ProgSnap 2: Towards a Standard Representation for Programming Process Data

David Hovemeyer, York College of Pennsylvania, dhovemey@ycp.edu
Kelly Rivers, Carnegie Mellon University, krivers@andrew.cmu.edu

Contributors

ProgSnap 2 is the work of a dedicated team! Significant contributors, in alphabetical order by last name:

- Cory Bart
- Brett Becker
- Luke Gusukuma
- David Hovemeyer
- Ayaan Kazerouni
- Andrew Petersen
- Thomas Price
- Kelly Rivers

ProgSnap 2

What it is:

- Standardized concrete data representation for programming process data (code edits, code executions, etc.)
- Programming language- and environment-agnostic
- File-based, main event table suitable for direct analysis using stats software, Python scripts, etc.

What it isn’t:

- Not suitable for truly large datasets (i.e., Blackbox)

Process

Working group members have experience in both producing and consuming/analyzing programming process data

So, ProgSnap 2 is informed by experiences from real systems and research

The standard will be validated by

1. Creating datasets from multiple systems/contexts
2. Doing analysis on datasets originating from multiple systems/contexts

Overview of a ProgSnap 2 dataset

Dataset metadata: CSV file specifying characteristics of overall dataset

Main event table: collection of records representing programming process events

Code states: code snapshots associated with specific programming process events

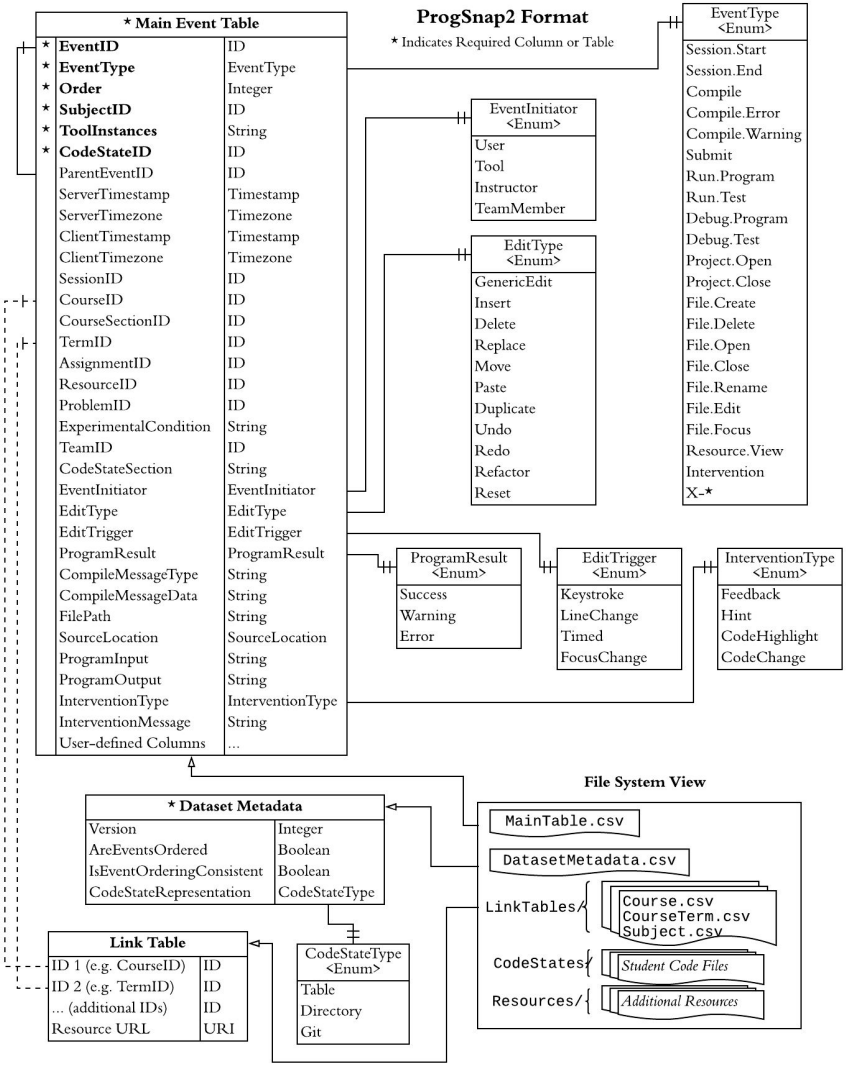
Link tables: associate combination of IDs with resource blobs, e.g. (TermID,CourseId)→information about course offering

Resources: data blobs associated with events (from main event table) or ID combinations (from link tables)

Data model

Diagram by Cory Bart

See: Thomas W. Price, David Hovemeyer, Kelly Rivers, Austin Cory Bart, Andrew Petersen, Brett A. Becker, and Jason Lefever. ProgSnap 2: A flexible format for programming process data. In *Proceedings of the 2nd Educational Data Mining in Computer Science Education (CSEDM) Workshop*, Tempe, AZ, 2019.



Where we are

We have a draft specification: see website (https://cssplice.github.io/progsnap2)

- Some remaining work to do, but is concrete enough to implement/use now

Prototype exporters for several systems are implemented

What happens next

Create and share example datasets

Make sure that exporters are following the specification (fix specification as needed to resolve ambiguities)

Create tooling for analyzing ProgSnap 2 datasets

We welcome participation!

> If you have data you could export, or if you have an interesting analysis to perform on exported data—contact us!