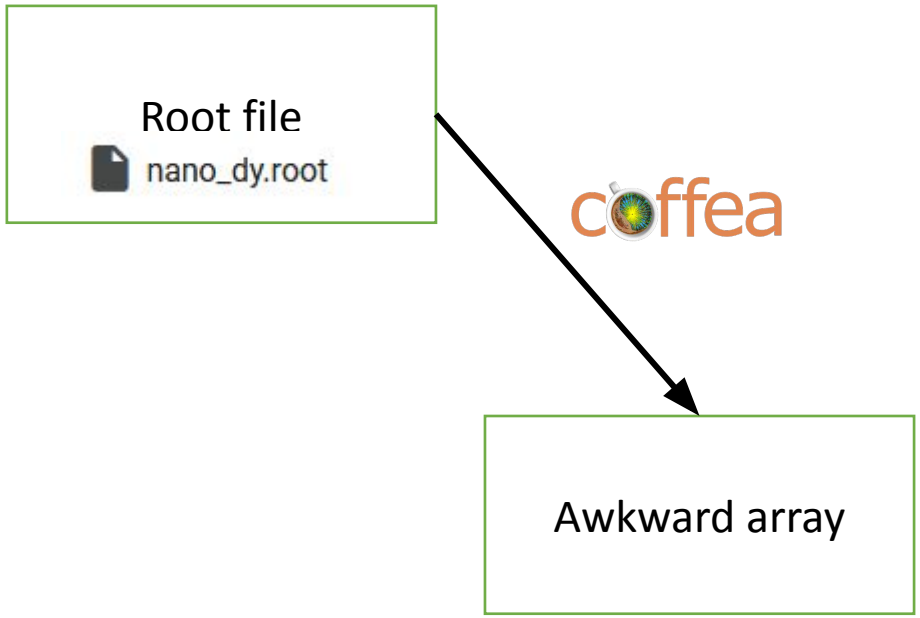
The development of coffea schemas

This project was made by the support of the IRIS-HEP program

Fellow: Maxym Naumchyk

Mentors: Ianna Osborne, Peter Fackeldey, Iason Krommydas, Nick Smith
(listing order is random)

HEP data analysis using coffea




```
[{PSWeight: [1.01, 1.26, 0.99, 0.791], SoftActivityJetNjets2: 5, ...},
 {PSWeight: [2.06, 0.872, ..., 0.962], SoftActivityJetNjets2: 5, ...},
 {PSWeight: [1.07, 0.887, 0.933, 1.02], SoftActivityJetNjets2: 7, ...},
 {PSWeight: [0.833, 0.827, 1.15, 1.1], SoftActivityJetNjets2: 4, ...},
 {PSWeight: [0.936, 0.622, 1.04, 1.17], SoftActivityJetNjets2: 16, ...},
 {PSWeight: [1.17, 1, 0.86, 0.978], SoftActivityJetNjets2: 6, ...},
 {PSWeight: [1.12, 1.59, 0.906, 0.688], SoftActivityJetNjets2: 4, ...},
 {PSWeight: [0.946, 0.922, ..., 0.997], SoftActivityJetNjets2: 12, ...},
 {PSWeight: [0.88, 0.855, 1.09, 0.811], SoftActivityJetNjets2: 3, ...},
 {PSWeight: [0.854, 0.518, 1.11, 1.16], SoftActivityJetNjets2: 20, ...},
 ...,
 ------------------------------------------------------------------------
 backend: cpu
 nbytes: 243.3 kB
 type: 40 * event
```

How coffea works

Root file → uproot → TTree structure

```
<TTree 'Events' (1499 branches) at 0x7b106caaa2d0>
```

?

Awkward array

```
[{PSWeight: [1.01, 1.26, 0.99, 0.791], SoftActivityJetNjets2: 5, ...},
 {PSWeight: [2.06, 0.872, ..., 0.962], SoftActivityJetNjets2: 5, ...},
 {PSWeight: [1.07, 0.887, 0.933, 1.02], SoftActivityJetNjets2: 7, ...},
 {PSWeight: [0.833, 0.827, 1.15, 1.1], SoftActivityJetNjets2: 4, ...},
 {PSWeight: [0.936, 0.622, 1.04, 1.17], SoftActivityJetNjets2: 16, ...},
 {PSWeight: [1.17, 1, 0.86, 0.978], SoftActivityJetNjets2: 6, ...},
 {PSWeight: [1.12, 1.59, 0.906, 0.688], SoftActivityJetNjets2: 4, ...},
 {PSWeight: [0.946, 0.922, ..., 0.997], SoftActivityJetNjets2: 12, ...},
 {PSWeight: [0.88, 0.855, 1.09, 0.811], SoftActivityJetNjets2: 3, ...},
 {PSWeight: [0.854, 0.518, 1.11, 1.16], SoftActivityJetNjets2: 20, ...},
 ...,
 ------------------------------------------------------------------------
 backend: cpu
 nbytes: 243.3 kB
 type: 40 * event
```

