Sharing displays and data from vision science research with Databrary

R.O. Gilmore, K.E. Adolph, D.S. Millman, L. Steiger, and D.A. Simon

Motivation

The open sharing of data and workflows from scientific research has gained increasing attention in many fields and has become a priority for research funders. Databrary (databrary.org) is a digital data library specialized for storing and sharing video and research data and metadata. Vision scientists often present temporally varying visual displays to human or non-human animal observers and may collect gaze position, kinematic data, or other physiological recordings simultaneously. Some vision scientists collect identifiable video recordings of participants for use in off-line tagging or coding. Databrary may be especially well suited for these use cases because of the library's focus on uploading, organizing, storing, and sharing research data and metadata collected in the course of behavioral resesearch. Databrary has also developed policies that enable identifiable information to be shared with the permission of research participants.

Privacy & Security

Videos may contain identifiable information such as faces, voices, and images of places. Databrary ensures that identifiable information can be shared in two ways:

▸ Restricting Access: Only those with independent researcher status at an institution may apply for authorization. Applicants promise to uphold ethical principles and sign an access agreement. The institution co-signs the agreement.

▸ Securing Permission to Share: Identifiable data may only be shared if participants have given permisssion. Databrary has developed a release template whose language may be adapted for specific research protocols (https://databrary.org/access/policies.html).

What Can Be Shared

Databrary 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. Databrary stores other data in native formats (e.g., .doc, .docx, .xls, .xlsx, .txt, .csv, .pdf, .jpg, .png).
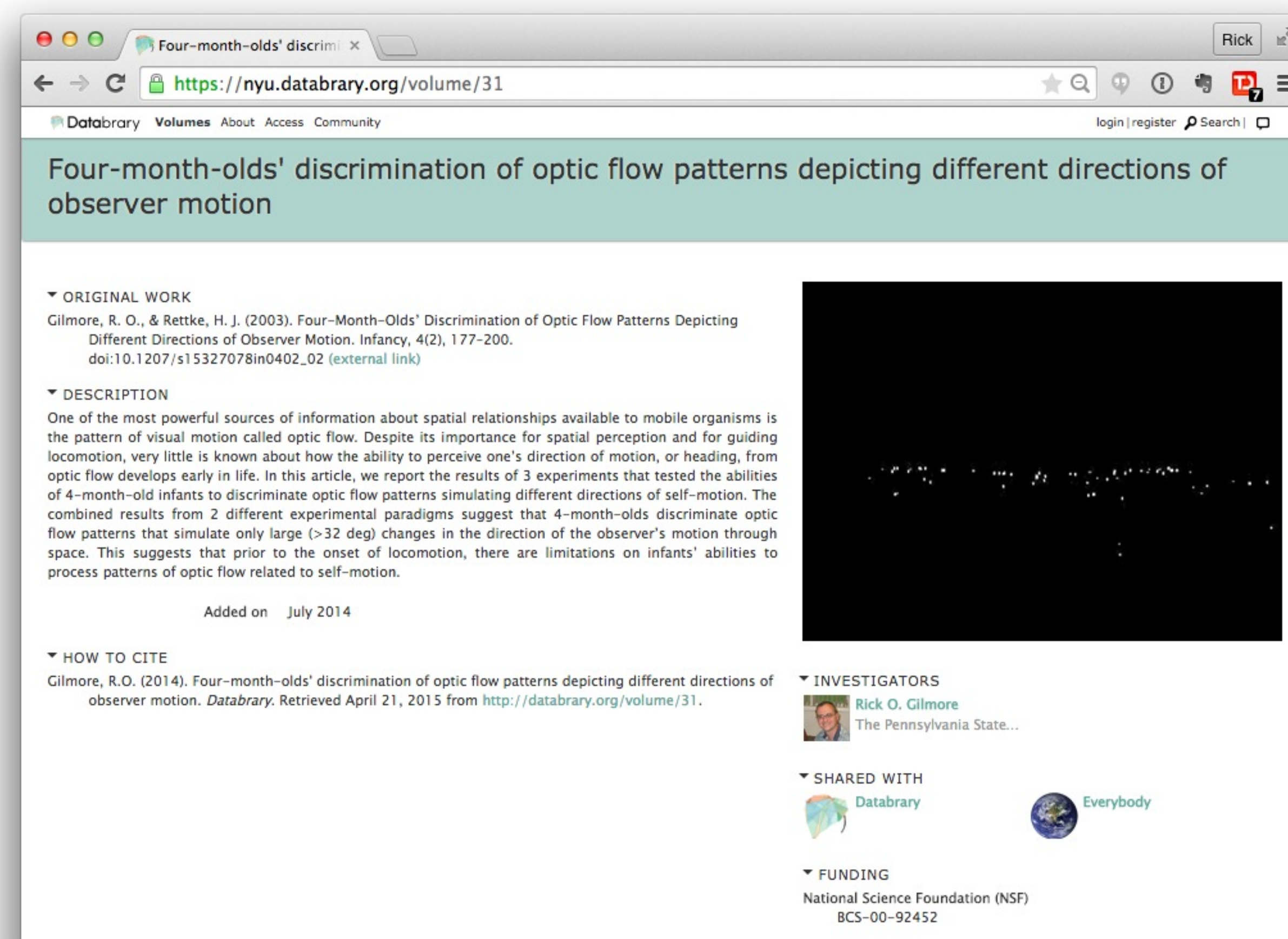
Technology Stack

