Reproducible Workflows, Continuous Integration, Analysis as a Service: Templated Analysis Collection on GitHub

Developing Greenfield Software that Kicks AaaS!

Wouter Deconinck

umanitoba.ca

Problem: Fragmentation of Analysis Code

Fractured codebase

- Used to be that almost all production code was in Athena

- Nowadays we have many, many repositories (truthfully, I am not sure how many), not all of which are in gitlab

- Some are not visible even to the SW coordinators

- Best we can do is to try to educate people about best practices and concentrate on Athena

(Edward Moyse, ATLAS Software Sustainability, Sustainable Software in HEP, 22 July 2020)

- Discussing a way to centralise bookkeeping of available analysis code
  - "Analysis Code Repository"
  - Ease discoverability
  - Enable collaboration and generate synergies

(Danilo Piparo, CMS Thoughts about Sustainable Software in HEP, ibid., 22 July 2020)

Need: Everything Sustainable & Connected

- In order to sustain research software, we can
  - Do things that reduce the amount of work needed
  - Do things that increase the available resources
  - Do things that do both reduce the amount of work needed and increase the available resources

- To reduce the amount of work needed
  - Train its developers, which involves finding or developing training material
  - Use best practices, which involves finding or developing best practices

- To increase the available resources
  - Create incentives so that people want to work on the software
    - Citations that help in existing career paths
    - Adjusted existing career paths that they reward software work
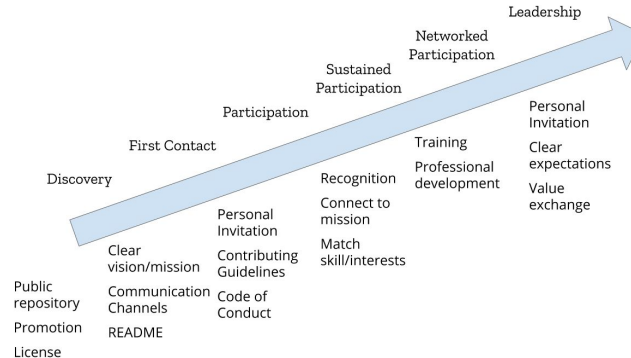   - New career paths

(Daniel Katz, Introduction to Software Sustainability, ibid., 22 July 2020)

Goal: fully connected research graph

- Community of practitioners
- Training grounds for new users

Aligning user incentives with goals

- Attribution and recognition
- Visibility, on github or platform X
- Opportunity for growth into RSE

umanitoba.ca

Proposal: Build Analysis Repos for Forking

Built-in CI/CD, containerization, documentation

- Maintain ability to pull updates, submit PRs, as tools mature
- Small analysis as CI, larger jobs by hash, full reproducibility
- Basic implementation development as service role to DWG/PWG

Experiences and Resources

- 2 M.Sc. student Sept 2021 - Aug 2023 (1 funded, 1 proposed)
- 30% FTE PI, EIC SC and EIC Canada member
- Experience with CI/CD, HCD, agile project delivery
- cvmfs, singularity, spack, OSG, ComputeCanada,...

* Forking = importing repo, adding upstream as remote.

umanitoba.ca

University of Manitoba