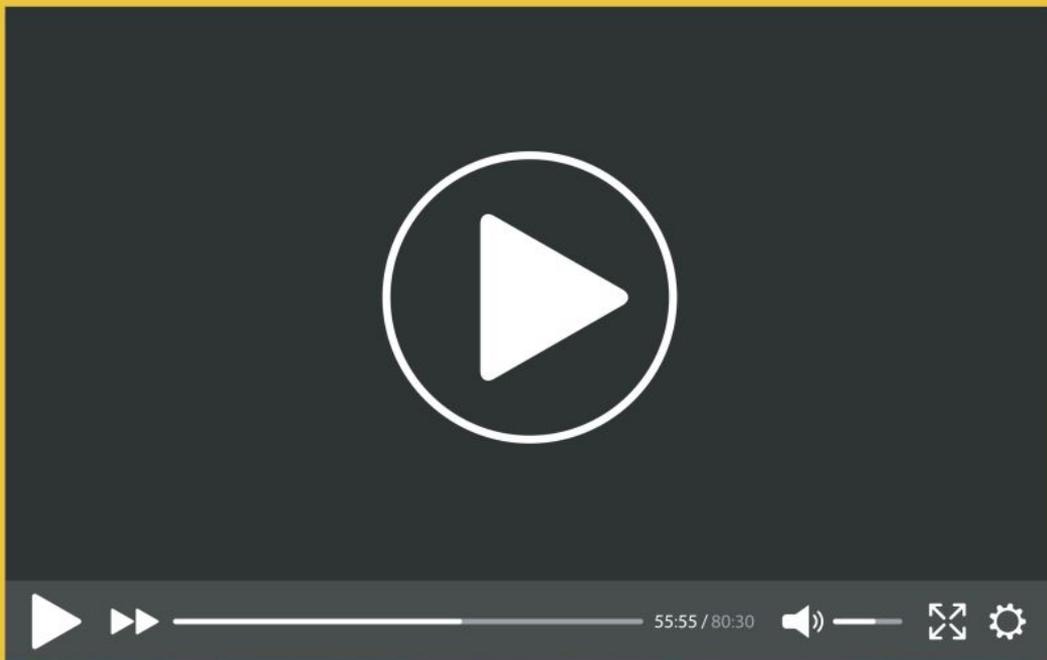


# Video as Data

February 29, 2016



Children's behavior is rich, complex, and fascinating. But it is transient. At every time scale, from milliseconds to months, behavior happens and then it vanishes. Newborns' "gas smiles" metamorphose into expressions of real pleasure, babies' babbles become reciprocal conversations and teen poetry jams, and infants' awkward toddling steps transform into ballet recitals and soccer goals. But all of it disappears into the ether as soon as the moment has passed.

Since the inception of developmental psychology, researchers have endeavored to transform the ephemeral nature of behavior into something more tangible than the written descriptions that Charles Darwin, William Preyer, and other eminent, scientifically minded parents relied on to describe their children's behavior. The founders of psychology, Wilhelm Wundt and G. Stanley Hall, argued that words are too amorphous and subjective to render behavior in sufficient clarity for detailed, reproducible analysis. So developmental researchers turned to visual media.

Arnold Gesell and Myrtle McGraw, two pioneers in the study of child development, developed sophisticated techniques to create objective, quantifiable, and reproducible observations on film. Perhaps more importantly, they showed how cinematic recording could manipulate space and time to facilitate scientific discovery. Panning and zooming bring particular aspects of behavior into focus. Wide-angle lenses bring the surrounding context into view. Multiple camera views create new perspectives on the same behavior. Running the film at different speeds slows down or speeds up time. And analysis of individual frames freezes time, allowing the anatomy of behavior to be dissected into its component parts. As Gesell said, "the cinema registers the behavior events in such coherent, authentic, and measurable detail that for purposes of psychological study and clinical research the reaction patterns of infant and child become almost as tangible as tissue" (1952, p. 132).

