

CAFAna overhaul to support updated StandardRecord: Status and Plans

María Martínez-Casales
Fermilab

Jeremy Wolcott
Tufts University

December 15, 2023

Reminder: StandardRecord overhaul

Public Attributes

| | | | | | |
|------------|--------------------|-------|----------------------|--------|-------------------------|
| int | meta_run | float | numu_pid | float | eRecoPim |
| int | meta_subrun | int | LongestTrackContNumu | float | eRecoPi0 |
| double | pot | float | Ev | float | eRecoOther |
| float | eRec_FromDep | float | Ellep | float | eDepP |
| float | Ev_reco | int | isCC | float | eDepN |
| float | Ev_reco_nue | int | nuPDG | float | eDepPip |
| float | Ev_reco_numu | int | nuPDGunosc | float | eDepPim |
| float | mvaresult | int | LepPDG | float | eDepPi0 |
| float | mvanue | int | mode | float | eDepOther |
| float | mvanumu | int | GENIE_ScatteringMode | float | NuMomX |
| float | cvnnue | int | nP | float | NuMomY |
| float | cvnnumu | int | nN | float | NuMomZ |
| float | cvnntau | int | nipi0 | float | LepMomX |
| float | cvnnc | int | nipip | float | LepMomY |
| int | reco_q | int | nipim | float | LepMomZ |
| float | Ellep_reco | int | niem | float | LepE |
| float | theta_reco | int | nikp | float | LepNuAngle |
| int | reco_lepton_pdg | int | nikm | ctor3D | LepEndpoint |
| float | RecoLepEnNue | int | nik0 | int | run |
| float | RecoHadEnNue | int | niother | int | subrun |
| float | RecoLepEnNumu | int | nNucleus | int | event |
| float | RecoHadEnNumu | int | nUNKNOWN | int | isFD |
| double | pileup_energy | float | Q2 | int | isFHC |
| SRNDBranch | nd | float | W | float | CVNResultIsAntineutrino |
| int | RecoMethodNue | float | Y | float | CVNResultNue |
| int | RecoMethodNumu | float | X | float | CVNResultNumu |
| int | TrackMomMethodNumu | float | vtx_x | float | CVNResultNutau |
| int | reco_numu | float | vtx_y | float | CVNResultNC |
| int | reco_nue | float | vtx_z | float | CVNResult0Protons |
| int | reco_nc | float | det_x | float | CVNResult1Protons |
| int | muon_contained | float | eP | float | CVNResult2Protons |
| int | muon_tracker | float | eN | float | CVNResultNProtons |
| int | muon_ecal | float | ePip | float | CVNResult0Pions |
| int | muon_exit | float | ePim | float | CVNResult1Pions |
| float | Ehad_veto | float | ePi0 | float | CVNResult2Pions |
| float | nue_pid | float | eOther | float | CVNResultNPions |
| | | float | eRecoP | float | CVNResult0Pizeros |
| | | float | eRecoN | float | CVNResult1Pizeros |
| | | float | eRecoPip | float | CVNResult2Pizeros |

```

/// Common Analysis Files
namespace caf
{
    /// brief The StandardRecord is the primary top-level object in the
    /// Common Analysis File trees.
    class StandardRecord
    {
    public:
        /// Metadata about the detectors
        SRDetectorMetaBranch meta;

        /// Information about the beam configuration and beam pulse for the
        SRBeamBranch beam;

        /// Truth information
        SRTruthBranch mc;

        /// Reconstructed info expected to be common to all (?) detectors
        SRCommonRecoBranch common;

        /// Reconstructed info unique to the FDs
        SRFDBranch fd;

        /// Reconstructed info unique to the ND complex
        SRNDBranch nd;
    };

```

Goals:

- “Hierarchicalize”
- Eliminate redundancy
- Add documentation
- *Make more user-friendly*

The CAF datastructure was reorganized to be “hierarchical” instead of “flat”

[from Jeremy's June 13th
update]

Major changes & consequences

One SR per *trigger* instead of one per event

- Considerations
 - “event” only makes sense in MC (== “true neutrino interaction”) and is especially ill-defined in ND spills w/ overlapping ν interactions
 - allows flexibility for, e.g., atmospheric ν analyses w/ no beam pulses
- Implications
 - **Organization** (of reco) is by “reco neutrino interaction”
 - Some **class methods necessary to connect reco objects** within ND stack (etc.). Works fine in C++; will need reimplementation in any Python context
 - CAFAna users will **need to use MultiVars instead of Vars** everywhere (unless every entry in a particular set of CAFs has only a single ν interaction...)

