ambitious vision to advance education policy and improve practice. Databrary is working toward a future where open video data sharing is the norm, a personalized science of teaching and learning is the goal, and what optimizes student learning is as tangible as tissue.

ACKNOWLEDGEMENTS

Databrary is based on work supported by the National Science Foundation under Grant No. BCS-1238599, the Eunice Kennedy Shriver National Institute of Child Health and Human Development under Cooperative Agreement U01-HD-076595, and the Society for Research in Child Development. Any opinions, findings, conclusions and recommendations expressed in the material contributed here are those of the authors and do not necessarily reflect the views of the National Science Foundation, the Eunice Kennedy Shriver National Institute of Child Health and Human Development, or the Society for Research in Child Development.

REFERENCES

Adolph, K.E., Gilmore, R.O., Freeman, C., Sanderson, P., & Millman, D. (2012). Toward open behavioral science. *Psychological Inquiry*, *23*(3), 244–247. http://doi.org/10.1080/1047840X.2012.705133.

Advances in Engineering Education. (n.d.). Retrieved 2016-02-07 from http://advances.asee.org.

AERA Code of Ethics: American Educational Research Association Approved by the AERA Council February 2011. (2011). *Educational Researcher*, *40*(3), 145–156. http://doi.org/10.3102/0013189X11410403.

Alibali, M.W., & Nathan, M.J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, *21*(2), 247-286. http://doi.org/10.1080/10508406.2011.611446.

Amso, D., Haas, S., Tenenbaum, E., Markant, J., & Sheinkopf, S. (2014). Bottom-up attention orienting in young children with autism. *Journal of Autism and Developmental Disorders*, *44*(3), 664-673. http://doi.org/10.1007/s10803-013-1925-5

Ascoli, G. A. (2006a). Mobilizing the base of neuroscience data: the case of neuronal morphologies. *Nature Reviews Neuroscience*, *7*(4), 318–324. http://doi.org/10.1038/nrn1885.

Ascoli, G. A. (2006b). The ups and downs of neuroscience shares. *Neuroinformatics*, *4*(3), 213–215. http://doi.org/10.1385/NI:4:3:213.

ATLAS. (n.d.). Retrieved 2016-02-07 from http://www.nbpts.org/atlas.

Baecher, L., Kung, S.-C., Jewkes, A.M., & Rosalia, C. (2013). The role of video for self-evaluation in early field experiences. *Teaching and Teacher Education*, *36*, 189–197. http://doi.org/10.1016/j.tate.2013.08.001.

Bakken, S.M., & Pierroux, P. (2015). Framing a topic: Mobile video tasks in museum learning. *Learning, Culture and Social Interaction*. http://doi.org/10.1016/j.lcsi.2014.12.001.

Blomberg, G., Sherin, M.G., Renkl, A., Glogger, I., & Seidel, T. (2013). Understanding video as a tool for teacher education: investigating instructional strategies to promote reflection. *Instructional Science*, *42*(3), 443–463. http://doi.org/10.1007/s11251-013-9281-6.

British Medical Journal Open. (n.d.). Retrieved 2016-02-07 from http://bmjopen.bmj.com/.

Databrary Access Agreement (n.d.). Retrieved 2016-02-07 from https://databrary.org/access/policies/agreement.html.

Databrary Access Policies (n.d.). Retrieved 2016-02-07 from https://databrary.org/access/policies.html.

Datavyu. (n.d.). Retrieved 2016-02-07 from http://datavyu.org.

Edthena. (n.d.). Retrieved 2016-02-07 from http://www.edthena.com.

F1000 Research. (n.d.). Retrieved 2016-02-07 from http://f1000research.com.

Fathi, A., Hodgins, J., & Rehg, J. (2012, June). Social interactions: A first-person perspective. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (p. 1226-1233). http://doi.org/10.1109/CVPR.2012.6247805.

Ferguson, L. (2014). *How and why researchers share data (and why they don't).* Retrieved from http://bit.ly/1A5mmEW.

FigShare. (n.d.). Retrieved 2016-02-07 from http://figshare.com.

Gates Foundation. (n.d.). Retrieved 2016-02-07 from http://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy.

GigaScience. (n.d.). Retrieved 2016-02-07 from http://www.gigasciencejournal.com.

Gilmore, R.O. (2016). From big data to deep insight in developmental science. *Wiley Interdisciplinary Reviews Cognitive Science, 7*(2), 112-126. http://doi.org/10.1002/wcs.1379.

Gilmore, R.O., & Adolph, K. E. (2012). *Video Use Survey of ICIS and CDS listserv members.*

Gilmore, R.O., & Adolph, K. E. (2016). *Video Use Survey of ICIS and CDS listserv members, Datavyu and Databrary users.*

GitHub. (n.d.). Retrieved 2016-02-07 from http://github.com.

Goldman, R., Pea, R., Barron, B., & Derry, S.J. (2014). *Video Research in the Learning Sciences*. Routledge.

Google Research. (2014). *A picture is worth a thousand (coherent) words: Building a natural description of images.* Retrieved 2015-05-08 from http://bit.ly/1wTMbk7.

Gordon, A., Millman, D.S., Steiger, L., Adolph, K.E., & Gilmore, R.O. (2015). Researcher-library collaborations: Data repositories as a service for researchers. *Journal of Librarianship and Scholarly Communication*, 3(2), http://doi.org/10.7710/2162-3309.1238.

Human Connectome Project (n.d.). Retrieved 2016-02-07 from http://www.humanconnectomeproject.org/.

