

# User analysis software in a large collaboration (xAODAnaHelpers) development, maintenance, and training

WLCG/HSF Workshop, DESY

---

**Tobias Fitschen**

14 May 2024

University of Manchester

 eossc |  EVERSE



European Research Council  
Established by the European Commission

## Tobias Fitschen

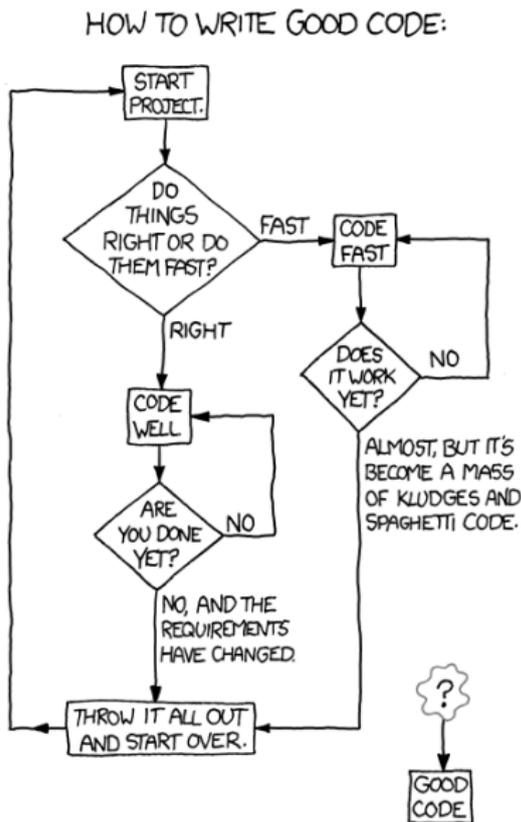
- PostDoc at University of Manchester (UoM)

### Interests

- DM search with alternative analysis strategies
  - Trigger level analysis (TLA)
  - Analysis contact/editor: Full Run 2 dijet TLA
  - Involved in Run3 TLA
- Jets and ML
  - Jet definition + MC calibration convener
  - Jet/EtMiss - ML-Forum Liaison
- Sustainable computing
  - **Environmentally**: Co-investigator of grant on sustainable ML, grant on sustainable computing at UoM
  - **Maintainability**: Maintainer of xAH (this talk): **EVERSE** pilot case (European **V**irtual **I**nstitute for **R**esearch **S**oftware **E**xcellence)



- Contributors are physicists, not software engineers
- Core software often maintained by single contributors
- Many users rely on the software to be maintained
- Too often code is written to "get this plot out real quick" rather than to last
- Recently committed to be **xAH** maintainer
  - Would have been discontinued end of 2023
  - Would lead to big delay in many analyses



## xAODAnaHelpers (🔗 [github](#))

- Framework to read and operate on AOD samples
- Apply selections, calibrations, scale factors...
- Produce Ntuples, histograms

