

What is DataShop?

DataShop is an open data repository and set of associated visualization and analysis tools for fine-grained, longitudinal learner data.

DataShop enables you to:

- Import and export datasets
- Visualize student performance
- Perform learning curve analysis
- Discover cognitive models of student knowledge
- Use data to improve instruction
- Programmatically retrieve data via web services

What data is in DataShop?

DataShop contains vast amounts of free data.

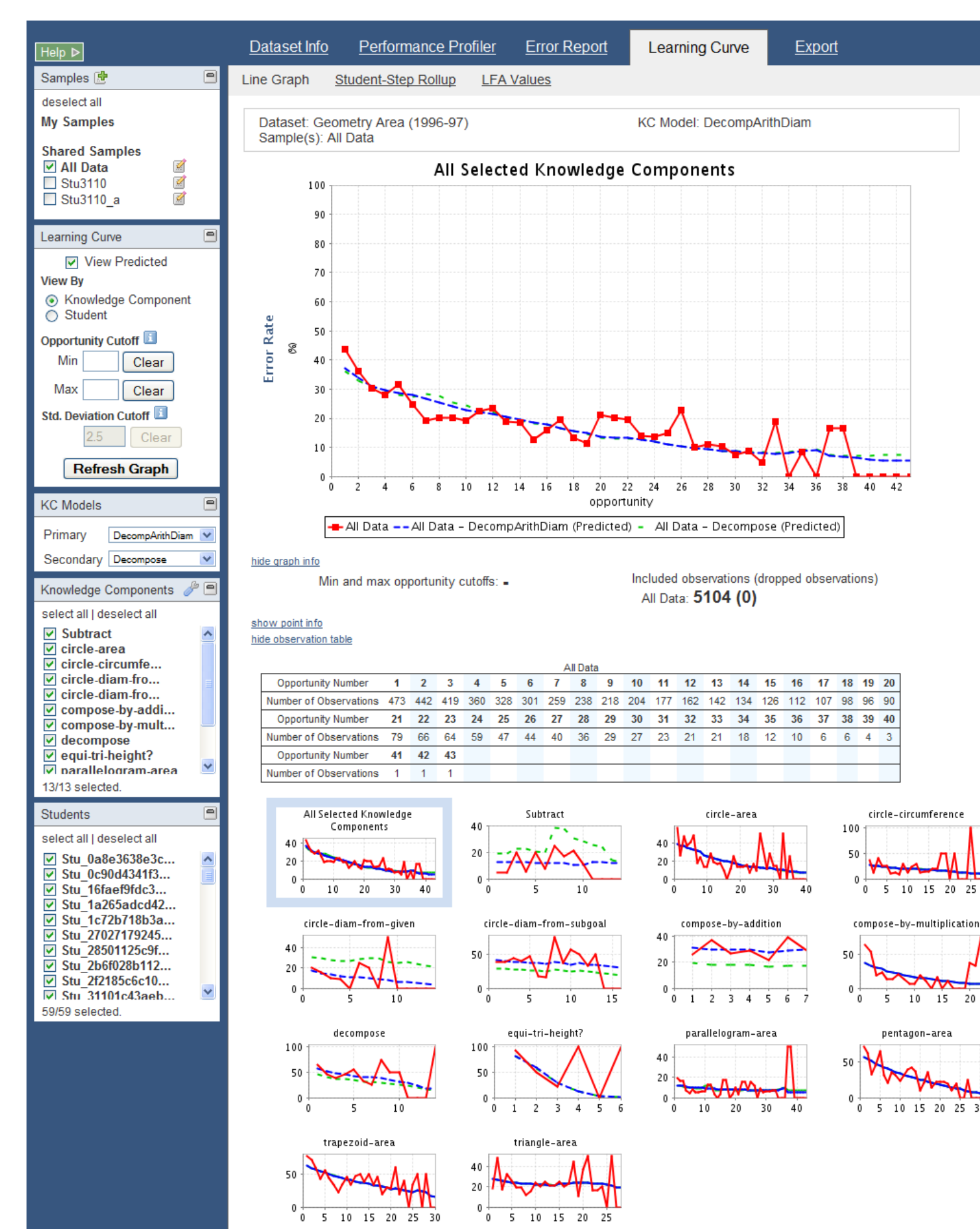
Domain	Datasets	Student Actions	Students	Student Hours
Language	60	2,838,843	3,263	5,661
Math	73	32,201,021	8,887	99,084
Science	30	3,675,770	3,890	19,830
Other	30	4,789,826	11,058	28,698
Total	193	43,505,460	27,098	153,273

Over 150,000 student hours!

