

The Role of Video Data Sharing in Advancing Education Policy and Practice

Rick O. Gilmore

The Pennsylvania State University, The Databrary Project

Karen E. Adolph, David S. Millman

New York University, The Databrary Project

Author Note

Rick O. Gilmore is in the Department of Psychology, The Pennsylvania State University, University Park, PA 16802, rogilmore@psu.edu. 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, and conclusions or recommendations expressed in the material contributed here are those of the author(s) 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.

Abstract

Video captures the complexity, richness, and diversity of behavior unlike any other measure. As a result, large numbers of people who study teaching and learning employ video. Video documents itself to a large degree, and therefore has significant potential for reuse by others. This potential remains largely unrealized because videos are rarely shared. Video often contains information about personal identities, so considerations about privacy and research ethics pose challenges to sharing. The relatively large size of video files, diversity of formats, and incompatible software tools for coding pose technical challenges. In this paper, we will describe how the Databrary data library has overcome the most significant barriers to sharing video within the developmental sciences community, including solutions to maintaining participant privacy, data tagging, and data management. Databrary suggests ways that video and other identifiable data collected in the context of education research might be shared. Widespread open sharing of high value, high impact data of the sort provided by video will advance education policy and improve practice.

The Role of Video Data Sharing in Advancing Education Policy and Practice

Introduction

The application of insights from scientific research to essential human needs such as improved educational policy and practice depends on the free flow of information. Open data sharing embodies many virtues: It bolsters transparency and peer oversight, encourages diversity of analysis and opinion, accelerates the education of new researchers, and stimulates the exploration of topics not envisioned by the original investigators. Open data sharing increases the impact of public investments in research, accelerates the pace of discovery, and leads to better and more effective public policy. Open data sharing is a scientific imperative (Board, 2011; of Mental Health, 2015) and common practice in many areas of biomedical (Kaye, Heeney, Hawkins, Vries, & Boddington, 2009), physical (Young, 2010), and earth sciences (Kleiner, 2011) where it has accelerated the pace of discovery.

Unfortunately, despite the efforts of funding agencies and professional organizations (“AERA Code of Ethics”, 2011) to make data sharing a norm, most research on human development and learning (Adolph, Gilmore, Freeman, Sanderson, & Millman, 2012) and in the social sciences in general (Nosek & Bar-Anan, 2012) remains shrouded in a culture of isolation. Rather than providing direct access to raw data, researchers in the developmental and learning sciences share interpretations of distilled data almost exclusively through publications and presentations. The path from raw data to finding to conclusion cannot be traced or validated by others, nor can new questions be posed by others building on the same, often expensive-to-collect, raw data. To realize significant advances in our understanding of teaching and learning and for those insights to have meaningful impact on educational practice and policy, open data sharing must become an established norm, not the rare exception. Moreover, we must find ways to share the most valuable raw data, the data with the greatest potential for reuse by others, and the data with the greatest potential for yielding meaningful insights into the learning process.

Video is a uniquely rich medium for capturing in real time the interactions that occur

in classrooms, museums, laboratory, home, and other informal learning settings. Many researchers and practitioners in the developmental, learning, and education sciences collect video as raw data. Video recordings capture what teachers, parents, tutors and peers say and do and what learners say and do in response. The recordings capture these dynamics in ways that offer the potential to discover who did what, and where, when, why, and how it influenced learning. Still, the open sharing of video data is rare, even within labs or among collaborators. Video data should be conducive to sharing. Video captures behavior that others can readily inspect. It captures multiple dimensions of complex real-world learning settings in ways that others can easily build upon. But, technical, ethical, and cultural barriers have made open sharing of video data rare. Videos come in varied formats and generate large files; these pose conversion and storage challenges. Personally identifying information can be removed from text-based data. Videos may contain faces, voices, names spoken aloud, views or voices of non-participants, views of classroom interiors and sometimes views of the homes of research participants. These elements cannot be removed without reducing the information content and reuse value to others. The collection of video or other identifiable or sensitive information requires approval by a research ethics board. It also requires informed consent from participants, and possibly parental, teacher, and school district approval. Researchers risk violating participants' privacy if digital images are viewed or released to others without authorization. For these and other reasons, no culture of widespread open sharing of video data has emerged in the education, learning, and developmental sciences.

