

# Apache POI

The challenges and rewards of a 15 year old codebase

David North <[dnorth@apache.org](mailto:dnorth@apache.org)>

# About me

- Sometime software engineer, and now Head of Engineering
- Member of the POI PMC
  - Not too active lately – apologies to all...
- Spent most of my career developing Java

# About Apache POI

- *A The* Java API for reading and writing MS Office documents
- Excellent support for Word and Excel (both binary and the newer .docx and .xlsx) and growing support for the others (Powerpoint et al)

# A bit of history

- 28 August 2001: the first 0.1 alpha
- 19 April 2017: 3.16 final
- 220,000 lines of code

# Examples: the code

# Examples: the bug reports

# Never forget

- Every “messy” bit of code captures knowledge
- Code can only be as clean as its problem domain
  - The specs for the XML MS Office formats are ... of significant size

# Example: knowledge

- The comments-as-a-giant-arrow bug



# Coping strategies

- Coding standards?
- Peer review
- The power of a decent IDE
- Clearly defined module boundaries
- Future plans

# Coding standards

- This is not the same as a corporate environment...

# Peer review

- We love getting patches, and we love them even more if...

# IDE usage

- When naming conventions aren't universal, syntax highlighting *matters*
- Demo

# Module boundaries

- We do the obvious thing and split by format (Excel, Word, Powerpoint, ...) and then again by “old” (binary) and “new” (XML)
  - Sandbox for newer ideas is separate
  - Also modules for the emerging common APIs between binary and XML formats

# Future plans

# More use of Git

- Already have a semi-transparent link to Github to receive pull requests

# Evolution, not revolution

- Cleaning up some of the older decisions
- Enums are a Java feature now!
- Not alienating our users on older JDKs
  - We won't be spraying lambdas everywhere just yet



# That's all he ~~wrote~~ said

- Questions?
- dnorth@apache.org
- <https://people.apache.org/~dnorth/>