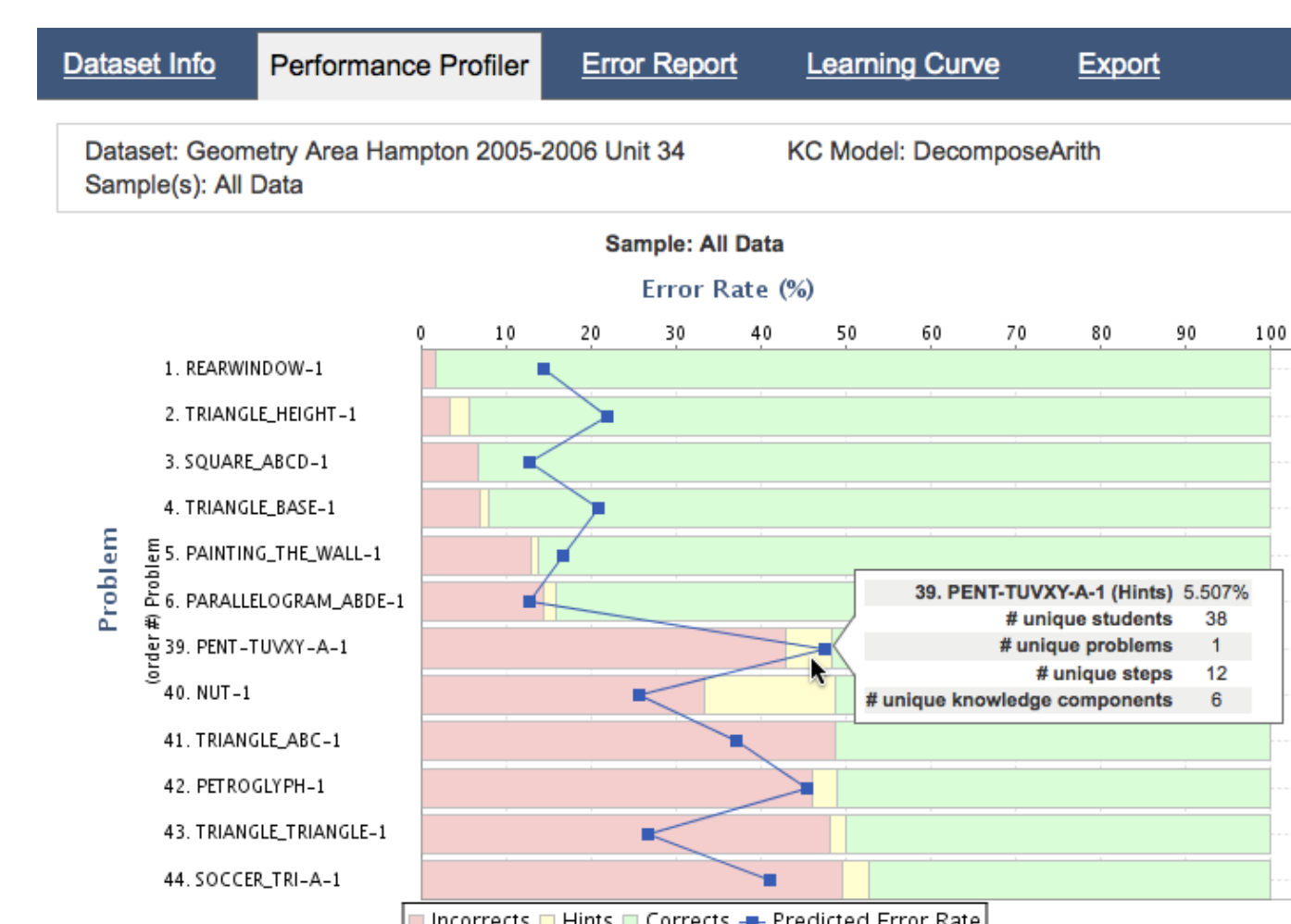
What are some analyses that have used DataShop data?

- Adapting to when students game an intelligent tutoring system (Baker et. al.)
- Self-explaining in the classroom: Learning curve evidence (Hausmann and VanLehn)
- Optimizing knowledge component learning of Chinese using a dynamic structural model of practice (Pavlik et. al.)
- Extending learning factors analysis to model reading transfer (Leszczenski and Beck)
- Exploring alternative methods for error attribution in learning curves analyses (Nwaigwe et. al.)