The Databrary project has built a digital data library specialized for the open sharing of video data from research in these areas of inquiry. In this paper, we will describe how Databrary has overcome the most significant barriers to sharing video, including solutions to maintaining participant privacy, data tagging, and data management. In developing technology and policies that enable the safe and secure sharing of video and other identifiable research data, Databrary suggests ways that video and other identifiable data

collected in the context of research on teaching and learning might be shared. Fostering the open sharing and reuse of high value, high impact data like video promises to advance education policy and improve practice.

The Promise of Video

More than 1 billion users worldwide upload more than 300 hours of video to YouTube every minute (YouTube, 2015). Google, YouTube's owner, has built an immense data infrastructure to upload, store, convert, tag, and stream video. The current scale of video collection in developmental and education contexts is undoubtedly smaller and more focused in purpose, but it is large and growing. The Gates Foundation-funded Measures of Effective Teaching (MET; <http://www.metproject.org/>) Project generated more than 1,000 full-length video segments from K-12 classrooms, covering core subjects such as mathematics, language arts from multiple camera angles. The data constituting tens of terabytes of storage are hosted at the University of Michigan (<http://soe.mivideo.it.umich.edu/>) and securely streamed to registered viewers across the country. Similarly, in just over a year of operation, the Databrary (<http://databrary.org>) digital library has collected more than 7,000 individual videos, representing 2,400 hours of recording, and featuring more than 1,800 infant, child, and adult participants. Databrary has more than 100 authorized researchers representing more than 50 institutions across the globe who access the videos. Video data is big, and the interest in recording and sharing video for research, education, and policy purposes continues to grow.

Video captures the richness and complexity of human behavior and interactions between people and their environment like no other measure. Video captures what people say and what they do. It captures when and where they look or gesture, when, where and how they move, and how their looking, gesturing, and moving corresponds to what others are doing. These characteristics mean that video has multiple use cases in the study of learning and development. In laboratory contexts, researchers use videos to study how

(Karasik, Tamis-LeMonda, & Adolph, 2014) infants, children, and adults behave in response to natural or experimenter tasks. In educational contexts, educators, teacher trainers and evaluators use video to record what other teachers do to help improve teaching practices or evaluate performance (Blomberg, Sherin, Renkl, Glogger, & Seidel, 2014; Masats & Dooly, 2011; Baecher, Kung, Jewkes, & Rosalia, 2013). Researchers record videos of learners in classrooms (Alibali & Nathan, 2012) to understand what teachers do and how audiences respond. Some explore how *using* video affects what museum-goers learn (Bakken & Pierroux, 2015). In lab (Kretch, Franchak, & Adolph, 2014) and classroom (Mason, Pluchino, & Tornatora, 2015; Prieto, Wen, Caballero, Sharma, & Dillenbourg, 2014) settings video-based eye tracking systems provide information about where people look and for how long. While not strictly video recordings, image streams from computer-based tasks or learning-rich video games may be combined with information about user keyboard or controller inputs to study real-time performance. When combined with other quantitative and qualitative measures about learning outcomes, video can provide information about what kinds of teaching techniques boost learning for what kinds of students. Video may even be used to study engagement by students enrolled in online/web-based courses or to monitor compliance with academic integrity rules.

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 another purpose. In other words, video requires less metadata and less manual curation than other data types, but offers greater potential for reuse. Moreover, the success, growth, and popularity of YouTube and other video-based social media (e.g., Vimeo, Periscope, Meerkat, SnapChat) demonstrates that video garners high levels of user interest and attention for interpersonal communication and entertainment uses. The success of these services also suggests that web-based video storage and streaming systems are now sufficiently well developed to satisfy large-scale demand. The question for researchers and policymakers is how to capitalize on this momentum in

order to realize video's substantial potential to improve teaching and learning.

The Challenge of Video

Sharing research video poses challenges. These include technical infrastructure, research ethics, privacy, and security, and data management. Further, making open data sharing in education research an established norm requires changing minds. Some researchers and research ethics officials believe that identifiable research data may not be shared at all, when in fact, open sharing of this sort of data can be achieved with proper informed consent.