How coffea works

Root file → uproot → TTree structure

```
<TTree 'Events' (1499 branches) at 0x7b106caaa2d0>
```

Coffea schemas → Awkward array

```
[{PSWeight: [1.01, 1.26, 0.99, 0.791], SoftActivityJetNjets2: 5, ...},
 {PSWeight: [2.06, 0.872, ..., 0.962], SoftActivityJetNjets2: 5, ...},
 {PSWeight: [1.07, 0.887, 0.933, 1.02], SoftActivityJetNjets2: 7, ...},
 {PSWeight: [0.833, 0.827, 1.15, 1.1], SoftActivityJetNjets2: 4, ...},
 {PSWeight: [0.936, 0.622, 1.04, 1.17], SoftActivityJetNjets2: 16, ...},
 {PSWeight: [1.17, 1, 0.86, 0.978], SoftActivityJetNjets2: 6, ...},
 {PSWeight: [1.12, 1.59, 0.906, 0.688], SoftActivityJetNjets2: 4, ...},
 {PSWeight: [0.946, 0.922, ..., 0.997], SoftActivityJetNjets2: 12, ...},
 {PSWeight: [0.88, 0.855, 1.09, 0.811], SoftActivityJetNjets2: 3, ...},
 {PSWeight: [0.854, 0.518, 1.11, 1.16], SoftActivityJetNjets2: 20, ...},
 ...,
 ------------------------------------------------------------------------
 backend: cpu
 nbytes: 243.3 kB
 type: 40 * event
```

