FDA U.S. FOOD & DRUG ADMINISTRATION

Open-Source Software for Regulatory Submissions: Challenges and Paths Forward

Paul Schuette
Scientific Computing Coordinator
FDA/CDER/OTS/OB
Paul.Schuette@fda.hhs.gov

Disclaimer

This presentation reflects the views of the author and should not be construed to represent FDA's views or policies.

Open Source

Open-source: denoting software for which the original source code is made freely available and may be redistributed and modified. (Oxford)

Some open-source software:

- R
- Python
- Linux

Statistical Software Clarifying Statement

FDA does not require use of any specific software for statistical analyses, and statistical software is not explicitly discussed in Title 21 of the Code of Federal Regulations [e.g., in 21CFR part 11]. However, the software package(s) used for statistical analyses should be fully documented in the submission, including version and build identification.

Statistical Software Clarifying Statement, cont

As noted in the FDA guidance, E9 Statistical Principles for Clinical Trials …"The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available." Sponsors are encouraged to consult with FDA review teams and especially with FDA statisticians regarding the choice and suitability of statistical software packages at an early stage in the product development process.

Link

CDER Submissions

- Submissions to CDER and CBER begun after December 17, 2016 are required to conform to CDISC standards (SDTM, ADaM).
- Other FDA Centers (CDRH, CTP, CVM and CFSAN) have different procedures and processes.
- The Study Data Technical Conformance Guide lays out technical expectations, in addition to CDISC.

Study Data Technical Conformance Guide

4.1.2.10 Software Programs

Sponsors should provide the software programs used to create all ADaM datasets and generate tables and figures associated with primary and secondary efficacy analyses. Furthermore, sponsors should submit software programs used to generate additional information included in Section 14 CLINICAL STUDIES of the Prescribing Information, if applicable. The specific software utilized should be specified in the ADRG. Refer to FDA Statistical Software Clarifying Statement for more information. The main purpose of requesting the submission of these programs is to understand the process by which the variables for the respective analyses were created and to confirm the analysis algorithms and results. Sponsors should submit software programs in ASCII text format. Executable file extensions should not be used. Link

Reasons to use Open-Source

1. Cost of software (although savings may not as great as some may claim).
2. Training/familiarity. Recent graduates are more likely to be familiar with R or Python than proprietary packages such as SAS.
3. Innovation. Thousands of R and Python packages are available, software for journal articles are often in R. Interactive data visualizations and dashboards created with Shiny are increasingly common.
4. Open-source solutions can be shared more readily, e.g. CRAN.
5. Performance. Can be easier to use open-source tools on clusters than proprietary tools. Licensing for cloud/cluster can be a concern for proprietary software.

Real and Perceived Challenges

1. Proprietary Software viewed as more stable.
2. Legacy Code written using proprietary software. Many companies have thousands of lines of code written using proprietary software.
3. Support. Proprietary software may have dedicated technical support that can be called upon.
4. Validation. Open-source products are not always perceived to be validated. Perception that open-source is more "buggy."

Challenges, continued

5. IT departments may need different skill sets to support open-source software.
6. Dissemination of results, making Shiny apps available to regulators and collaborators poses potential logistical issues.
7. Copyright/copyleft and intellectual property issues.

Solutions to Challenges

1. Stability. Most open-source software does change faster than proprietary products. Package management systems, well written programs can mitigate some/much of the instability.
2. Legacy code. Consider hybrid workflows. Use proprietary software for some purposes, open-source for others.

Solutions/Suggestions cont.

3. Support. Technical support may be an issue. Anecdote: found a bug in an R package developed by a NIH statistician, but was fixed over a weekend. Informed proprietary software developer of an issue, for which there has never been any followup.

4. Validation. Defer to Andy Nichols, R Validation Hub

Solutions/Suggestions

5. IT departments. IT departments will need to change and develop, CI/CD solutions may need to be implemented, change control processes streamlined.

Solutions/Suggestions cont.

6. Dissemination of results. Consider using a package archive such as CRAN as a method of distribution for both R programs and Shiny apps. Example: gsDesign, an R package on CRAN developed by statisticians and programmers at Merck, was used by Moderna for Covid-19 vaccine trials.

GitHub can also be a potential distribution venue, even with regulatory authorities.

Solutions/Suggestions

7. Intellectual Property issues. Generally, not a problem. However, if desired have your legal department look at MIT and other copyrights. Most government employees’ work is “public domain” and is subject to Freedom of Information Act (FOIA) requests.

Items for discussion

To date, CDER has not had a submission that is completely open-source from industry.

Is this a problem?

Potential paths forward:

1. Hybrid workflows. Use proprietary tools for some parts of the workflow, open-source for others.

2. Purely open-source solutions. May require replication of proprietary solutions. (Work in progress for some companies).

Emerging Trends

Software as a Service (SaaS)

Some proprietary software developers have begun to use cloud platforms to distribute software to users. Users pay access and/or usage fees on the cloud platform, as well as software fees.

This approach has also been proposed by some sponsors to distribute Shiny apps, etc.

Emerging Trends, cont.

Concerns with SaaS

1. Will this end up costing more?
2. Can software developers/providers guarantee confidentiality and privacy of queries and use?
3. Potential governance issues.
4. Transparency, reproducibility, intellectual property are potential concerns.

The time may come when only open-source can be run locally!

Emerging Trends

Simulations

- Complex Innovative Clinical Trial Designs (CID) Pilot Program
- Bayesian Designs and Analyses

Specialized open-source products such as JAGS and Stan can be used in combination with R, Python.

Emerging Trends

Larger, messier data sets are being used more widely:

- Real World Evidence
- Wearable technologies
- Genomic analyses
- NLP and Text Mining/Analytics
- AI/ML

Questions and Comments?

FDA U.S. FOOD & DRUG ADMINISTRATION