Technical Challenges

Videos are large. One hour of HD video (1080p, 1440x1080 @ 60 Hz) can consume more than 10 GB. The size of the file varies by format and resolution, and recordings vary widely, making it essential for video storage systems to convert files to a standard format. Many settings, like classrooms, require multiple camera views to capture the desired behaviors. Some questions involve the collection of multiple, temporally dense data streams (physiological recordings, motion tracking, eye tracking, verbal transcripts, keyboard/controller responses) alongside video. These require tools for synchronization across camera views and data streams. The use of video creates a data explosion: A typical lab studying infant or child development collects 8-12 hours of video/week in widely varied formats (Gilmore & Adolph, 2012). Thus, sharing digital video requires substantial storage capacity, significant computational resources for transcoding videos into common, preservable formats, and powerful search and streaming tools.

After videos are recorded, human raters evaluate videos and apply text-based annotations or numeric codes to videos. Raters use a wide-range of commercial (Noldus Observer, Mangold Interact, MaxQDA, V-Note), and academic (Transana, Elan, Datavyu) software tools, spreadsheets or even pencil and paper. Some tools (Datavyu, Mangold, V-Note) offer text-based export to interoperable formats like comma-separated value

(CSV) files, but the tools lack standard formats for data storage or export. Some tools offer web/cloud-based storage (e.g., Datavyu via Databrary, Transana, V-Note), but otherwise coded data are not easily shared between users employing the same tools, much less between them.

Human-centered video annotation is likely to remain the primary mode of data coding and reduction for some time given the unique abilities of trained observers to capture important behavioral dimensions. However, there has been explosive growth in the use of machine learning and computer vision tools to tag images and videos (Research, 2014) . With few exceptions (Amso, Haas, Tenenbaum, Markant, & Sheinkopf, 2014; Yu & Smith, 2013; Raudies & Gilmore, n.d.; Fathi, Hodgins, & Rehg, 2012), these techniques have not been applied to video data coding in the developmental and learning sciences. The application of machine-generated tags to research videos presents an as-yet-untapped opportunity to apply big data tools to educational and developmental research.

Ethics and Privacy

Sharing video recordings poses a challenge to existing data sharing policies. De-identified data (of Health & Services, 2010) may often be openly shared, but since the faces and voices contained in video are considered identifiable information, de-identifying video requires their removal. This severely diminishes the reuse value of video and poses additional costs and burdens on researchers. So, sharing video data openly requires the development of new policies that protect the privacy of research participants while preserving the integrity of raw video for reuse by others. Databrary has developed an access model that achieves this by building on the well-established principle of informed consent and restricting access to authorized individuals who meet specific criteria. We discuss the access model in detail below.

Data Management

As in other fields (Overpeck, Meehl, Bony, & Easterling, 2011), developmental and education science is inundated by an explosion of digital data, most of which is inaccessible to other researchers. A primary impediment to sharing is researchers' lack of time to find, label, clean, organize, and copy their files into formats that can be used and understood by others (G. A. Ascoli, 2006). Even scientists committed to data sharing can find it difficult to do. Data management systems that make sharing convenient, reliable—and optimally automatic—would help.

But, even with the availability and widespread adoption of suitable data management tools, existing data management practices pose challenges for video sharing. Despite similar research methods, study designs vary widely. No two labs manage data in the same way. Some studies are longitudinal, where researchers observe the same participants at multiple sessions. Some are cross-sectional, where researchers observe each participant at only one session. The timing of observations may be determined by participants' age (e.g., kindergarteners) or abilities (e.g., readers), experimentally determined variables (e.g., pre- and post-intervention), or other factors. Some researchers maintain central, organized file systems, while others allow trainees to keep separate records, making it difficult to share data even within the lab. Some researchers keep videos, metadata, and analyses together, and some do not. Idiosyncratic terms, record-keeping, and data management systems are the norm. Raw data without information about workflow and data provenance is essentially “orphaned” and cannot be understood by subsequent researchers who weren't part of the original team (Curry, 2011). Moreover, even though video requires minimal metadata to be suitable for reuse, video files should be electronically linked to whatever relevant metadata do exist, and there must be information linked to the videos that determines what level of access is permitted. We describe below Databrary's data management system that enables users to manage the uploading of research videos and related metadata as these are collected.