coffea schemas

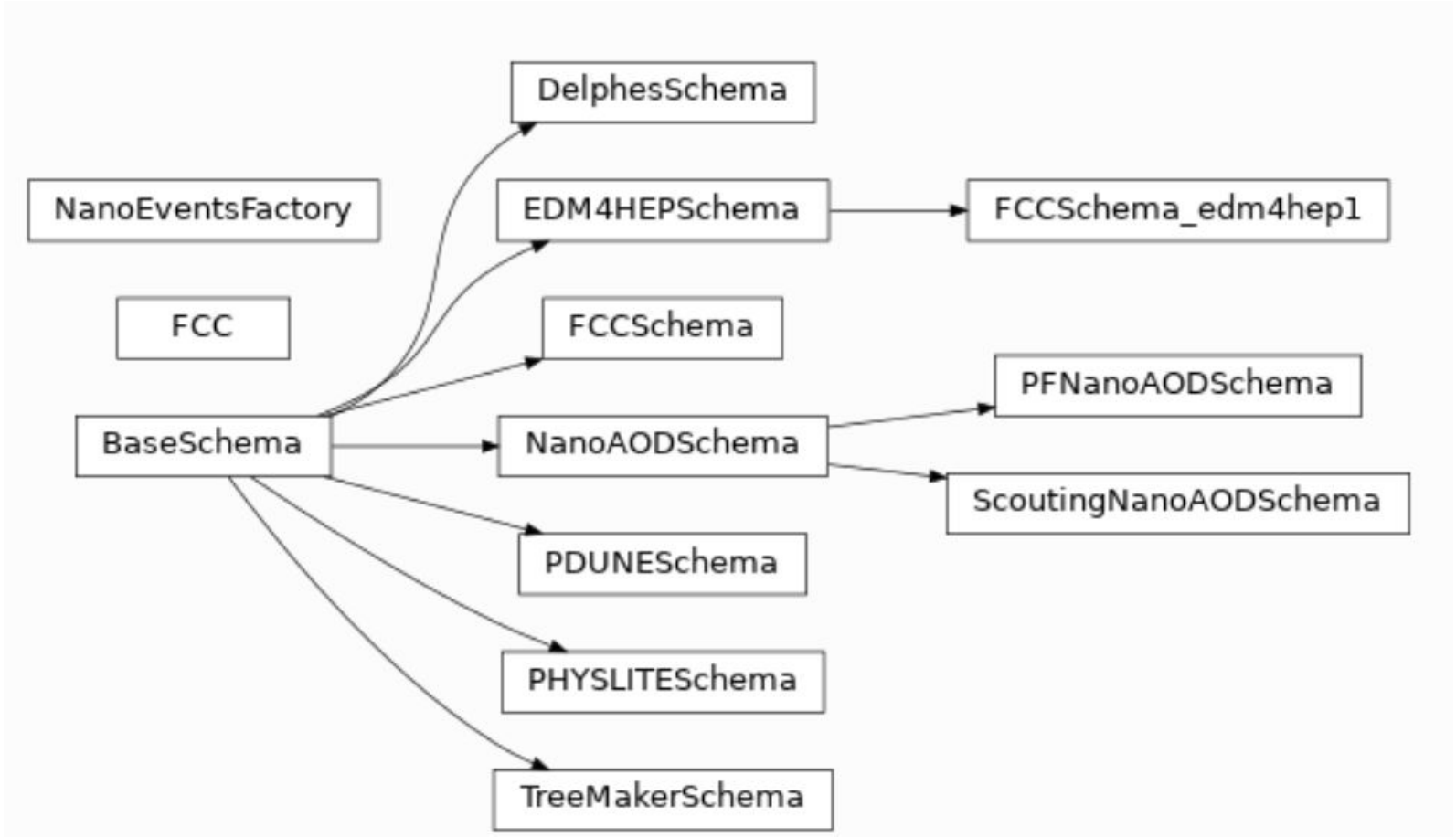
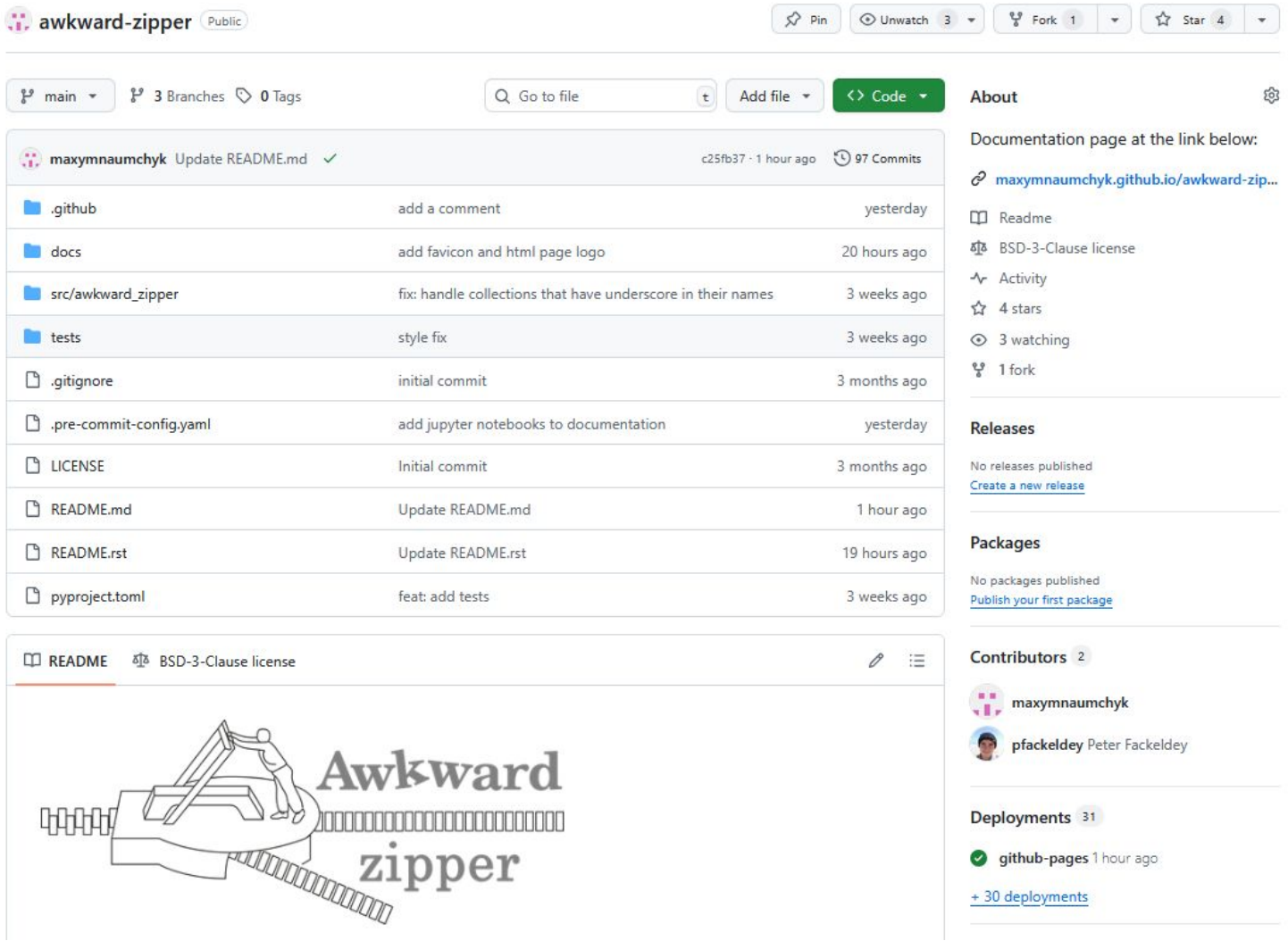