Over the ensuing decades, recording technologies have vastly improved, but whether behavior is filmed, videotaped, or digitally captured, visual records are vulnerable to careless preservation and cataloging. Much of the vast film libraries compiled by Gesell and McGraw have decayed, become a scrambled mess, or crumbled into dust. Similarly, most modern developmental researchers have lost digital video files thanks to corrupted hard drives, DVDs that are no longer readable, or irretrievable file formats. Other files are forgotten on a computer in the corner. The gray-haired among us also have boxes of research videotapes moldering away in a storage closet — VHS, Betacam, and those tiny tapes that no one can remember the name of or find the camera for. And where are the metadata that describe who is on each video, when the recording was collected, and to what study the video data belong?

Our research history is disappearing, along with the children's behaviors that we worked so hard to capture.

Enter [Databrary](#), a web-based video library funded by the National Science Foundation and the National Institutes of Health to enable sharing and reuse of research videos among developmental scientists. Databrary stores and preserves the videos in the most current standard for digital file formats. The shared videos are accessible, searchable, and reusable by a rapidly growing group of developmental researchers who are authorized by their institutions with oversight by their ethical review boards. Databrary is housed at New York University, securely protected by the university information-technology services, and supported by the university libraries. All recordings relevant to developmental research are welcome.

## **The Value of Video**

Unfortunately, most research on behavioral development remains shrouded in a culture of isolation. Rather than providing direct access to raw data, developmental researchers typically share interpretations of distilled data through publications and presentations. A notable exception in the developmental community is the Child Language Data Exchange System ([CHILDES](#)), which has

provided child language researchers with a digital repository of child language transcripts and raw audio files since 1984 and supported several thousand publications through the use of shared data. In the larger social science community, the Inter-university Consortium for Political and Social Research ([ICPSR](#)) has provided a repository of demographic data, including data about children's development, since 1962.

Distilled behavioral data require relatively extensive documentation to be interpretable. Flat-file data (e.g., numbers or string variables in tabular spreadsheets), imaging data (e.g., MRI, EEG), and physiological data (e.g., heart rate, motion tracking, galvanic skin response) require detailed descriptions of the researcher's workflow (how the data were collected and for what purpose, what the numbers or images represent, and how the data were processed so as to produce those values).

In contrast, video is largely self-documenting. Merely viewing a research video provides vast amounts of information about who the participants were, where they were, what they were doing, and how the data were collected. Details about data provenance are often unnecessary, and a small amount of metadata (e.g., child's birth date and gender, test date, identity of the research team) goes a long way.

Moreover, video is a uniquely rich form of behavioral data. In contrast to other forms of time-series data or static images, video enables developmental researchers to see behaviors unfold. Typically, videos:

- contain a sound track, so researchers can hear, not just see, behavior;
- depict the context, so researchers can see where behaviors are happening and, in many cases, whether the surrounding environment contributes to the expression of the behavior; and
- include both facial and body information, so researchers can score where participants are looking, what they are doing, and what facial expressions they are exhibiting while doing it.

Databrary has developed a novel active-curation framework. Researchers are encouraged to upload videos (and, if desired, files in other formats) after each session of data collection and to fill relevant metadata fields in a flexible, modifiable spreadsheet to aid in data management and to facilitate search and reuse. With this active-curation framework, the cost in time and labor to researchers is equivalent to current lab practices of storing a copy of the video on a hard drive and entering the associated metadata into a spreadsheet. Moreover, with this method, Databrary acts as the researcher's personal lab file server and cloud storage, enabling web-based sharing among members of the protocol and ensuring secure backup.

Reuse is the ultimate aim of Databrary, because the richness and self-documenting nature of video make it uniquely suited for reuse. New researchers can reuse old research videos to ask questions beyond the scope of the original study. In my lab, for example, we collect videos of infants at play to examine the quantity and quality of natural locomotion and to understand the causes and consequences of locomotor exploration. We typically watch our videos with the sound off. Other developmental researchers could use the same videos of infants at play to study infant-caregiver social interaction, maternal responsiveness, linguistic input to infants, or object manipulation. Scholars with a more integrative bent could ask whether crawling and walking affect social interaction, whether social interaction and maternal responsiveness are related to linguistic input, or whether physical activity trades against object play. "Throw-away" sections of video for one lab (e.g., babies crying) could be gold mines for another lab (e.g., one whose researchers study the acoustics of infant cries). Trivial segments of video (e.g.,

those collected to ensure fidelity to a protocol or the reliability of an online code) or subsidiary videos (e.g., intended correlates of a larger question) could be reused as the primary data for a new study. Previously collected videos could be reused to grow one's sample size, expand the study population, or serve as preliminary data to show the feasibility of the approach in a grant submission.