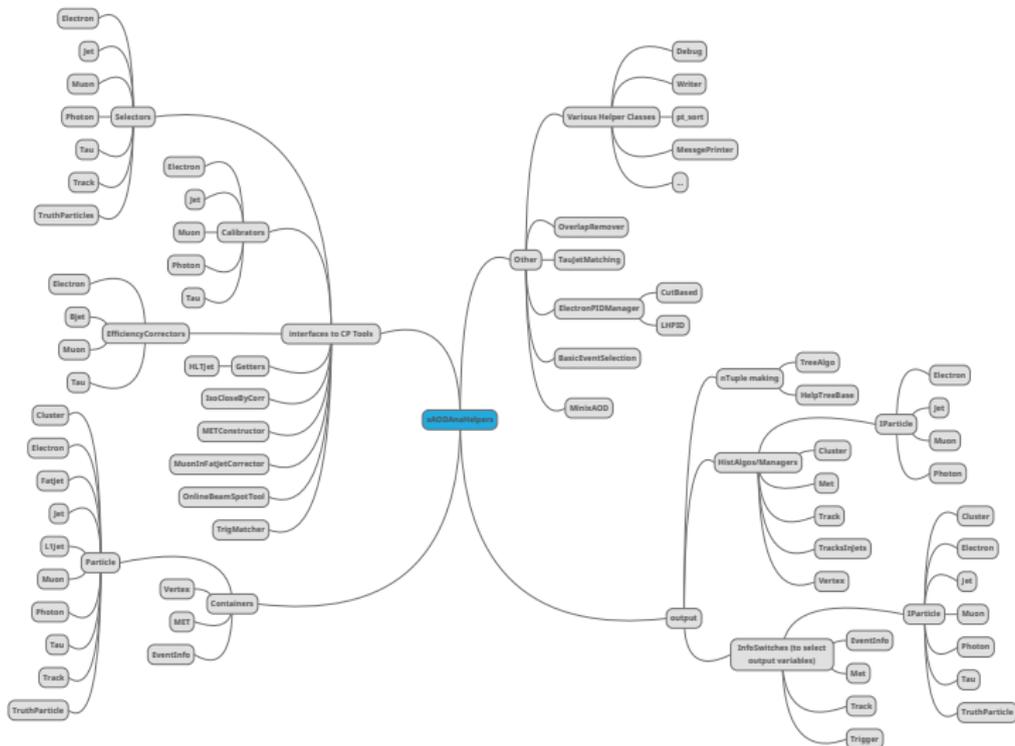
## Configuration

- Each job has one config file
  - Easy book-keeping & reproducibility
  - Same locally or on the grid
- Passing containers (e.g. jets) between algorithms

example\_config\_calib.py

```
#adapted from: https://xaodanahelpers.readthedocs.io/en/latest/usingus.html
1
2
3 from xAODAnaHelpers import Config
4 c = Config()
5
6
7 c.algorithm("BasicEventSelection",
8 {
9     "n_truthLevelOnly": False,
10    "n_doPurweighting": False,
11    "n_vertexContainerName": "PrimaryVertices",
12    "n_PVWTrack": 2,
13    "n_name": "myBaseEventSel",
14 })
15
16
17 c.algorithm("JetCalibrator",
18 {
19     "n_name": "JetCalibrator:Offline",
20     #------ Container Flow -----#
21     "n_inContainerName": "AntiKt4EMPFLOWJets",
22     "n_jetAlgo": "AntiKt4EMPFLOW",
23     "n_outContainerName": "AntiKt4EMPFLOWJets_calibrated",
24     "n_sort": True,
25     "n_redoJVT": False,
26     #------ Systematics -----#
27     "n_sysName": "Nominal",
28     "n_sysVal": 0,
29     #------ Calibration -----#
30     "n_calibConfigData": "PreRec_R22_PFlow_ResPU_EtaJES_GSC_February23_230215.config",
31     "n_calibSequence": "JetArea_Residual_EtaJES_GSC_Insitu",
32     #------ Cleaning -----#
33     "n_doCleaning": False,
34     "n_jetCleanCutLevel": "LooseBad",
35 })
36
37
38 c.algorithm("JetHistsAlgo",
39 {
40     "n_inContainerName": "AntiKt4EMPFLOWJets_calibrated",
41     "n_detailStr": "kinematic",
42     "n_name": "NoPreSel"
43 })
44 }
```

## Algorithms themselves are C++ interacting with ATLAS Athena CP tools



tools: mindmap, doxygen, dot

Roughly corresponds to inheritance graph, cleaned up and reordered for visibility

## Who is using us?

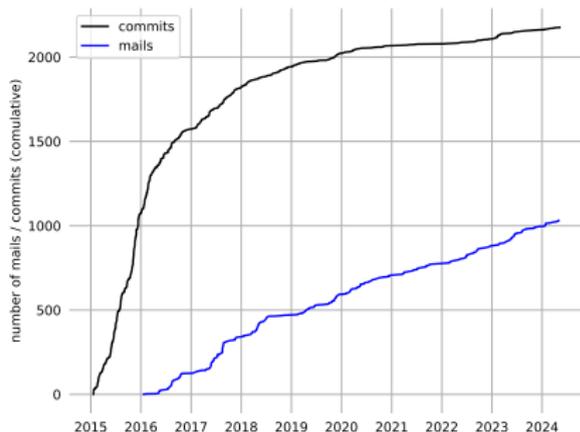
- [ttH->multileptonic final state](#)
  - [HTop](#) - former HSG8 group
- [dijet](#)
- [multijet](#)
- [hh->4b](#)
- [VBF + invisible](#)
- [g->tt susy multi-b-jet](#)
- [Jet/MET](#)
- [Jet Cleaning](#)
- [jet inputs to reconstruction](#)
- [Punch-through studies](#)
- [Multijet balance](#)
- [Standard Model inclusive jet cross section](#)
- [Voronoi Area Pileup Subtraction](#)
- [Trigger-Level Analysis](#)
- [Jet trigger group performance studies](#)
- [Dijet+ISR Analysis](#)
- [SM Full Run 2 Z+HF analysis](#)
- [SM Full Run 2 W+jets analysis](#)

