# LabFlow-1: a Database Benchmark for High-Throughput Workflow Management*


**Anthony J. Bonner**[1] bonner@db.toronto.edu

**Adel Shrufi**[1] shrufi@db.toronto.edu

**Steve Rozen**[2] steve@genome.wi.mit.edu

[1]University of Toronto
Department of Computer Science
10 King's College Rd
Toronto, ON, Canada
M5S 1A4

[2]Whitehead/MIT
Center for Genome Research
One Kendall Square
Building 300, Floor 5
Cambridge, MA 02139, USA


October 8, 1996


**Abstract**

Workflow management is a ubiquitous task faced by many organizations, and entails the coordination of various activities. This coordination is increasingly carried out by software systems called *workflow management systems* (WFMS). An important component of many WFMSs is a DBMS for keeping track of workflow activity. This DBMS maintains an audit trail, or event history, that records the results of each activity. Like other data, the event history can be indexed and queried, and views can be defined on top of it. In addition, a WFMS must accommodate frequent workflow changes, which result from a rapidly evolving business environment. Since the database schema depends on the workflow, the DBMS must support dynamic schema evolution. These requirements are especially challenging in high-throughput WFMSs—*i.e.*, systems for managing high-volume, mission-critical workflows. Unfortunately, existing database benchmarks do not capture the combination of flexibility and performance required by these systems. To address this issue, we have developed *LabFlow-1*, the first version of a benchmark that concisely captures the DBMS requirements of high-throughput WFMSs. LabFlow-1 is based on the data and workflow management needs of a large genome-mapping laboratory, and reflects their real-world experience. In addition, we use LabFlow-1 to test the usability and performance of two object storage managers. These tests revealed substantial differences between these two systems, and highlighted the critical importance of being able to control locality of reference to persistent data.



Portions of this report appear in:

- *Proceedings of the International Conference on Extending Database Technology (EDBT)*, 1996.
- *The OOPSLA Workshop on Object Database Behavior, Benchmarks, and Performance*, 1995.
- *The NSF Workshop on Workflow and Process Automation in Information Systems*, 1996.

These papers and the entire report are available at the following web page:

`http://www.db.toronto.edu:8020/people/bonner/bonner.html`

*This work was supported by funds from the U.S. National Institutes of Health, National Center for Human Genome Research, grant number P50 HG00098, and from the U.S. Department of Energy under contract DE-FG02-95ER62101.