Watching research videos provides insight into the details of procedures and video-coding rules that never make it into the method or data-coding sections of a published paper. Showing excerpts from research videos to students illustrates phenomena and makes findings come alive in a way that words alone cannot.

## Surmounting Ethical and Technical Barriers

Video sharing and reuse, however, present new technical and ethical challenges. One reason why Databrary is the first large-scale repository for open video-data sharing is that video contains identifiable data — participants' faces are viewable, their names are spoken aloud, and home data contain potentially identifiable information about the family. With guidance from legal counsel, representatives of ethical review boards and offices of sponsored programs, and information-technology experts, Databrary has addressed the ethical concerns about sharing identifiable data. Central to our policy framework is a code of conduct in which researchers promise to respect participants' wishes about sharing, to treat other people's data with the same high level of ethical care that they treat their own, and to be responsible for their students' and collaborators' use of other people's data. This code of conduct extends the zone of trust beyond the boundaries of individual laboratories to a community of like-minded authorized investigators. To become authorized, a researcher must be sponsored by the office of grants and contracts at their institution, which attests that they have researcher privileges and that their actions are governed by an ethical review board. Doctoral students, postdocs, and research staff must be sponsored by an authorized investigator who is responsible for their conduct.

We have developed a framework for requesting permission from adult participants and from children's parents to share and reuse their identifiable video data and associated metadata (birthdates, race/ethnicity, geographic location, etc.). The guiding premise is that raw research videos can be shared on Databrary if participants understand the potential risks to privacy and agree to share. Whereas consent to participate must be obtained *before* data are collected, we advise researchers to request participants' permission to share *after* the observational session has ended. Most developmental studies are fairly innocuous, and most participants readily agree to share. Parents of children with disabilities are typically eager to share in the hope that reuse will speed progress in understanding and treatment.

Researchers make the final decision about whether and when to share their study while respecting the release levels agreed to by the participants. The default release level (termed "Private") allows only the researchers named on the original protocol to access the data. Because Databrary is a secure facility with protections against unwarranted viewing or downloading, these data can be stored on Databrary to facilitate work within the lab group and to keep the data set intact. The other release levels include:

- **Authorized users**, which allows for sharing recordings and metadata with all authorized investigators on Databrary;
- **Excerpts**, which additionally allows authorized investigators to show selected excerpts from shared research videos in public settings for research or instructional purposes; and

- **Public**, which allows for open sharing (few researchers currently request consent for public sharing from their participants).

With the appropriate technical infrastructure and community mind-set for video sharing and reuse, researchers can render children's behavior in a tangible form, preserve it indefinitely, and exploit the richness of video to increase scientific transparency, accelerate the pace of scientific discovery about child learning and behavioral development, and facilitate insights into the causes of health and disease. The scientific contribution of a particular data set will no longer depend on the private activities of one researcher, but will instead benefit from the imagination of many researchers with different viewpoints. And, in a real and meaningful way, researchers will continue the tradition of devising new ways to make visible and comprehensible the changing form of behavior that is the stuff of developmental science.

To learn more about the Databrary project and how you can join the Databrary community, [click here](#). œ

## References and Further Reading

Adolph, K. E., Gilmore, R. O., Freeman, C., Sanderson, P., & Millman, D. (2012). Toward open behavioral science. *Psychological Inquiry*, *23*, 244–247.

Curtis, S. (2011). "Tangible as tissue": Arnold Gesell, infant behavior, and film analysis. *Science in Context*, *24*, 417–442.

Darwin, C. (1877). A biographical sketch of an infant. *Mind*, *2*, 285–294.

Gesell, A. (1952). Arnold Gesell. In E. G. Boring (Ed.), *A History of Psychology in Autobiography*, Vol. 4 (pp. 123–142). Worcester: Clark University Press.

MacWhinney, B. (2001). From CHILDES to TalkBank. In M. Almgren, A. Barreña, M. Ezeizaberrena, I. Idiazabal, & B. MacWhinney (Eds.), *Research on child language acquisition* (pp. 17–34). Somerville: Cascadilla.