Cooperation in Software: models and needs

Axel Naumann, 2020-04-02

IDT-UM / ErUM-Data

This is not just about ROOT

- ROOT as an example and personal motivation

- (Lack of) cooperation as general issue in the community

  - Few major software teams suffocate other smaller ones: not enough "meat" left to get funding. Lack of distribution of tasks.

  - Causes duplication of efforts, segregation, non-technical opposition.

- In an environment of scarce resources, difficult sustainability, and major challenges!

Facilitators

- WLCG as coordination body

- Projects such as ACTS

- Personal connections

Phases of Successful Software

- Talking

- Coding

- Rolling out

- Maintaining

Phases of Successful Software

- Talking: cooperatively, happy to share ideas if start of a cooperation

- Coding: distribute and share tasks

- Rolling out: benefit from existing, very visible channels

- Maintaining: make use of labs' existing long-term teams

Goal: Win-Win

- Better software for community

- Less duplication of effort

- More visibility and relevance for all contributors

  - Creating software for production use, not for CHEP papers

Collab + Coop Models

- Formal: Geant

- Cooperative: ROOT

- Ad-hoc: HSF

- Others (US IRIS) mix this: formal for money, cooperative in coordination, seemingly ad-hoc in topics (depending on available contributors?)

- What should we aim for, in the future?

What brings me here?

- Community has problems to solve

- Would like to have even higher visibility of e.g. German contributors

  - Embedded in international projects

- How? What do you need? What can projects, what can CERN change to enable this? Can we find a general answer or is it "depends" (which hinders scale)?

Module	Code Owner
I/O	Philippe Canal (**Fermilab**)
• Compression	Oksana Shadura (**Nebraska-Lincoln**)
Math, Statistics, Machine Learning	Lorenzo Moneta (CERN)
• RooFit	Wouter Verkerke (**NIKHEF**), Stephan Hageboeck (CERN)
• New histograms (RHist)	Hadrien Grasland (**LAL**), Axel Naumann (CERN)
Graphics	Olivier Couet (CERN), Sergey Linev (**GSI**)
RDataFrame	Enrico Guiraud (CERN)
Web-GUI, http Server	Sergey Linev (**GSI**), Bertrand Bellenot (CERN)
Event Display	Matevz Tadel (**UCSD**), Alja Mrak-Tadel (**UCSD**)
PyROOT	Enric Tejedor (CERN)
cling	Axel Naumann (CERN)
Build System	Oksana Shadura (**Nebraska-Lincoln**)
C++ Modules	Vassil Vassilev (**Princeton**), Oksana Shadura (**Nebraska-Lincoln**)
Documentation	Olivier Couet (CERN)

Why? Why now?

- We have major roadblocks for instance for HL-LHC

- Game changer for analyses:

  - High statistics

  - Sampling multi-dimensional theory parameters

  - Systematics must be under control to match stat precision (correlations correlations correlations)

- Need to counterweight today's focus on national funding agency

Context: National Facilities

- Facilities and software should be designed for each other

  - Decisions on storage, networking, CPU/memory need input from software

- For software, this is also a question of evolution

  - Designing a facility should motivate software evolution; plans in software evolution change recommendations on facility design

- How much, on which scale is this information exchange happening? (see Exascale)

ROOT-Related Tasks (Examples!)

- I/O: TTree succession (RNTuple)
  E.g. remote I/O, benchmarking, key/value store, metadata (parentage, luminosity), parallel writing, early adoption, tooling, layout vs GPU, STL support, schema evolution

- Math
  E.g. generate CUDA kernels at runtime (CUDA@cling!); RooFit; multithreaded ML inference (TMVA); RANLUX++ for ARM; new histograms

- Analysis: RDataFrame
  E.g. heterogenous scheduling; category handling; evaluation of systematics / correlations; PROOF succession; domain-specific language

ROOT-Related Tasks

- Just examples; we know of a long list of R&D and development tasks that should be addressed, also in visualization, platform, tooling

- We would be happy to engage - obvious ErUM-Data subject candidates: heterogeneous computing; heterogenous testing; "Proof++" deployment vs virtualization (debugging!); ML, esp fast inference

- You probably have your own lists that could benefit from cooperation with existing projects

- How can we increase awareness and cooperation beyond existing personal connections?