Databrary is built on a PostreSQL database using the Scala Play framework and JavaScript. Data are preserved indefinitely in a secure data storage facility at NYU managed by the university's central information technologies organization. Central IT staff handle storage, network, and backup systems. NYU does routine tape backups that are stored off site. There is no cost to use the system; an institutional subscription model is under development.

Databrary.org
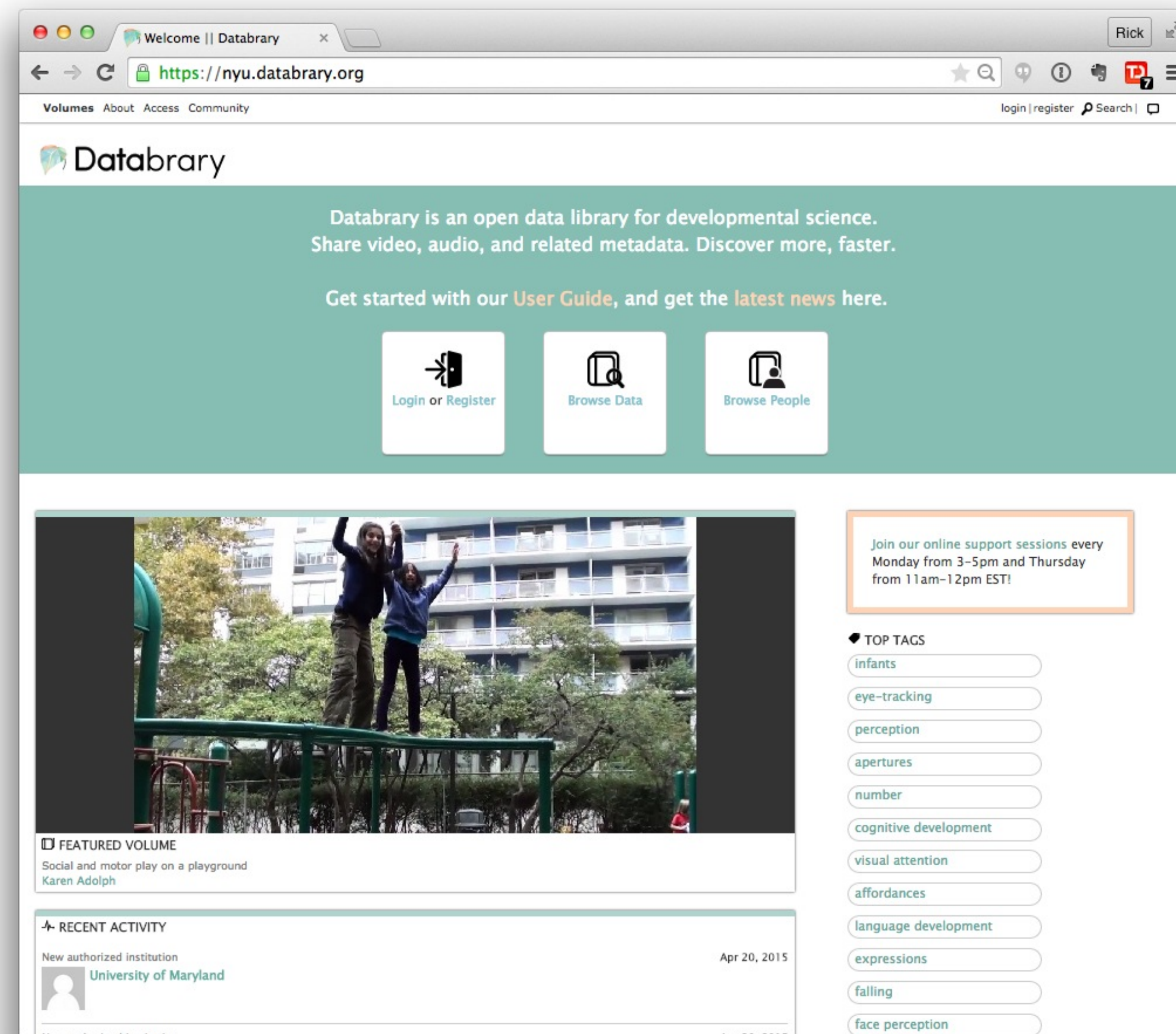
Databrary's splash page shows a timeline of recent activity, a set of clickable tags that link to studies with those search terms, a featured dataset, and links that enable users to search for data or manage their own data.
