Tools for reproducible developmental science

Rick O. Gilmore[1,2], Swapnaa Jayaraman[2], Shohan Hasan[2], Jesse Lingeman[2]

[1] The Pennsylvania State University
[2] Databrary.org

Overview

Many fields of scientific research face daunting challenges of reproducibility. The science of infant development is no exception. We describe a set of free and open source software tools we have developed that support fully reproducible data collection, cleaning, visualization, and analysis workflows.

Case 1: PLAY Project

The Play & Learning Across a Year (PLAY) project (Adolph, Tamis-Lemonda, & Gilmore, 2020) is a collaborative research initiative by 65 researchers from 45 universities across the United States and Canada. PLAY focuses on recording and revealing the behaviors of infants and mothers during natural activity in their homes.

Figure 1 shows the project's multi-step workflow.

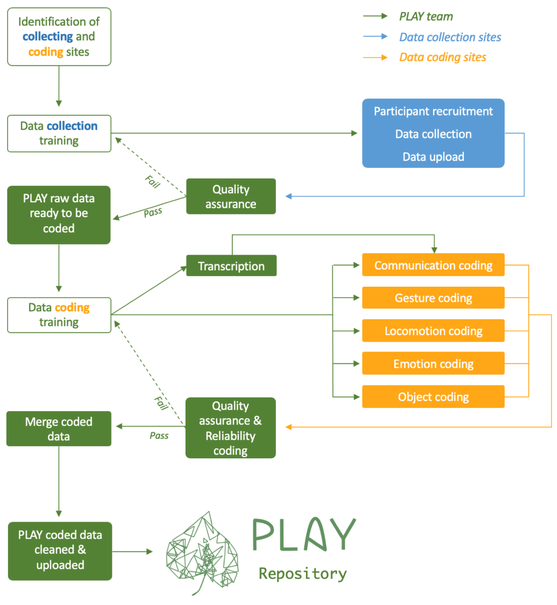

Figure 1: PLAY workflow

Quality assurance (QA)

Data collection labs upload 1) videos and other documents to Databrary and 2) survey information to the KoBoToolbox server. Staff then run an R script to check that all data were entered correctly. Figures 2 and 3 show some of the outputs.

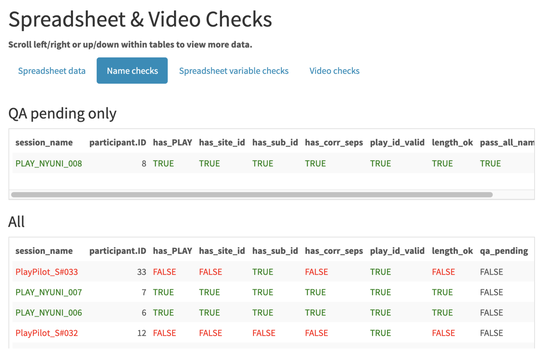

Figure 2: Checking PLAY file names

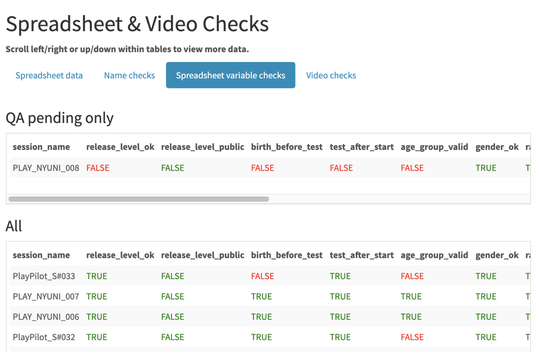

Figure 3: Checking PLAY spreadsheet data

In addition to using the `databraryapi` package (Gilmore, 2020), we have also been developing packages in Python (Hasan, Nezzar, & Lingeman, 2020; Lingeman & Hasan, 2020) as part of our work to automate workflows on PLAY.

Case 2: Databrary.org

Databrary (*Databrary*, 2020) is a restricted access data library offering scientists a secure way to store and share identifiable research data, especially video and audio recordings. Databrary has an application program interface (API). Using scripts that call the API (Gilmore, 2020; Lingeman & Hasan, 2020), researchers can write **reproducible** code that gathers data from Databrary and visualizes or analyzes it.

Charting Databrary's growth

Every week, the Databrary staff run an R script that generates an HTML-formatted report (see Figure 4) about the number of users, institutions, and new projects.

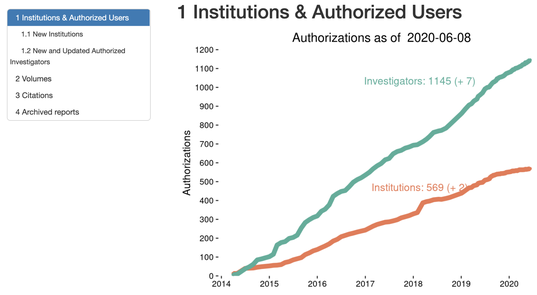

Figure 4: Charting growth in Institutions and Investigators

The `databraryapi::get_db_stats()` from the `databraryapi` package retrieves most of the critical data for this report.

Conclusions

Using R (R Core Team, 2019), R Markdown (Allaire et al., 2020), Python, GitHub, KoBoToolbox, Box, and Databrary (*Databrary*, 2020), we are able to create reproducible workflows for complex projects like PLAY and Databrary. Our code can be found on GitHub (Gilmore, 2020; Gilmore & Seisler, 2020), and we are happy to work with researchers who would like to make use of it.

Support

We acknowledge support from the National Institutes of Health, the National Science Foundation, the Alfred P. Sloan Foundation, the James S. McDonnell Foundation, the Society for Research in Child Development, the LEGO Foundation, and the Defense Advanced Research Projects Agency.

References

Adolph, K. E., Tamis-Lemonda, C. T., & Gilmore, R. O. (2020). *The play & Learning Across a Year (PLAY) Project*. Retrieved from https://play-project.org

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., … Iannone, R. (2020). *R Markdown: Dynamic documents for r*. Retrieved from https://CRAN.R-project.org/package=rmarkdown

Databrary: A restricted access data library. (2020). Retrieved from https://databrary.org

Gilmore, R. O. (2020). *Databraryapi: An r package for Databrary*. Retrieved from https://github.com/PLAY-behaviorome/databraryapi

Gilmore, R. O., & Seisler, A. R. (2020). *Databrary analytics*. Retrieved from https://github.com/gilmore-lab/databrary-analytics

Hasan, S., Nezzar, R., & Lingeman, J. M. (2020). *Pyvyu: Python commands for interacting with databrary*. Retrieved from https://github.com/databrary/pyvyu

Lingeman, J. M., & Hasan, S. (2020). *Db-playmate*. Retrieved from https://github.com/databrary/db-playmate

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/