LearnLab's DataShop: A Data Repository and Analytics Tool Set for Cognitive Science

Kenneth R. Koedinger, John C. Stamper, Brett Leber, Alida Skogsholm

Human-Computer Interaction Institute, Carnegie Mellon University

---

In *What Should Be the Data Sharing Policy of Cognitive Science?* Pitt and Tang (2013) make the case for an open data-sharing policy in Cognitive Science and highlight the use of online data repositories to store and share raw research data. One such data repository is the LearnLab DataShop (http://pslcdatashop.org) hosted at Carnegie Mellon University. DataShop is part of LearnLab, a NSF-funded Science of Learning Center started in 2004. DataShop is a major resource for researchers in educational data mining and the learning sciences, including the educational arm of Cognitive Science. DataShop is both an open repository of learning data and a web application for performing exploratory analyses on those data. DataShop specializes in data on the interaction between students and educational software, including online courses, intelligent tutoring systems, virtual labs, online assessment systems, collaborative learning environments, and simulations. As of March 2013, DataShop offers 385 datasets under 116 projects. Across these data sets, there are 97 million software-student transactions, representing over 238,000 student hours.

A key feature relevant to the Cognitive Science community is DataShop's set of tools for exploring cognitive models both visually and statistically. In DataShop, a cognitive model is a mapping between hypothesized "knowledge components"—a more general term for skill, concept, schema, production rule, misconception, or facet—and steps in the procedural completion of an online activity. A researcher can define a hypothesized model in a spreadsheet and upload it to DataShop, where it becomes available for analyses. Visual analyses include learning curves and an error report, while statistical analyses include a logistic regression model that describes how well alternative cognitive models predict student learning.

DataShop has been valuable to both primary and secondary researchers in the learning sciences fueling over 100 secondary analysis studies and associated papers. For

---

Correspondence should be sent to K. R. Koedinger, Human-Computer Interaction Institute, Carnegie Mellon University, Newell-Simon Hall 3601, 5000 Forbes Ave., Pittsburgh, PA 15213. E-mail: koedinger@cs.cmu.edu

researchers who add their data to DataShop, access controls allow them to keep the data entirely private, share selectively, or make the dataset accessible to all registered users. DataShop enables secondary research by allowing registered users to view public datasets and request access to private ones.

Researchers in the Cognitive Science community have used DataShop data and tools in their analyses. For example, at least eight papers from past Cognitive Science proceedings or journal articles make use of data in DataShop (see http://pslcdatashop.org/about/cogsci.html).

We would be happy to support greater use of LearnLab's DataShop within the Cognitive Science community. More generally, DataShop provides examples of open data-sharing strategies and policies that could be the subject of community reflection toward the goals expressed in Pitt & Tang (2013).

References

Pitt, M. A., & Tang, Y. (2013). What should be the data sharing policy of cognitive science? *Topics in Cognitive Science*, *5*, 214–221. doi: 10.1111/tops.12006.