Changing Community Practices

Finally, we note that while data sharing is an established best practice in some areas of educational research (“AERA Code of Ethics”, 2011), it is not widespread. The reasons are many (G. Ascoli, 2006; Ferguson, 2014). 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. Researchers also note that data shaving diverts energy and resources from other scholarly activities that are more heavily and frequently rewarded (G. Ascoli, 2006; Ferguson, 2014). Put another way, data sharing involves mostly sticks and few carrots. In order to make data sharing become a scientific norm, these factors must be addressed.

Databrary.org

The Databrary project (databrary.org) arose to meet the challenges of sharing research video and deliver on the promise of open data sharing in educational and developmental science. With funding from NSF (BCS-1238599) and the National Institute of Child Health and Human Development (NICHD U01-HD-076595) the project has focused on building a data library specialized for video, creating data management tools, creating new policies to enable video sharing, and creating a community of researchers who embrace video sharing. A fifth aim focuses on the development of a free, open-source video annotation tool, Datavyu (<http://datavyu.org>). Our interdisciplinary team brings expertise in developmental and neural science, information technology, library science, software development, data curation, and community engagement. The project receives guidance from a board of expert advisors composed of developmental scientists, library scientists, and leaders of data repositories in the behavioral and social sciences provide guidance. The project received funding in 2012-2013, began a private beta testing phase in the spring of 2014 and opened for full use in October 2015.

System Design

No data sharing system would be complete without an information technology infrastructure that enables large numbers of video and related files to be uploaded, organized, stored, streamed, and tagged. Considerations about what file formats to support are also paramount. Databary is built on a PostgreSQL database using the Scala Play framework and JavaScript. Data are preserved indefinitely in a secure data storage facility at NYU managed by central IT. There is no cost to use the system, but an institutional subscription model is under development. All code is hosted on GitHub (<http://github.com/databary>). Databary can house video, audio, PDF, spreadsheet, image, and text-based files (along with associated metadata), as well as executable scripts in Ruby, R and Matlab that facilitate data analysis. Video and audio data are transcoded into standard and HTML5-compatible formats, currently H.264+AAC as MP4. This ensures that video and audio data can be streamed and downloaded by any operating system that supports a modern browser. Preservation copies of original video and audio files are also stored. Databary stores other data in native formats (e.g., .doc, .docx, .xls, .xlsx, .txt, .csv, .pdf, .jpg, .png). Since common video coding/tagging file formats are required for users to build on earlier coding efforts, Databary plans to develop tools for converting files from other coding tools into a common format.

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 and contains one or more recordings or related flat files (assets). Researchers can group sessions in different ways, using whatever groupings are appropriate for the particular dataset. These groupings may also identify particular time segments of a recording to distinguish tasks, events, or participants that comprise the session. The aggregate of these flexible groupings is called a *volume*. In practice, researchers combine sessions or segments of sessions within and across datasets to form the raw material that are subsequently described in published articles and presentations. Thus, Databary

contributors can combine sessions or segments within and across datasets with ancillary materials, such as coding manuals, coding spreadsheets, statistical analyses, questionnaires, IRB documents, computer code, sample displays, and links to published journal articles in an aggregate structure investigators call a *study*. Like datasets, studies are represented as volumes in the data model.

The discovery and browsing interfaces offer content items at the volume (study or dataset) level to users searching the library. The contributor-assigned groupings may include metadata about participants, tasks or measures, and locations. This enables users to quickly identify data that may be relevant to their own interests or research. Users can also add keyword tags at the study/dataset, session, or segment level, and can endorse or deprecate keywords that have been suggested by others. Different users may have very different reasons for citing a study or using data, so these features are critical for fostering reuse. Future implementations may extend the ability of authorized researchers to annotate studies with other kinds of metadata.

The Databrary team recognizes that other data repositories enforce strict metadata ontologies and that these Doing may have benefits in sub-domains of research when there is community consensus on core concepts. The Databrary team has chosen not to require strict metadata ontologies for several reasons. We hope to reduce the pre-deposit curational demands on contributors and foster the rapid adoption of data sharing. Developing and achieving agreement on metadata ontologies can take significant amounts of time. 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 standard tasks, procedures, or terminology. However, Databrary empowers users to re-use tags and terminology by suggesting common or similar terms, without confining users to these suggestions. User communities within Databrary may eventually converge on common conceptual and metadata ontologies based on the most common (and commonly endorsed) keyword tags, but standardized ontologies are not necessary for

browsing and searching in most of the use cases we envision.