i.e.: for 2x2, every entry in the TTree in the file corresponds to a beam spill, and might contain *many* neutrino interactions

design of current version of CAFAna assumes there is a *single* interaction in each entry of the TTree. MultiVar is a (clumsy) way to work around this

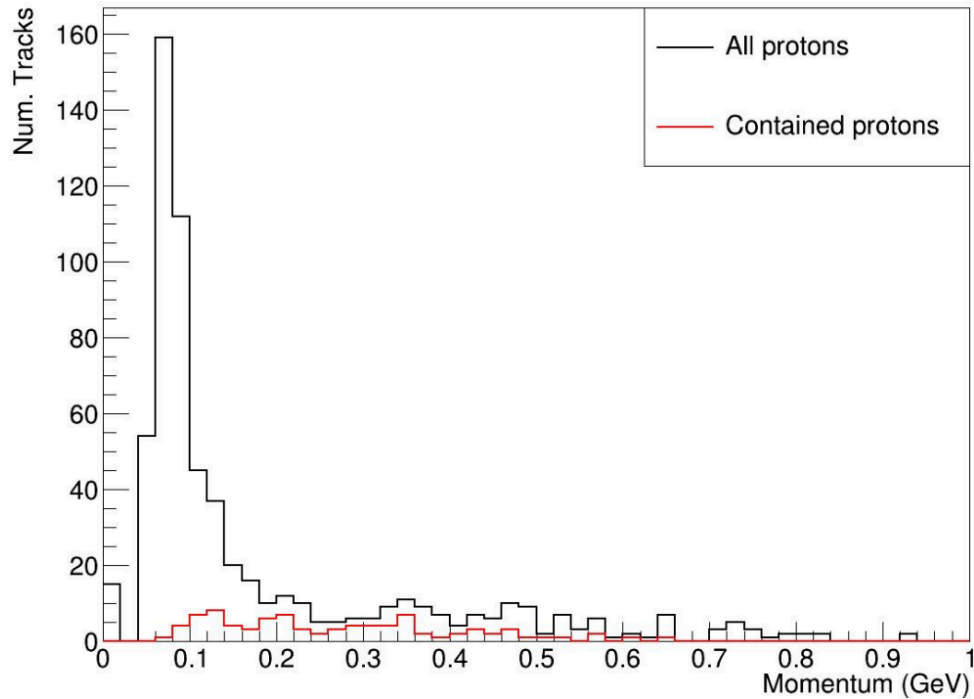
Current CAFAna usability

- Already fleshed out the MultiVar approach (see [Doc 29520](#))
 - It works, but is somewhat error-prone because *selections* get intermingled with *variable definitions*
 - Good enough for a stopgap for immediate future, but probably not maintainable long-term
 - Branch is here: [feature/mcasales CAFAna 2x2](#)

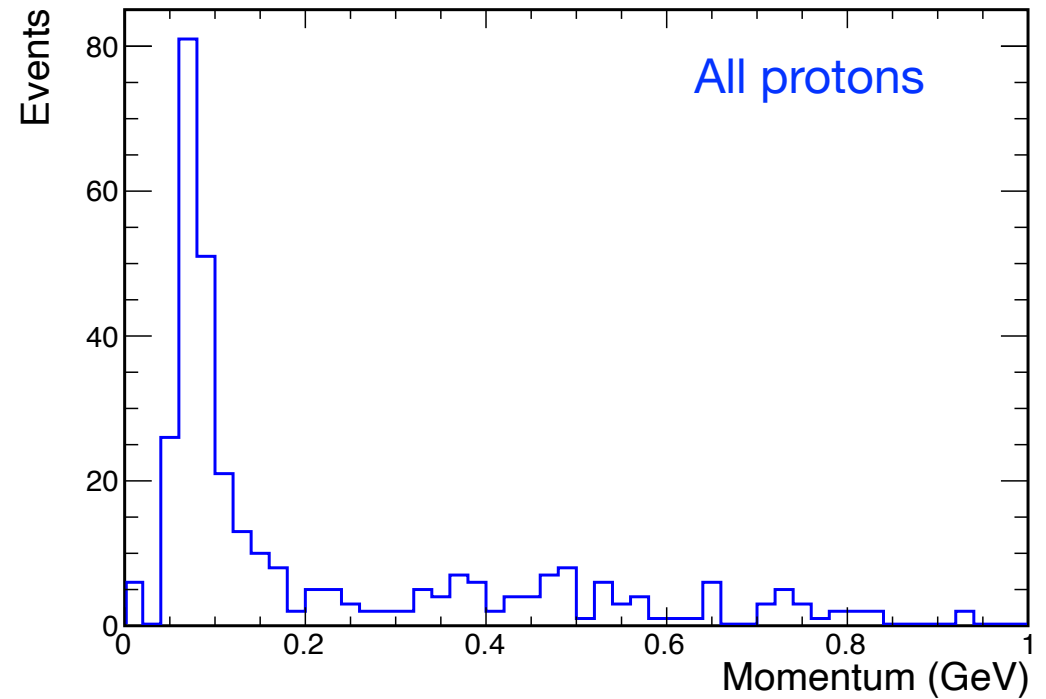
```
const MultiVar kRecoE([](const caf::SRProxy* sr){
    std::vector<double> energy = {-5000.};
    if (sr->common.ixn.ndlp>0){
        for (int i=0; i<sr->common.ixn.ndlp; i++){
            for (int j=0; j<sr->common.ixn.dlp[i].part.ndlp; j++){
                energy.push_back(sr->common.ixn.dlp[i].part.dlp[j].E);
            }
        }
    }
    return energy;
});
```

