IRIS-HEP / IMPRESS-U Summer 2026 Fellows Program Project Proposal

Modifying existing TTrees in Uproot

Applicant: Jokūbas Mačiulis

Mentor: Andres Rios-Tascon

Project duration: June 22, 2026 – September 2026 (10 weeks)

Project Description

ROOT is the standard data storage and analysis framework used in High Energy Physics (HEP), developed at CERN [1]. Data is stored in .root files as TTrees — tree-like structures where branches represent data columns (e.g. particle properties) and entries represent individual collision events. Uproot is a widely-used Python library in the scikit-hep ecosystem that enables reading and writing ROOT files without requiring CERN's ROOT software, making HEP data accessible in modern Python data science workflows [2].

While Uproot supports creating new ROOT files and writing new TTrees, it currently lacks the ability to modify existing TTrees in place. This limitation affects workflows such as backfilling data — for example, adding a new trigger branch with default values to previously recorded data. Uproot does provide uproot.update for adding new objects to existing ROOT files [3], however modifying TTrees themselves is not yet supported.

As a previous IRIS-HEP fellow, Zoë Bilodeau took on the challenge of implementing the ability to add new branches (columns) to existing TTrees, resulting in PR #1155 [4]. This contribution brought the feature close to completion, but a few corner cases remain that need to be tested and resolved before the PR can be merged. Adding new entries (rows) to existing TTrees is a separate, entirely new problem with no prior implementation to build on.

Planned Work

The project has two phases:

Phase 1 — Completing TBranch addition (columns): Review and understand Zoë's implementation in PR #1155 [4]. Identify remaining corner cases, write comprehensive tests, debug issues, and work toward merging the PR into the main Uproot codebase.

Phase 2 — Implementing TTree entry addition (rows): Design and implement the ability to append new entries/events to existing TTrees. This involves understanding ROOT's binary basket format, correctly appending data, and updating all relevant metadata. This is entirely new functionality with no prior implementation to build on.

Both phases will include writing tests, documentation, and ensuring compatibility with CERN's ROOT software.

Software Deliverables

- Merged PR #1155 (adding columns to existing TTrees) with full test coverage.
- New implementation for adding rows/entries to existing TTrees.
- Unit tests for all new functionality.
- Documentation and usage examples.

- If time permits: equivalent functionality for RNTuples, ROOT's next-generation data format.

Timeline

Weeks 1–2: Set up development environment, study Uproot codebase and PR #1155.

Weeks 3–4: Debug and test remaining corner cases in PR #1155, prepare for merge.

Weeks 5–6: Design approach for row-addition feature, study ROOT basket format.

Weeks 7–9: Implement and test row-addition functionality.

Week 10: Final testing, documentation, and cleanup. If time permits: equivalent functionality for RNTuples.

References

[1] CERN ROOT framework: https://root.cern

[2] Uproot documentation: https://uproot.readthedocs.io

[3] uproot.update: https://uproot.readthedocs.io/en/latest/uproot.writing.writable.update.html

[4] PR #1155 — Add branches to existing TTrees: https://github.com/scikit-hep/uproot5/pull/1155