Future challenges include building the capacity to searching for tagged segments inside of videos. Some of within-video tagging functionality exists in the current software, with more extensive search and preview capabilities on the near horizon. A related challenge involves developing mechanisms that enable files from video coding tools to be imported into Databrary so that user supplied codes may be visualized independent of the desktop coding tool deployed in a particular project. We envision a parallel set of export functions that permit full interoperability among coding tools. Since some tools use proprietary data formats, the priority will be to create interoperability with tools using open formats. The Databrary team believes that the effort spent creating this powerful and flexible system will enable it to be extended to a diverse set of communities engaged in educational research who collect and seek to share video.

Policies for Safe & Secure Video Sharing

Equally important for the success of a video data sharing enterprise are policies for openly identifiable data in ways that securely preserve participant privacy. As discussed previously, videos cannot be easily or inexpensively deidentified without undermining their reuse value. Databrary has chosen to maximize the potential for data 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 i) restricts access to authorized individuals, and ii) enables access to identifiable data only with the permission of the people depicted.

To gain access to Databrary a person must register on the site and electronically sign an access agreement, which must be co-signed by the applicant's institution. Applicants must agree uphold common practices concerning the responsible and ethical use of identifiable and sensitive data. Full privileges will be granted only to researchers with independent researcher status at their institutions. Others may be granted privileges if

they are affiliated with researcher who agrees to sponsor and supervise their application. Ethics board or IRB approval is not required to gain access to Databrary since many use cases do not involve research. Approval is required for uses that involve research. Once authorized, a user may have full access to the site's shared data, and may browse, tag, download for later viewing, and conduct non- or pre-research activities.

Unique among data repositories, the Databrary access agreement authorizes data use and contribution. However, users may only store materials on Databrary for which they have approval from an ethics board or IRB and for identifiable data like video, permission from the people depicted. Databrary has extended the commonly understood principle of informed consent to participate in research to encompass permission to share data with other researchers. In other words, identifiable data may be shared if the participants agree.

Data from a particular session may be stored in Databrary for the contributing researcher's use whether the records are shared with other scientists or not. When a researcher chooses to share, Databrary makes records openly available to the community of authorized researchers only if the people depicted in the recordings have given permission to release the data for sharing. To formalize the process of acquiring permission, Databrary has developed a Participant Release Form Template, based on photo or video release language many researchers use currently. The template release form has standard language that Databrary recommends investigators should use with study participants. This language makes it easy for participants to understand what is involved in sharing their video data, with whom it will be shared, and the potential risks associated with releasing their video and other identifiable data to other researchers. Use of the template also allows for the standardization of language associated with the release of identifiable or sensitive participant data.

Some IRBs may deem an investigator's existing, approved video or photo release form equivalent to the Databrary release. This enables a researcher to share with Databrary recordings they have already collected. However, most researchers will need to modify their

research protocols, by adding the Databrary sharing permission procedures, prior to collecting new shareable video data. Databrary staff are available to advise potential data contributors about how to amend existing research protocols so that the information acquired is Databrary-compliant. Protocol amendments involve seeking approval for use of the Databrary template release form and modifying the time period over which collected data will be made available. Specifically, researchers must remove any clauses in research consent documents that require data destruction after some fixed period of time since Databrary intends to store shared data indefinitely.

We developed these ideas in close collaboration with the NYU and PSU research ethics and administration official, and we took inspiration from the Human Connectome Project and Open Genome Project that have similarities. As of May 2015, Databrary has secured agreements with 55 institutions on behalf of 105 researchers representing a diverse array of entities in North and South America, Europe, and Australia. We believe that this model of sharing identifiable video data can be extended to other data types and other contexts. In the realm of education research, we recognize that the access model will have to adapt to meet the needs of users who are not affiliated with an institution as currently conceived and to address the role of other parties involved in the educational research enterprise, namely teachers, schools, and school districts. Nevertheless, we believe that the core ideas provide a strong foundation for enabling the widespread sharing of video and other identifiable data to the educational research community.