<https://xaodanahelpers.readthedocs.io/en/latest/Community.html>

## User contact: via email-list

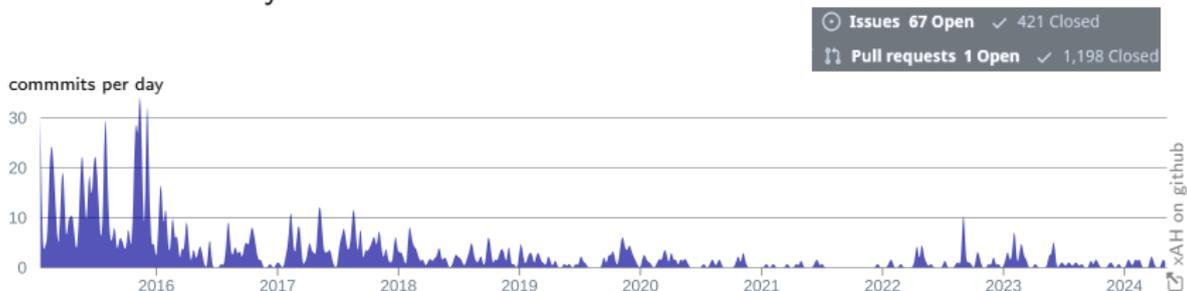
✉ [atlas-sw-xAODAnaHelpersFW@cern.ch](mailto:atlas-sw-xAODAnaHelpersFW@cern.ch)

- 1k mails to date
- **Mail activity** remains constant
- While commit frequency decreased
  - develop → maintain



data from ✉ [mail archive](#) and ✉ [git repo](#)

> 2k commits by 80 contributors



As with most projects, large activity in the start

- Now work shifted towards maintaining rather than adding new features
  - Support for users
  - Fixing eventual bugs that were not discovered earlier
  - Keeping up-to-date with updates upstream (athena) and downstream (dependencies)
- 2023: peak in activity due to major update of athena for parallelism

Documentation created with [doxygen](#), hosted by [ReadTheDocs](#)

- Parts written manually
- Enhanced by doxygen-readable comments for automatic additions

The screenshot shows the documentation page for xAODAnaHelpers. On the left is a dark sidebar with a blue header containing the project name and a search bar. The sidebar lists navigation items: Introduction, Installing, xAH\_run.py, Algorithms, Development, Community, xAH FAQ, API Reference, and Doxygen API. The main content area has a light blue header with 'Docs > xAODAnaHelpers' and an 'Edit on GitHub' link. The main heading is 'xAODAnaHelpers', followed by a version badge '001 10.5281/review.7335120 build unknown'. The text describes the framework as 'The xAOD analysis framework, born out of ProofAna...or not.' and includes a 'Welcome to the xAODAnaHelpers wiki!' section. Below this is a 'Supported releases' section with a note about R21 and R22 documentation. A 'Latest Version' section contains a 'Note' box with instructions on how to add new releases. At the bottom, it lists supported releases: AnalysisBase,25.2.6, AnalysisBase,25.2.5, and AnalysisBase,25.2.4. A vertical watermark 'xAH documentation' is visible on the right side of the page.

Hosted hands-on tutorial during [2024 ATLAS UK week](#)

[Instructions on gitlab](#)

Name	Last commit	Last update
📁 .vscode	<a href="#">Add VS Code settings</a>	4 months ago
📁 config	<a href="#">cosmetic changes to configs</a>	4 months ago
📁 run	<a href="#">added (hopefully) publicly accessible test file</a>	4 months ago
📁 src	<a href="#">Merge remote-tracking branch 'origin/dev-j...</a>	4 months ago
🔖 .gitignore	<a href="#">using recommended calib config</a>	4 months ago
📄 README.md	<a href="#">improved instructions in README</a>	3 months ago
📄 setup.sh	<a href="#">merge into master</a>	4 months ago

📄 README.md

### Table of Contents

---

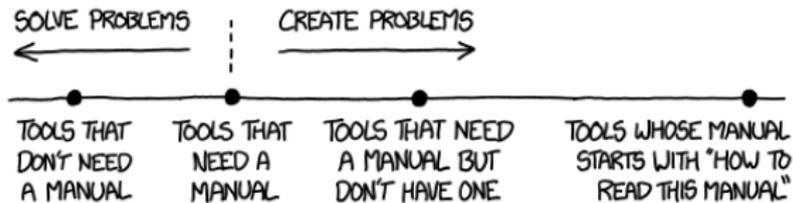
- [0. Further Resources](#)
- [1. Setup and Making Histograms with xAH](#)
- [2. Advanced Algorithms: Re-calibrating Jets](#)
- [3. xAH in an Analysis Framework: TLASteeringFullRun3](#)
- [4. The Future of xAH: Integration into CPTools](#)
- [5. Contacts](#)