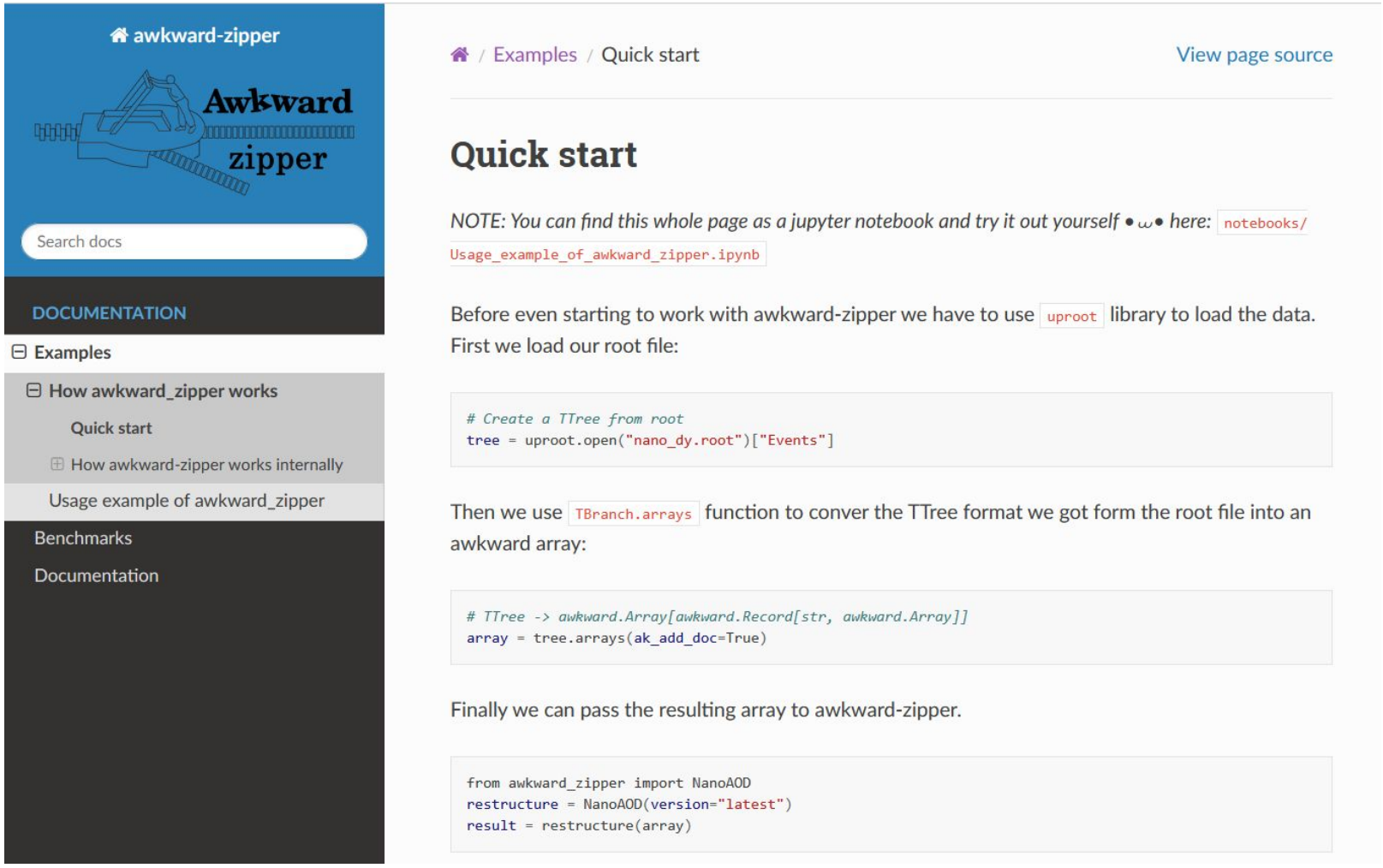
Image link: https://coffea-hep.readthedocs.io/en/latest/modules/coffea.nanoevents.html#classes

Awkward-zipper



Link: https://github.com/maxymnaumchyk/awkward-zipper

Documentation



Link: https://maxymnaumchyk.github.io/awkward-zipper/usage_example.html

Advantages of a standalone package

- Smaller package
- Only a few dependencies
- Improvements(simplifications) in how the ‘schemas’ work internally (makes it easier for users to create their own ‘schemas’)

Plans of development

- Nanoaod format works with eager (materialized) arrays
- Nanoaod format works with virtual (not materialized) arrays
- Eager and virtual modes are well tested and documented

Currently here ->

- First release
- Awkward-zipper supports all *‘schemas’* from coffea
- Coffea uses awkward-zipper instead of its own *‘schemas’*
- All the new schemas are developed on zipper directly

Thank you for your attention!