

Expanding Software Collaboration beyond HEP

Fons Rademakers

Main Problem

- Each science/profession/field has their own computing culture (from Univ up to professional life)

HEP Case

- HEP community is quite small but has unique use cases:
 - Extremely big data
 - Very large code bases
 - Very long experiment life times
- A lot of low level programming to maximize usage of resources
- Try to use everything new we can... bleeding edge
- Some stuff we can share: ROOT, G4, XROOT, EOS, cernvmfs ...

Others

- Biologists
 - Minimal low level computing experience
 - SQL queries
 - One button interfaces
- Typically a large usage impedance mismatches

R-e-Inventing the Wheel...

- R started as a statistical package for engineering students with a simple, slow custom language for small in-core data sets
- After 22 years these limitations are still there
- Typically many social science, engineering and biology (pre genome) data sets are much smaller than ours, and perfectly fine for R
- For ROOT joining R at the time was not an option as R was not addressing any of our issues and vice versa for R

Why do others not profit from HEP

- We disseminate our work badly, only: CHEP, ACAT, ...
- Typically not allowed (or very difficult) to visit conferences in other fields (at least in PH-SFT) like VLDB, Super Computer, Google I/O, WWDC, etc, etc

Results of our Self Imposed Autism

- Google reinventing the wheel by making Dremmel and Map/Reduce
- CERN reinventing the wheel by not not learning enough from external sources
- Not rewarded to collaborate on projects outside HEP (there is also a fear that we will end up having to provide the support due to our long experiment life time)
- At most Our stuff typically trickles down in other fields by HEP people moving on to other fields (e.g. finance)

What do we need to Collaborate

- Spend time and effort to adapt our solution to other communities environment
 - money and effort not spend on HEP (but having a larger community pays back in a more robust system)
- Spend time and effort to adapt their solution to our environment
 - typically the other community is larger and has less interest to cater to HEP
- Be the ones willing to maintain the stuff as we probably will be using it the longest