Current CAFAna usability

- Mesonless $\nu \mu$ CC Cross Section preliminary selection
 - Selected protons momentum



Andrew Cudd, Elise Hinkle,



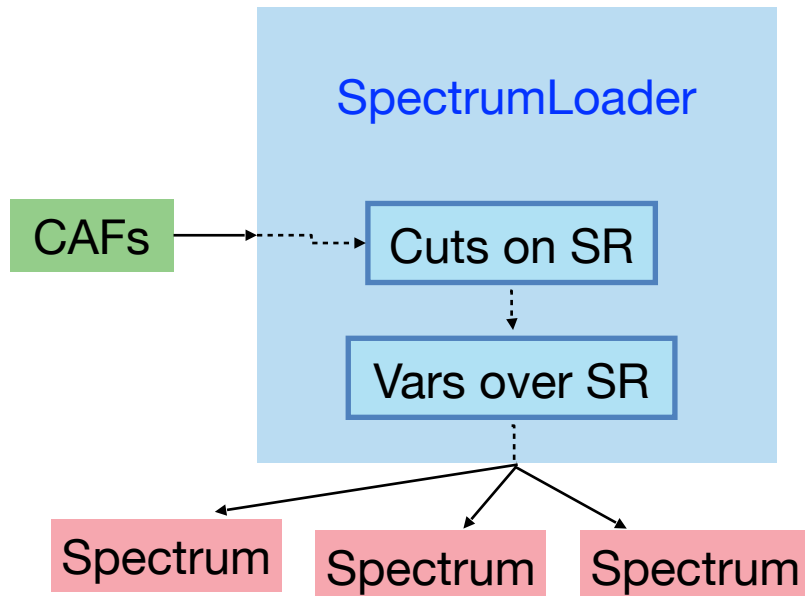
CAFAna

In progress...