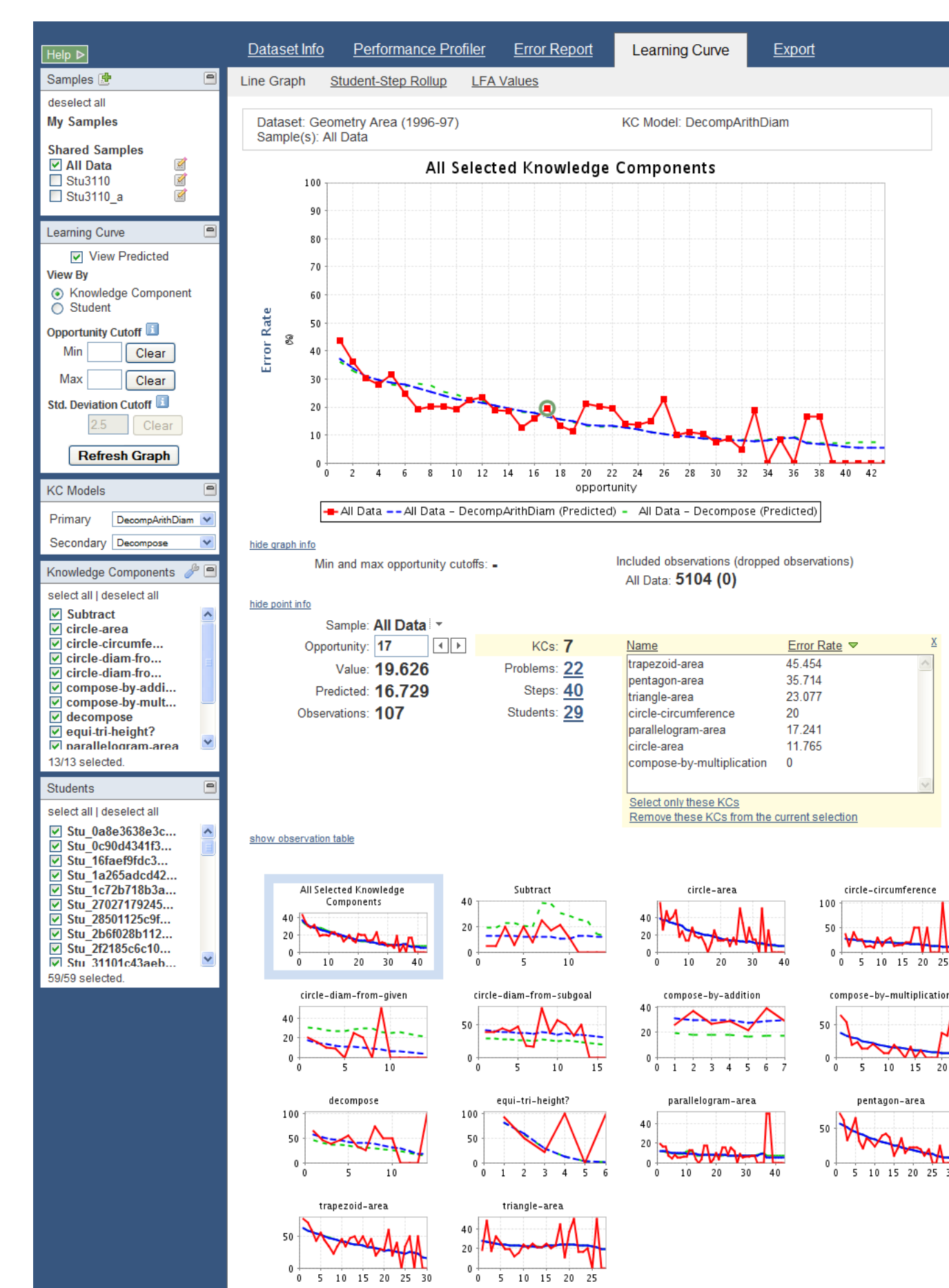
Current features