[tutorial's gitlab](#)

- Basic usage: selecting events, calibrating jets, writing histograms/nTuples
- How to use xAH in a larger project/analysis
- Some words on development/future of xAH

## Lessons learned from tutorials:

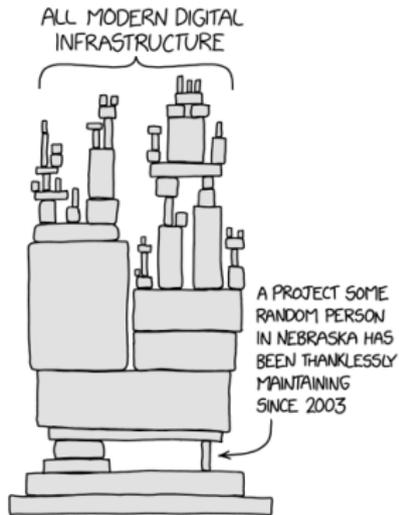
- Very important to provide solid setup instructions
- Be prepared for MAC, Linux, Windows users
- Spend first 30 min to get all participants fully setup
- Send out instructions well in advance before course
- Work together with participants to improve them
- Preparing tutorials is a great way to update documentation
- Recognise that not everyone has to learn your specific framework



✉ xkcd.com/1343

## From the Linux foundation's [guide for open source maintainers](#)

- Create a simple page that aggregates all entry points & communication channels
- Strive to lower the bar for contributions
  - Don't reject; offer improvements
  - But don't be afraid to say no
- Establish a triage process for issues
- Radical transparency in communication
- Set boundaries and stick to them
- A failing CI pipeline offers an objective reason to reject a suggestion
- Constantly look for efficiency hacks to optimize and automate
- Step away if you feel burned out, both for you and for the project
- Recognize you will never "finish the job"



## xAODAnaHelpers, v1.0.0

Giordon Stark ; Marco Milesi ; John Alison ; and 46 others

Flexible C++ xAOD analysis, driven by python, on top of ATLAS offline software  
16 more versions exist for this record  1040

## DOI on zenodo: DOI:10.5281/zenodo.596498

- 17 versions, updated since 2015
- Linked to on github and in documentation
- 1k views, citations: None!

## Contributions to software are often not recognized (in traditional ways)

- Open source software often seen as infrastructure that is "just there"
- Need framework for proper accreditation → [EVERSE](#)

ATLAS is moving to **CP Algorithms**: ↗ (see this CHEP23 talk)

- Common wrapper around **CP Tools**
  - Harmonises application of recommendations for: calibrations, selections, scale factors, etc
  - Developed/maintained by respective domain experts
  - Avoids errors by analysts who cannot be experts for all
- **xAH**, like most frameworks, interacts with **CP Tools** directly
  - Will either have to change to **CP Algorithms** or be deprecated
  - Major refactoring needed
- Major projects take a lot of time → needs accreditation → **EVERSE**



## EVERSE

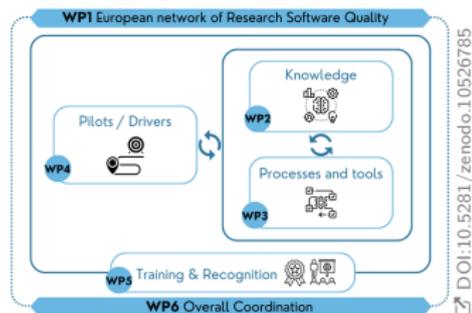


- **E**uropean **V**irtual **I**nstitute for **R**esearch **S**oftware **E**xcellence

- Website with more information: [EVERSE](#)
- More on EVERSE in [dedicated talk on Friday](#)
- Funded as part of [Horizon Europe Initiative](#)
  - EU key (100 € billion) funding programme
  - Strong focus on environmental sustainability

- Manchester HEP: WP4 leader for [Science Clusters pilot cases](#)

→ including **xAODAnaHelpers**



## Goals

- Build community-led structure for evaluating and improving code
- Establish sustainable and reliable ecosystem of stakeholders
- Create framework for appropriate recognition for software careers

## Analysis software in a large collaboration (ATLAS)

- Most contributors are physicists, not software engineers by trade
- Code is often written to "get this plot out real quick"
- Little reward for sustainability, focus often on quick results

## Towards sustainable software practices

- Need to reward sustainable, long-lasting software
- Framework for accreditation of good software practices
- **EVERSE** strives to be just that