Longer-range vision

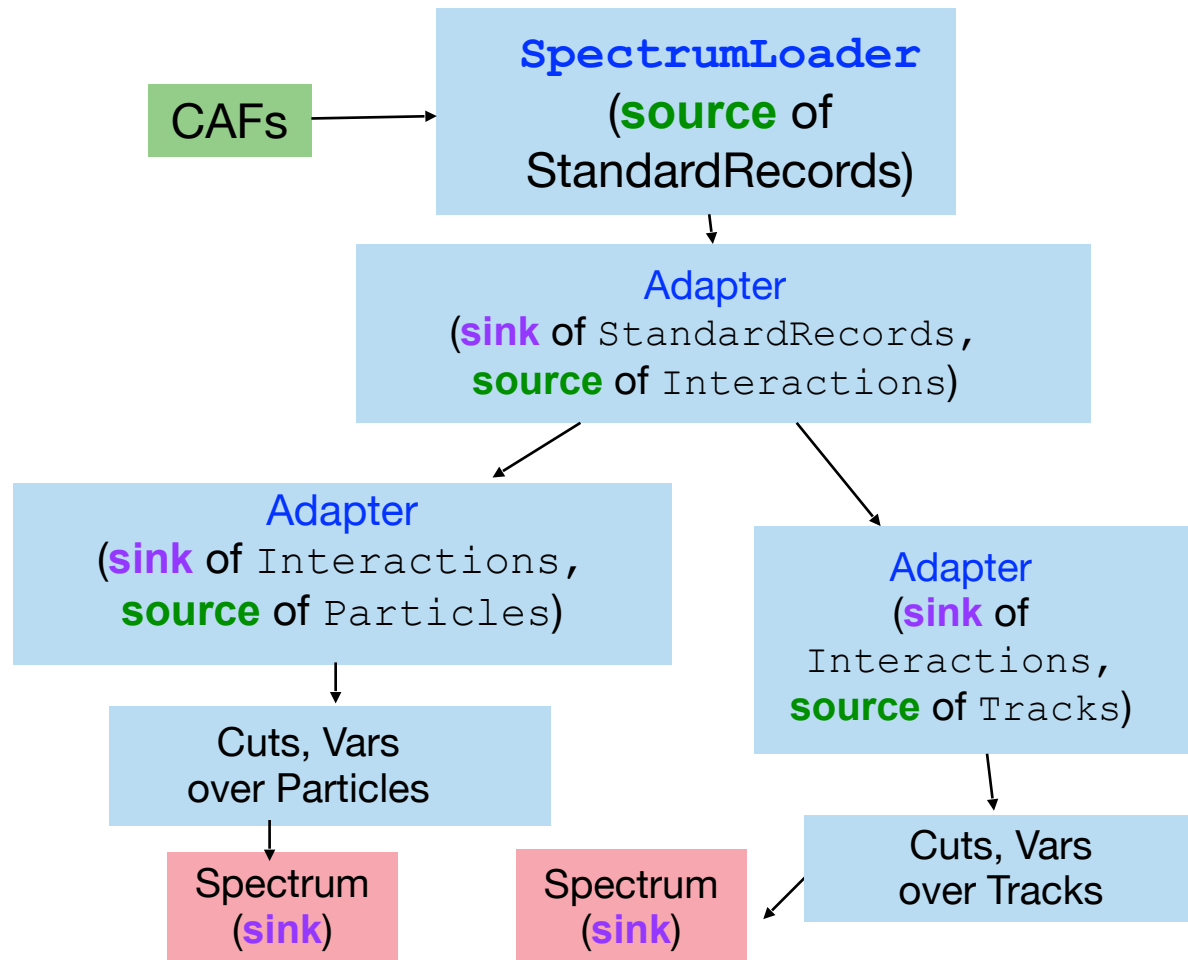
- We set out to chart out how a “holistic” solution would work, and evaluate how feasible it is
- We knew of some work that C. Backhouse (previous CAFAna lead developer) had left partially completed
- We expected that this would provide the foundation we needed, but that it would take significant new work to bring it to completion
- We were pleasantly surprised to discover that most of what we needed had already been prototyped for use in SBN context
 - Has a similar “one-to-many” problem for records & interactions
 - One analyzer used the updated approach for a xsec analysis
 - Doesn't map perfectly onto DUNE case, but significantly clarifies the path forward (+ reduces expected time to completion)

Sketch of what's different



In the old model, a "SpectrumLoader" both manages the interface to files and directly applies selections and variable definitions to the records that come out of the file to fill in Spectrums.