Managing Data for Sharing

As we discussed earlier, even when researchers are willing to share data and have the required approvals to do so, varying data management practices make curating data for sharing after a research project has finished a difficult, often unrewarded chore. In order to make the raw data informative to others, it must be linked with coding spreadsheets, codebooks, protocols, statistical analyses, and manuscripts. But, the organization scheme

that works for one research team does not necessarily work for general audiences unfamiliar with the nuances of a research project. Our experiences curating and ingesting archival datasets have highlighted the considerable value of contributors entering raw video data into Databrary as soon as recordings are acquired. 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. Accordingly, Databrary has developed a data management system that empowers researchers to curate their own data *as it is collected*.

The system employs familiar, easy-to-use spreadsheet and timeline-based interfaces that allow users to upload videos, add metadata about tasks, settings, and participants, link related files, document workflow and data provenance, and tag files with appropriate permission levels for sharing. To encourage immediate uploading, Databrary provides a complete set of controls so that researchers can restrict access to their data to only their own labs or to other users of their choosing. Datasets can be shared 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.

It is too early to evaluate the impact of Databrary's data management tools on researcher's willingness to share data. But, the developmental science community is hungry for new practices that will enhance research productivity, as demonstrated by the rapid growth in the number of authorized investigators. We are confident that consistent data management practices will strengthen work within labs and among collaborators, enable convenient and reliable file uploading, and enhance the value of shared data in the Databrary.

Building A Community That Embraces Sharing

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 enacting that support in their own work flows. Those efforts involve many interacting components. They include active engagement with professional associations, conference-based exhibits and training 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 and labor-intensive, but we believe that they are critical if we are to change community attitudes toward data sharing in the educational and learning sciences. Looking ahead, it will be critical to engage professional organizations that have already made fostering video data sharing a priority in the effort to forge community consensus about the importance, feasibility, and potential of open video data sharing.

Conclusion

Imagine a time in the near future when researchers interested in studying classroom teaching and learning can mine an integrated, synchronized, interoperable, and open data set. That data set includes multiple measurements – video from multiple cameras, eye tracking, motion, and physiological measurements, and both historical and real-time student performance data. 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 made keenly sensitive to privacy, confidentiality and ethical issues involved in research involving sensitive, identifiable information. We then 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 (News, 2014). 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 could be the core around which other measures of teaching and learning cluster. This requires reducing barriers to sharing video and related data and one that fosters new community values around ethical data sharing. The Databrary project has built technology and policies that overcome many of the most significant barriers to widespread sharing within the developmental sciences community, including solutions to maintaining participant privacy, data tagging, and data management. 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 will have 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 video data 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 advance education policy and improve practice. The Databrary team welcomes collaborations with other like-minded parties who share our vision (Adolph et al., 2012) for a future where open video data sharing is the norm and a personalized science of teaching and learning is the goal.

References

- Adolph, K. E., Gilmore, R. O., Freeman, C., Sanderson, P., & Millman, D. (2012). Toward open behavioral science. *Psychological Inquiry*, 23(3), 244-247. Retrieved from <http://doi.org/10.1080/1047840X.2012.705133>
- Aera code of ethics. (2011). *Educational Researcher*, 40(3), 145-156. Retrieved from <http://edr.sagepub.com/content/40/3/145.short> doi: 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. Retrieved from <http://dx.doi.org/10.1080/10508406.2011.611446> doi: doi: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. Retrieved from <http://dx.doi.org/10.1007/s10803-013-1925-5> doi: 10.1007/s10803-013-1925-5
- Ascoli, G. (2006). The ups and downs of neuroscience shares. *Neuroinformatics*, 4(3), 213-215. doi: 10.1385/NI:4:3:213
- Ascoli, G. A. (2006). Mobilizing the base of neuroscience data: the case of neuronal morphologies. *Nature Reviews Neuroscience*, 7(4), 318-324. doi: <http://doi.org/10.1038/nrn1885>
- 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(0), 189 - 197. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0742051X13001315> doi: <http://dx.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*(0), -. Retrieved from

<http://www.sciencedirect.com/science/article/pii/S2210656115000021> doi:

<http://dx.doi.org/10.1016/j.lcsi.2014.12.001>

Blomberg, G., Sherin, M., Renkl, A., Glogger, I., & Seidel, T. (2014). Understanding video as a tool for teacher education: investigating instructional strategies to promote reflection. *Instructional Science*, 42(3), 443-463. Retrieved from

<http://dx.doi.org/10.1007/s11251-013-9281-6> doi: 10.1007/s11251-013-9281-6

Board, N. S. (2011). *Digital research data-sharing and management* (Tech. Rep.).

