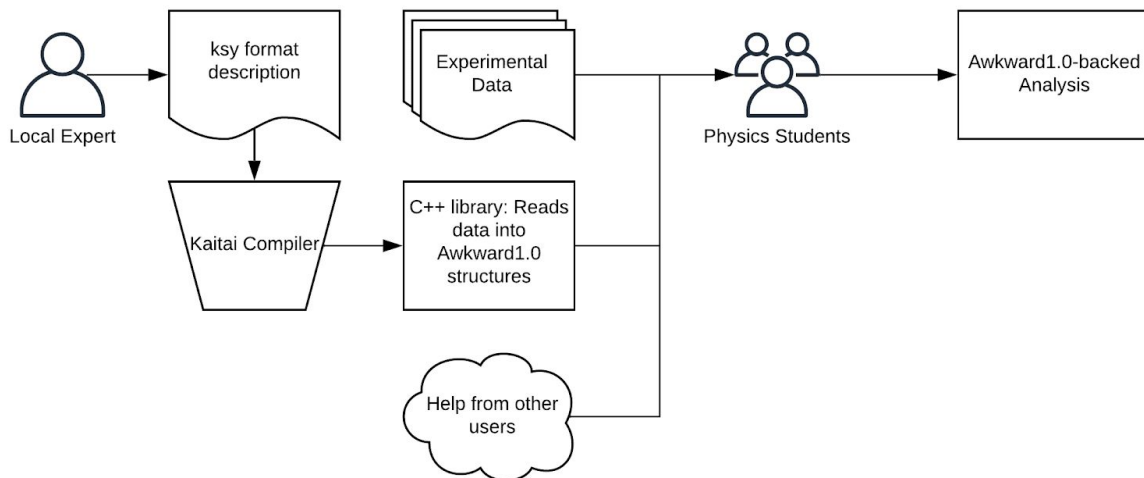# Awkward1.0

Awkward1.0 will provide data structures in C++ and python that allow efficient event-by-event analysis - but only for data in standard formats such as HDF5 and ROOT.  Scientists whose data is in a custom binary format cannot use the event-friendly Awkward1.0 library.

# Kaitai Struct

Kaitai Struct is an open-source project that generates data-reading code in target languages like python, C++, and Java given a description of file format.  We've found that the Kaitai Struct language is able to describe XIA and MIDAS-format data, two common data formats in the nuclear physics and dark matter communities.



# The Project

We propose to write another "target" language to the Kaitai Struct project: we would write code that generates C++ code that stores data in Awkward1.0 objects.  While Kaitai currently has a compiler that targets C++, it stores data in objects that are awkward to use for event-based physics analysis.  Adding an Awkward1.0 target to Kaitai-Struct would allow scientists to use Awkward1.0 for data with any format, provided only that the format is described with the kaita-struct language.  This would increase the potential user base of Awkward1.0 to smaller-scale collaborations that often use custom data acquisition systems that do not write to standard data formats.

To make this project feasible to complete in half a year, the katai extension would target only the subset of features needed to describe a sample dataset from an XIA digitizer.  This dataset is widely used in the nuclear physics community and represents a reasonable range of features, such as repeating structures, so that even with no additional development the code would be useful for many custom scientific data formats.

# Timeline

The proposed plan of work is done in four stages: (1) Becoming familiar with the Awkward1.0 API, (2) developing an Awkward1.0 kaitai target for example data with an extremely simple binary format, (3) improving the Awkward1.0 kaitai target to read XIA data, and (4) final testing and documentation.

My work will focus on updating kaitai-struct to target Awkward1.0. This requires writing code that will generate data-reading code in the target language - here, C++ code that will read data into Awkward1.0 structures. To make this meta-programming easier, I will first write C++ code that reads the target data into Awkward1.0 structures. This hand-written C++ code can serve as a template when updating kaitai-struct to automatically generate the data-reading code.

To make the code as broadly usable as possible, we will release python packages on test-pypi that allow reading the example data and the XIA data through pybind11 and write documentation for how to release python modules to read custom data. Mary Span has expertise in releasing python packages and will lead the effort on these releases. In addition, Nikki Ramirez will take lead on developing tests for implemented kaitai features.

Below is a proposed timeline for the project which breaks down the various steps in more detail.

| Task | Assignee | Delivery date |
|---|---|---|
| **Awkward and Kaitai** | | |
| Test awkward1.0 API | Jim | 10/04/19 |
| Update Kaitai test documentation | Nikki | 10/24/19 |
| C++ code to interface hdf5 data with awkward1.0 | Nikki | 10/17/19 |
| **Example Raw Data ("Toy Data")** | | |
| Identify minimal set of features needed in awkward1.0 to access toy data | Josh | 12/20/19 |
| C++ code to interface toy data with awkward1.0 | Josh | 12/13/19 |
| Implement awkward1.0 Kaitai target for toy data | Josh | 01/17/20 |
| publish python package for toy data on test-pypi | Mary Span | 02/28/20 |
| **XIA Data** | | |
| Review XIA kaitai-struct and update documentation | Josh | 11/15/19 |
| Write C++ code that reads in XIA data into awkward1.0 data structure | Josh | 02/21/20 |
| Identify minimal set of awkward1.0 features to access XIA data | Josh | 03/06/20 |
| Implement awkward1.0 Kaitai target for XIA | Josh | 04/03/20 |
| **Wrapping Up** | | |
| Try interfacing C++ code in Python using pybind | Josh | 01/31/20 |
| Update Kaitai tests for implemented features | Nikki | 04/17/20 |
| Presentation at DIANA/HEP meeting | Josh | 04/24/20 |
| Publish python package for XIA data on test-pypi | Mary Span | 05/01/20 |