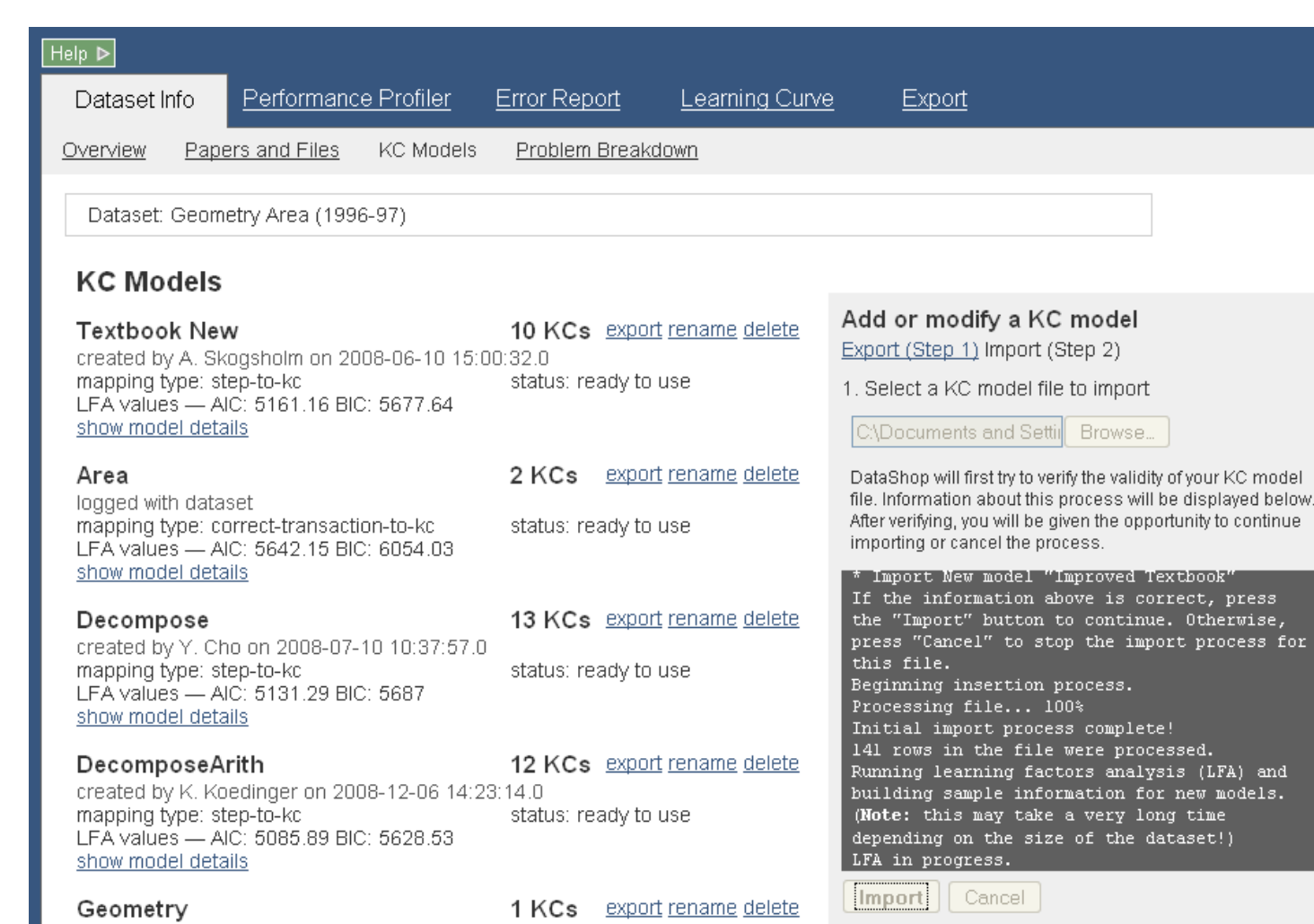
Knowledge component model analysis with learning curves