Retrieved from <http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf>

Curry, A. (2011). Rescue of old data offers lesson for particle physicists. *Science*, 331(6018), 694-695. Retrieved from

<http://doi.org/10.1126/science.331.6018.694>

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). doi: 10.1109/CVPR.2012.6247805

Ferguson,

L. A. (2014). *How and why researchers share data (and why they don't)*. Retrieved from

<http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-w>

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

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. Retrieved from <http://dx.doi.org/10.1111/desc.12129> doi:

10.1111/desc.12129

Kaye, J., Heeney, C., Hawkins, N., Vries, J. d., & Boddington, P. (2009). Data-sharing in genomics - - re-shaping scientific practice. *Nature Reviews Genetics*, *10*, 331-335.

Kleiner, K. (2011). Data on demand. *Nature Climate Change*, *1*, 10-12.

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development*, *85*(4), 1503–1518. Retrieved from <http://dx.doi.org/10.1111/cdev.12206> doi: 10.1111/cdev.12206

Masats, D., & Dooly, M. (2011). Rethinking the use of video in teacher education: A holistic approach. *Teaching and Teacher Education*, *27*(7), 1151 - 1162. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0742051X11000497> doi: <http://dx.doi.org/10.1016/j.tate.2011.04.004>

Mason, L., Pluchino, P., & Tornatora, M. C. (2015). Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology*, n/a–n/a. Retrieved from <http://dx.doi.org/10.1111/bjet.12271> doi: 10.1111/bjet.12271

News, B. (2014). *Health data boom heralds new era of personalised medicine*. Retrieved 2015-05-10, from <http://www.bbc.com/news/business-26133269>

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. opening scientific communication. *Psychological Inquiry*, *23*(3), 217–243. doi: 10.1080/1047840X.2012.692215

of Health, U. D., & Services, H. (2010). *Guidance regarding methods for de-identification of protected health information*. Retrieved 2015-05-08, from <http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification>

of Mental Health, N. I. (2015). *Data sharing expectations for clinical research funded by nimh* (Tech. Rep.). Retrieved from <http://grants.nih.gov/grants/guide/notice-files/NOT-MH-15-012.html>

Overpeck, J., Meehl, G., Bony, S., & Easterling, D. R. (2011). Climate data challenges in

- the 21st century. *Science*, 331, 700–702. doi: 10.1126/science.1197869
- Prieto, L. P., Wen, Y., Caballero, D., Sharma, K., & Dillenbourg, P. (2014). Studying teacher cognitive load in multi-tabletop classrooms using mobile eye-tracking. In *Proceedings of the ninth acm international conference on interactive tabletops and surfaces* (pp. 339–344). New York, NY, USA: ACM. Retrieved from <http://doi.acm.org/10.1145/2669485.2669543> doi: 10.1145/2669485.2669543
- Raudies, F., & Gilmore, R. O. (n.d.). Visual motion priors differ for infants and mothers. *Neural Computation*, 26(11), 2652–2668.
- Research, G. (2014). *A picture is worth a thousand (coherent) words: Building a natural description of images*. Retrieved 2015-05-08, from <http://googleresearch.blogspot.com/2014/11/a-picture-is-worth-thousand-coherent.htm>
- Scott, C. S. (2011). ‘tangible as tissue’: Arnold gesell, infant behavior, and film analysis. *Science in Context*, 24(3), 417-42.
- Young, J. R. (2010). Crowd science reaches new heights. *The Chronicle of Higher Education*. Retrieved 2015-05-09, from <http://chronicle.com/article/The-Rise-of-Crowd-Science/65707/>
- YouTube. (2015). *Youtube statistics*. Retrieved 2015-05-07, from <https://www.youtube.com/yt/press/statistics.html>
- Yu, C., & Smith, L. B. (2013, 11). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS ONE*, 8(11), e79659. doi: 10.1371/journal.pone.0079659