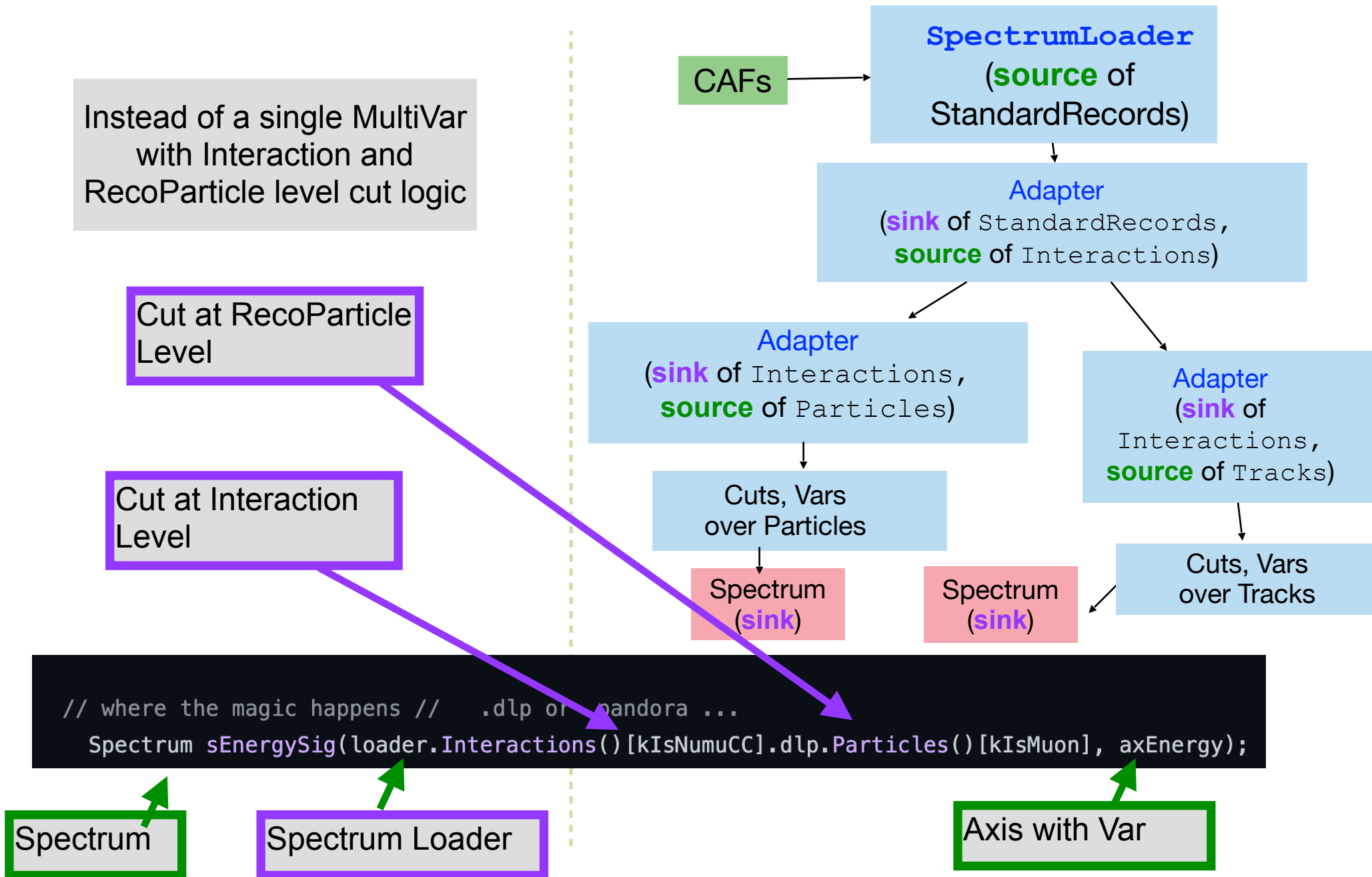
This tightly ties the Cut/Vars to the structure of the CAF record type (StandardRecord).



The new model switches to a much more generic "source-sink" architecture (at CAFAnaCore level) of which the SpectrumLoader is just one component.

The result is that the CAF record type can be decoupled from the Var/Cut management.
Thus Cuts and Vars can be genericized to work with any object in the StandardRecord hierarchy.

Sketch of what's different



Work plan

- We've laid out the roadmap for the necessary adaptations to DUNE CAFAna
 - Structural code changes
 - Developing in branch [feature/source-sink-overhaul](#)
 - Tests
 - Basic [Test script](#) we're working toward is written
 - Eventually can cross check with distributions made with “old CAFAna” (see slide 5)
 - Documentation of CAFAnaCore
 - Currently in [fork](#)

🔒 **Update CAFAna to Source-Sink architecture to handle new CAF structure**

View 1 + New view

Filter by keyword or by field

| | Title | ... | Assignees | ... | Status | ... | + |
|---|--|-----|--|-----|-------------|-----|---|
| 1 | 🔗 Sketch of moving to the full RecordSource infrastructure #46 | | | ▼ | | ▼ | |
| 2 | 🟢 Structural changes to SpectrumLoader & friends to admit one-to-many (SR -> SRNeut... #52 | |  chenel and mcasal... | ▼ | In Progress | ▼ | |
| 3 | 🟢 Documentation for Source-Sink approach to CAFAnaCore #16 | |  chenel and mcasal... | ▼ | In Progress | ▼ | |
| 4 | 🟢 Port EnsembleSpectrum from SBN CAFAna #53 | |  chenel and mcasal... | ▼ | Todo | ▼ | |
| 5 | 🟢 Doxygen pages for CAFAnaCore #15 | |  chenel | ▼ | In Progress | ▼ | |

+ You can use `Control + Space` to add an item

[Tracked as DUNE GitHub [“project” #14](#)]

Timeline guesstimates

- There are two major stages in the work plan:
 - “Central value” tools—producing single distributions.
 - Path here is completely laid out, expect to complete by Jan./Feb.
 - Systematics toolkit
 - Path is still under construction—some aspects clearer than others (reweights: easy? shifts: harder?)
 - Harder to guess timeline, but Spring '24??