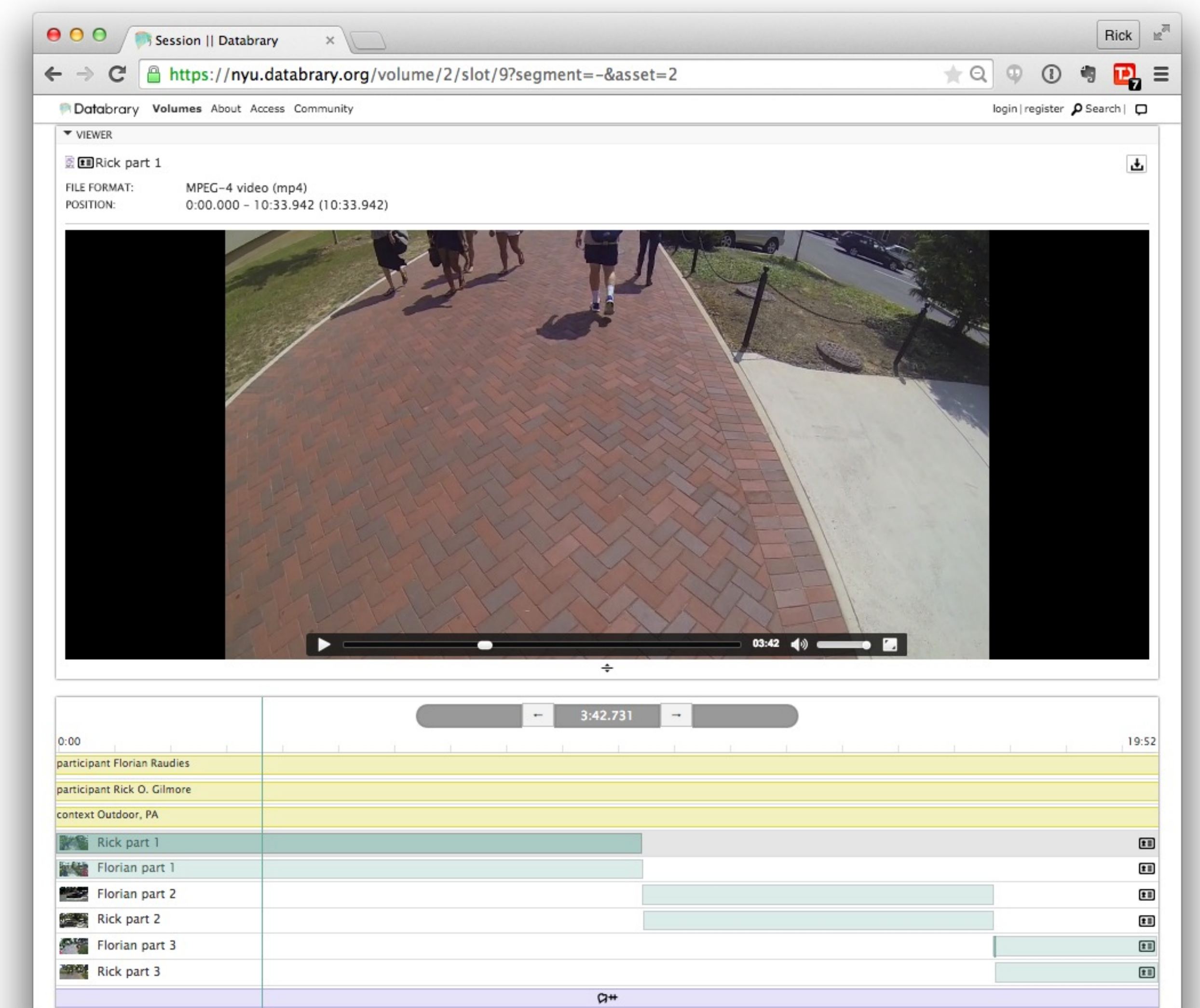
Each Study/Dataset Has Its Own Page

Data Management for Ongoing Projects

Users may upload video and related data *as they collect it*, recording information about individual data collection sessions such as participant characteristics, sharing permission levels, and study conditions.

Organize, View. Tag Data in Timeline

Within a session, data are represented on a timeline that reflects when events occurred. This is particularlly useful for studies with multiple camera views (e.g., eye-tracking) or with combined behavioral and image/display streams.