Karasik, L.B., Tamis-LeMonda, C.S., & Adolph, K.E. (2014). Crawling and walking infants elicit different verbal responses from mothers. *Developmental Science*, *17*(3), 388–395. http://doi.org/10.1111/desc.12129.

Kaye, J., Heeney, C., Hawkins, N., Vries, J. de, & Boddington, P. (2009). Data sharing in genomics -- re-shaping scientific practice. *Nature Reviews Genetics*, *10*(5), 331–335. http://doi.org/10.1038/nrg2573.

Kleiner, K. (2011). Data on demand. *Nature Climate Change*, *1*(1), 10–12. http://doi.org/10.1038/nclimate1057.

Masats, D., & Dooly, M. (2011). Rethinking the use of video in teacher education: A holistic approach. *Teaching and Teacher Education*, *27*(7), 1151–1162. http://doi.org/10.1016/j.tate.2011.04.004.

Methods of Effective Teaching Learning Exploratory (n.d). Retrieved 2016-02-07 from http://soe.mivideo.it.umich.edu/.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D.P., Humphreys, M., Imbens, G., Laitan, D., Madon, T., Nelson, L, Nosek, B.A., Petersen, M., Sedlmayr, R., Simmons, J.P., Simonsohn, U., & Laan, M. V. der. (2014). Promoting Transparency in Social Science Research. *Science*, *343*(6166), 30–31. http://doi.org/10.1126/science.1245317.

National Science Board. (2011). *Digital Research Data Sharing and Management* (No. NSB-11-79). Retrieved from http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf.

Nature Scientific Data (n.d.).

Nosek, B.A., & Bar-Anan, Y. (2012). Scientific Utopia: I. Opening Scientific Communication. *Psychological Inquiry*, *23*(3), 217–243. http://doi.org/10.1080/1047840X.2012.692215.

Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., ... Wagenmakers, E.J., Wilson, R., & Yarkoni, T (2015). Promoting an Open Research Culture. *Science*. http://doi.org/10.1126/science.aab2374

Open Humans Project. (n.d.). Retrieved 2016-02-07 from https://www.openhumans.org/.

Open Science Framework. (n.d.). Retrieved 2016-02-07 from http://osf.io.

PeerJ. (n.d.). Retrieved 2016-02-07 from http://peerj.com.

Perneger, T.V. (2011). Sharing raw data: another of Francis Galton's ideas. *BMJ*, *342* (May 17 2), d3035–d3035. http://doi.org/10.1136/bmj.d3035.

Personal Genome Project. (n.d.). Retrieved 2016-02-07 from http://www.personalgenomes.org/.

PLOS One. (n.d.). Retrieved 2016-02-07 from http://www.plosone.org.

Poldrack, R.A., & Gorgolewski, K.J. (2014). Making big data open: data sharing in neuroimaging. *Nature Neuroscience*, *17*(11), 1510–1517. http://doi.org/10.1038/nn.3818.

Poline, J.B., Breeze, J. L., Ghosh, S., Gorgolewski, K., Halchenko, Y.O., Hanke, M., ... Kennedy, D. N. (2012). Data sharing in neuroimaging research. *Frontiers in Neuroinformatics*, *6*, 9. http://doi.org/10.3389/fninf.2012.00009.

Python (n.d.). Retrieved 2016-02-07 from https://www.python.org.

R Project. (n.d.). Retrieved 2016-02-07 from http://cran.r-project.org.

Raudies, F., & Gilmore, R.O. (2014). Visual motion priors differ for infants and mothers. *Neural Computation*, *26*(11), 2652–2668. http://doi.org/10.1162/NECO_a_00645

Reichman, O.J., Jones, M.B., & Schildhauer, M.P. (2011). Challenges and Opportunities of Open Data in Ecology. *Science*, *331*(6018), 703–705. http://doi.org/10.1126/science.1197962.

Research Data Alliance. (n.d.). Retrieved 2016-02-07 from https://www.rd-alliance.org/groups/domain-repositories-interest-group.html.

Scott, C.S. (2011). 'Tangible as tissue': Arnold Gesell, infant behavior, and film analysis. *Science in Context*, *24*(3), 417–42. http://dx.doi.org/10.1017/S0269889711000172

SlideShare. (n.d.). Retrieved 2016-02-07 from http://www.slideshare.net.

Software Carpentry (n.d.). Retrieved 2016-02-07 from https://software-carpentry.org.

Young, J.R. (2010, May 28). Crowd Science Reaches New Heights. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/The-Rise-of-Crowd-Science/65707/.

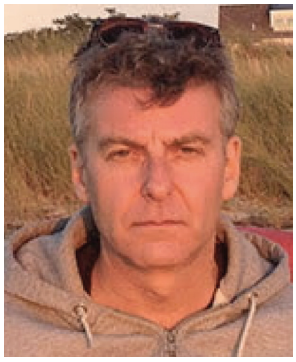
YouTube (n.d.). Retrieved 2016-02-07 from https://www.youtube.com/yt/press/statistics.html

Yu, C., & Smith, L.B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS ONE*, *8*(11), e79659. http://doi.org/10.1371/journal.pone.0079659.

AUTHORS

Rick O. Gilmore is in the Department of Psychology, The Pennsylvania State University, University Park, PA 16802, rogilmore@psu.edu. Gilmore is the Associate Director of Databrary.

Karen E. Adolph is a Professor in the Department of Psychology and the Center for Neural Science at New York University (NYU). Adolph is the Director of the Databrary Project.



David S. Millman is the Assistant Dean of Digital Library Technology Services at NYU, and is Co-Investigator on Databrary.

Andrew Gordon is the Information Engineer with Databrary.