EDWARD MOYSE

ATLAS: COPYRIGHT AND LICENCES

INTRODUCTION

▸ We have been discussing for >1 year whether we should open-source our code, and if so, what technicalities are involved.

▸ Key advantages of open-sourcing code:

  ▸ Useful for ATLAS contributors to be able to point to the work they've done

  ▸ Protect right of people to re-use their code if they move experiment / leave the field

  ▸ Collaborate more easily on software projects

ATLAS PLAN TO OPEN SOURCE SOFTWARE

▸ Plan, as presented to CB (and accepted by it):

  ▸ Add copyright to all code:

    ▸ `Copyright (C) 2002-2018 CERN for the benefit of the ATLAS collaboration`

  ▸ Add further Acknowledgements to Readme.md (if necessary)

  ▸ Make code open source under liberal licence - preferably Apache 2, but GPL if necessary.

    ▸ Analysis code remains private

    ▸ Remove code incompatible with our licensing goals

  ▸ Follow CERN code of conduct - no swearwords etc

  ▸ Binaries will be GPL

▸ Target date is ~ summer 2018

COPYRIGHT

```
Copyright (C) 2002-2018 CERN for the benefit of the ATLAS
collaboration
```

▸ Have to have CERN hold copyright because ATLAS isn't a legal entity in itself

 ▸ Having ONE copyright owner makes licensing much easier

▸ We have added this to (more or less) all of the code we plan to open-source

 ▸ Bit of a balance between adding it everywhere and unacceptable code bloat - we don't add it to python files for example

▸ Was added as part of git migration - we now require our Merge Request shifters to check that all files accepted to repository have correct text.

▸ Still a few places where personal copyright statements need removing

LICENSING

▸ Default licence would be "all rights reserved" - no good for opening our software

▸ Graeme gave a good summary in his [talk](#) back in August 2017:

   ▸ Talked about tools (e.g. fossology) to examine code for licences

   ▸ We concluded that we need to write our own scanner - partly because we need to track GPL 'infection' which requires understanding of our dependency graph

▸ He found that all of the generators, plus FastJet are released under GPLv2. Xerces (in externals) also a problem.

   ▸ Basically all open source licences compatible with GPLv2, except Apache 2 and GPLv3 (!)

▸ Since then, HepMC have agreed to move to LGPL

CLEANING THE REPOSITORY

▸ When we started thinking about opening our repository we had a look and found a few things we didn't like

  ▸ Personal copyright statement

    ▸ In many cases these were autogenerated by the editor

  ▸ Incompatible licences

  ▸ Swearwords, defamatory statements etc

▸ **ATLAS management is very keen for this to be cleaned up before we make our software public**

CLEANING THE REPOSITORY(2)

▸ However removing e.g. swearwords from history isn't so easy.

▸ Options include

   ▸ **filter-tree** - extremely powerful but slow

   ▸ **BFG** - much faster, but good for e.g. removing certain words (ATLAS has used an unfortunate shorthand for 'silicon hit' which makes this approach trickier)

   ▸ **rebase** / **replace**

▸ However changing the repository changes the hashes. Not a problem for e.g. tags, since we can update them too, but will break links to external tools (such as Jenkins and JIRA)

▸ Also potentially means that forks of atlas/athena are invalidated - at best, need to rebase. At worst, need to re-fork. Potentially very painful

▸ Another option we discussed is having a separate public repository with cherrypicks to update it, but this was generally not a popular choice (maintenance cost, not in spirit of opening sw)

CONCLUSIONS

▸ ATLAS is committed to open-sourcing software by summer 2018

▸ Much work, e.g. copyrights etc, already done.

▸ Still some work to be done, notably in cleaning the repo and checking the licence dependency.