Performance Profiler tool for exploring the data



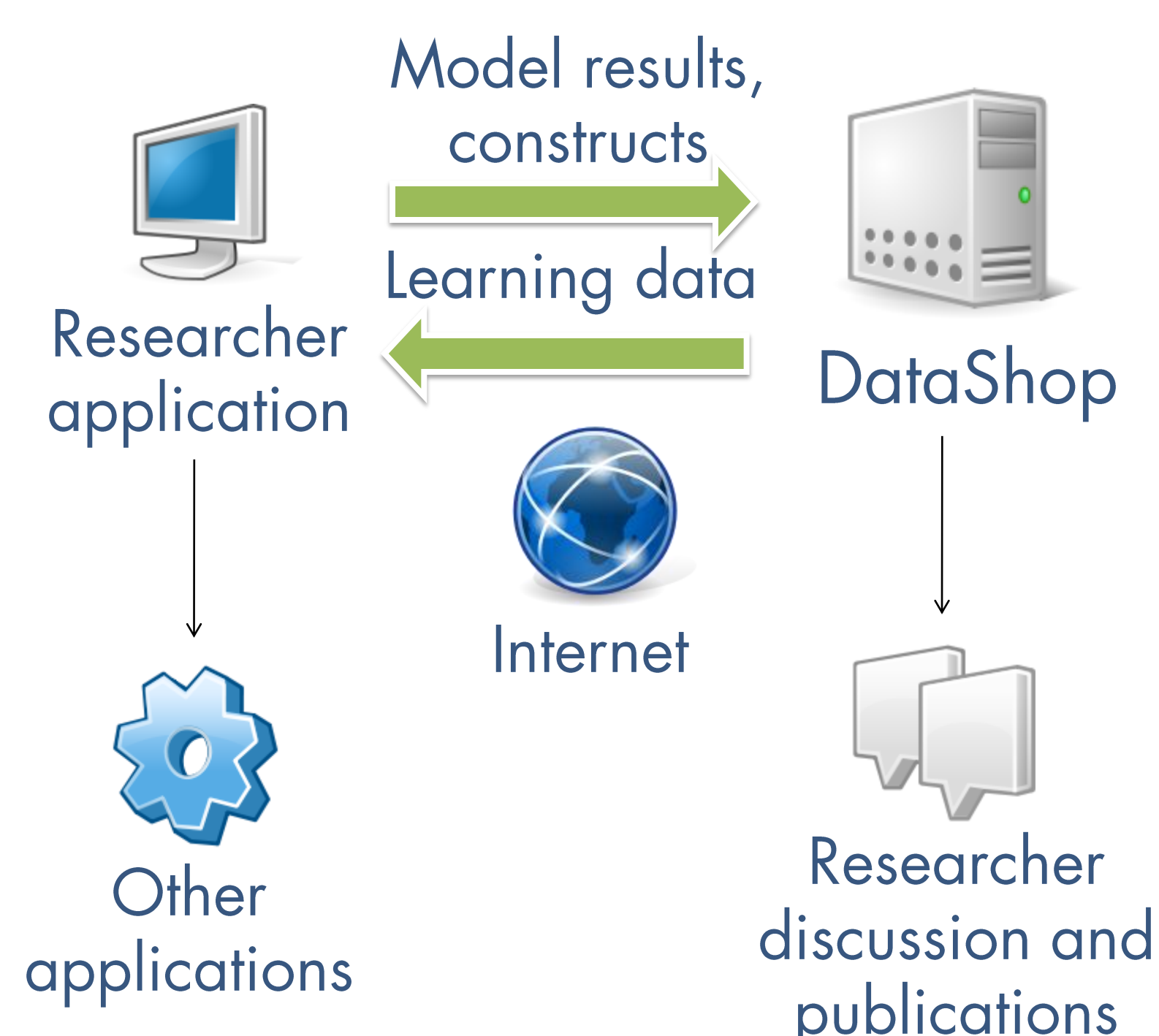
Learning curve point decomposition



Easy knowledge component model creation

New features

Web services architecture



Coming in future releases

- Easy-to-use API for using web services to add fields to an existing dataset
- Ability to upload models to DataShop for dissemination, making it easier to build on each other's work
- Improved scalability and speed: accommodate a planned 100x increase in quantity of data in DataShop
- Support for timestamps with millisecond precision
- Error bars on learning curves

KDD Cup 2010 Educational Data Mining Challenge

> <http://pslcdatashop.org/KDDCup>

Awarded to the PSLC and DataShop, this year's challenge asked participants to predict student performance on mathematical problems from logs of student interaction with Intelligent Tutoring Systems.

The KDD Cup is the premier data mining competition that attracts top data mining and machine learning experts from industry and academia.

The competition ended on June 8. There were:

655 registered participants

130 participants who submitted predictions

3,400 submissions

The datasets used for the challenge were:

Dataset	Students	Steps	File size
Algebra I 2008-2009	3,310	9,426,966	3 GB
Bridge to Algebra 2008-2009	6,043	20,768,884	5.43 GB

The competition addressed questions of both scientific and practical importance. Improved models could be saving millions of hours of students' time (and effort) in learning algebra. These models should both increase achievement levels and reduce time needed to learn.