Databrary: Enabling Sharing and Reuse of Research Video

Andrew S. Gordon (*drew@databrary.org*), Rick O. Gilmore, Karen E. Adolph, David S. Millman, Lisa Steiger, and Dylan A. Simon

JCDL 2015

Abstract

Video and audio recordings serve as a primary data source in many fields, especially in the social and behavioral sciences. Recordings present unique opportunities for reuse and reanalysis for novel scientific purposes, but also present challenges related to respecting the privacy of individuals depicted. Databrary is a web-based service for sharing and reusing the video data created by researchers in the developmental and learning sciences. By investigating how researchers organize, analyze, and mine their own recordings, we have implemented a system that empowers researchers to capture, store, and share recordings in a standardized way. The Databrary service promotes storage, management, sharing, and reuse of research data, controls access privileges to restricted human subject data, and facilitates browsing and discoverability of datasets.

Access Restrictions

Databrary ensures that identifiable information in videos can shared under four different release levels:

Private: Content has not been shared and is restricted to collaborators

Authorized Users: Content is restricted to authorized Databrary users

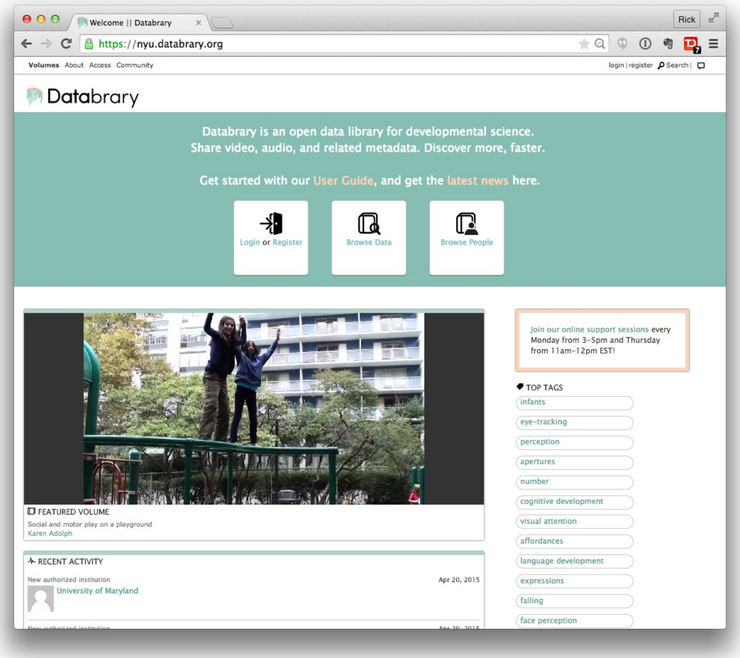
Excerpts: Content is restricted to authorized Databrary users, who may use it in presentations for informational or educational purposes

Public: Content is available openly to anyone

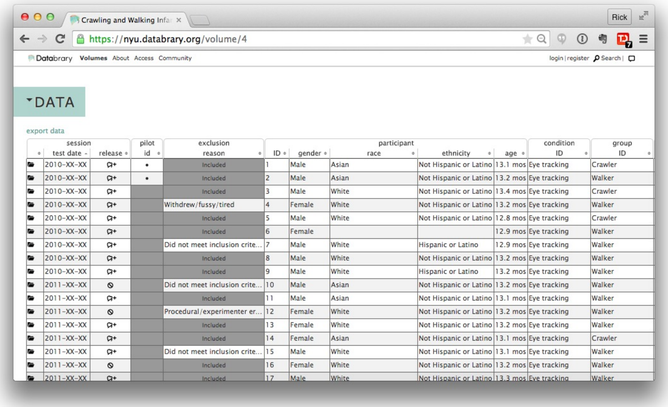
What Can Be Shared

Databrary can house video, audio, PDF, spreadsheet, image, and text-based files along with associated metadata and annotations. Video files are transcoded into standard, HTML5-compatible formats, currently H.264+AAC in MP4.

Databrary.org

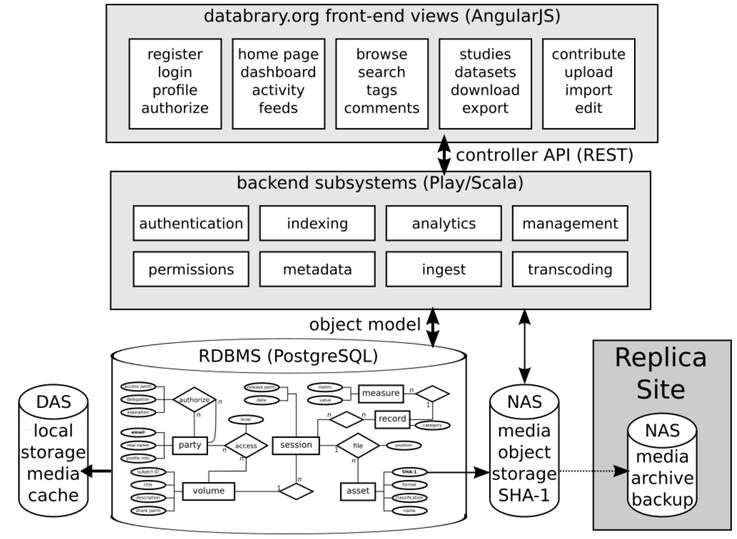


Session-level Data Management

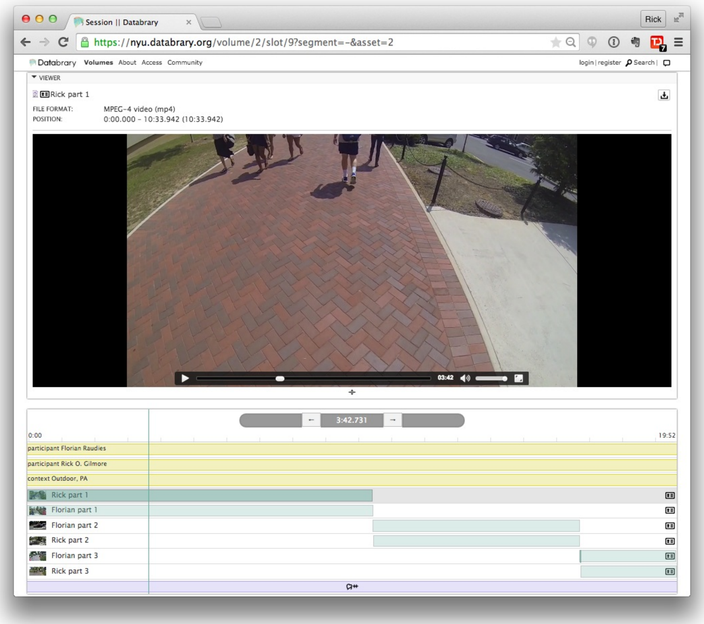


Technology Stack

Databrary is built on PostreSQL using the Scala Play framework and AngularJS. Data are preserved indefinitely in a secure data storage facility at NYU. There is no cost to use the system; a sustainability model is under development. All code is GPL on GitHub.



Timeline



Acknowledgments

Databrary is based on work supported by the National Science Foundation under Grant No. BCS-1238599 and the Eunice Kennedy Shriver National Institute of Child Health and Human Development under Cooperative Agreement U01-HD-076595. 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 or the Eunice Kennedy Shriver National Institute of Child Health and Human Development.