

EP-SFT Input

Axel Naumann axel@cern.ch for the EP-SFT group
2021-09-20, Open Science Policy WG Meeting



Outline

- EP-SFT's Role
- Open Science Activities
- Opportunities
- Conclusion



EP-SFT

- Develops common software used by majority of HEP experiments world-wide. Rationale: synergy between experiments, focus of expertise. Examples:
 - Geant4: simulates particles in matter; simulation ~ 50% compute resources
 - ROOT: statistical analysis of huge amounts of data; ~ all HEP data in ROOT
 - cvmfs: software distribution at global scale; ~ all HEP software through cvmfs
- Long project lifetimes >25 years; ~ 30 highly skilled developers
- Open Source projects with CERN being main contributor / maintainer

Vision: making analysis accessible

- Given skills and data, everyone should be able to do a HEP analysis
 - Open source software: no need to pay licenses, no secret sauce
 - Documented and accessible: no need to get a USB drive and the knowledge of your fellow physicist in the neighboring office
 - Supported: no unfair advantage to insiders
 - Commodity platforms: no special investment in niche systems
- Limited by projects' resources, experiments' data and processes

Open Science and software

- At CERN, software is a tool for doing physics: it can support Open Science
- Software carries knowledge itself, is part of the intellectual wealth of CERN
 - Is sharing CERN's software and related expertise part of Open Science?
- Not all software will change the world, but some might
 - Can we distinguish "relevant" from "irrelevant"? Whose call is it?

EP-SFT software use beyond HEP

- Natural extension to HEP synergy: transcend beyond HEP
 - Geant4: companies, medical (GATE, TOPAS) + space sciences
 - ROOT: companies, finance (HighLO) + pharmaceutical research
 - Cvmfs: companies, EUCLID, LIGO, EESSI, LSST, GenAP
 - ScienceBox for deploying analysis ecosystem: companies
- Generally through adoption of software, with feedback on evolution; wider exposure taps into additional expertise / innovation

Making software usable - outside HEP

- Requires additional investment; increased relevance helps with sustainability; generally no direct return
 - Visibility: unique features must be advertised to be noticed
 - Support: from "open forum also for non-HEP" to commercial support model
 - Collaboration: stakeholders want (to understand how) to influence project evolution; range from "on GitHub" to "collaboration agreement"
 - Sustainability: buy-in requires statements on lifetime
 - Interfaces: documented, interoperable, extensible; indication: lively ecosystem

Societal impacts

- Collaboration and contributions to society's fundamental software components, such as compilers, Linux, Jupyter
- Sharing of expertise
 - Standing participation to ISO standard > 10 years
 - COVID-19: SFT projects major contributors to Folding@Home and Rosetta@Home; consulting WHO data processing
 - Industry training / consulting through KT

Indirect societal impacts

- SFT develops enabling technologies (e.g. SWAN, CernVM, ROOT, LCG stacks) for educational outreach + HEP heritage
 - Physics Masterclasses; Open Data; analysis preservation
- Training of developers: scientific software, multi-stakeholder, best practices,...
- Part of CERN's center of expertise for analysis / machine learning
- Software as vehicle for interdisciplinary research
- Software as additional output and impact of HEP

Center of Innovation

- Important role: fight "Not Invented Here" syndrome
 - Scout technology / software advances
 - Initiate collaborations, from exploratory (e.g. ALPAKA) to production (e.g. LLVM)
 - Acquire expertise in new developments, adapt and adopt

Challenges (1/2)

- Visibility, outreach, collaboration, coordination, evolution beyond HEP at a cost: balance between HEP / non-HEP investment
 - SFT must prioritize HEP, KT requires commercial focus: spin-off gap
 - Reality: unless someone finds us and invests, SFT can only do HEP
- CERN's infrastructure supporting small non-HEP usage: scale limit? Citizen science needs a service

Challenges (2/2)

- CERN essential for sustainability of HEP's common software projects and software expertise, current major issue: generational hand-over
- No defined marketplace (forum) for scientific software despite global challenges
- Potential for improved synergies, optimized solutions, increased international relevance

Conclusion

- EP-SFT software projects aren't really just CERN's: they are global
- Exclusive use of Open Source for science tools
- Long history of "outside" collaboration and cooperation
 - Open Science is not a one-way street; dissemination *and adoption* come at a cost to software projects, with significant benefits to CERN, HEP, society
- Software projects benefit from Open Data, Open Access
- Appreciate inclusion in Open Science WG, as integral constituent of today's interdisciplinary science with societal impact