SOCIETY OF ACTUARIES

Article from:

Actuary of the Future

April 2013 – Issue 34

The Virtues of Open Source

By David Januszewski

David Januszewski, ASA, M.Sc., is an actuarial and Base SAS Certified statistics professional based in Toronto, Canada. He can be reached by email at dave.januszewski@gmail.com or from his LinkedIn page http://www.linkedin.com/in/davidjanuszewski.

What do a statistical programming language, a mathematical typesetting system, and a multi-paradigm computer programming language have in common? Answer: They are all licensed under the GPL.

The sources of R, LaTeX and Python are made available to everyone and they are completely free. These languages are licensed under the GNU's Not Unix (GNU) General Public License (GPL)—a very special license that has created a community adherent to the open source philosophy—one that prides itself in making resources available to beginners.

The open source community has been gaining acceptance across many businesses and fields of study. If we take a quick look at TIOBE.com, the website of a company specializing in tracking the quality of software, we could see that Python and Perl are in the top-10 programming languages. Both Python and Perl are licensed under the GNU GPL, which guarantees the end users of the software the freedom to copy, share, study and edit the software. This license was written by Richard Stallman, the same computer programmer who wrote the GNU Manifesto.

According to Stallman, all computer users will benefit because:

1. The sources will be available to everyone. So users do not need to contact the owner to make changes. Any available programmer can make changes.

2. Everyone can study the system and its code, and everyone is encouraged to make changes.

3. No one has to worry about who owns the software or what one is entitled to do with it.

Stallman writes "*I consider that the Golden Rule requires that if I like a program I must share it with other people who like it…. Copying all or parts of a program is as natural to a programmer as breathing, and as productive. It ought to be as free.*"[1]

As for statistical programming languages, it seems that R has made real advancement beyond academia and into the corporate world in recent years. R has caught up to its commercial counterpart, the Statistical Analysis System (SAS). Both appear to be equally used in business intelligence, data mining, research and many other areas (both hover around the 23rd to 25th most popular programming language as ranked on TIOBE.com).

Dr. Vincent Goulet from the University of Laval has made an interesting actuarial package called "actuar" on the Comprehensive R Archive Network (CRAN). However it seems that there are not too many other packages with actuarial models and functions on the CRAN network. I have not seen open source software sweep across actuarial science like it has with other fields. Nevertheless, I think it is possible for open source software to be used in pricing, valuation and modeling of complex insurance products.

GGY produces AXIS, SunGard produces Prophet, Milliman produces ALFA, and Towers Watson produces MoSes and TAS, but where are the open source competitors? Having an open source financial modeling and insurance valuation system would allow beginners and students to study and analyze the formulas and algorithms that are used in modeling complex insurance products. Actuaries would be able to edit the source code and make modifications based on their own judgment. Such a system would provide a greater exposure of the actuarial profession, it would be a means to validate the results of the commercial software competitors, and it would allow a greater involvement of academia in the profession.

Open source software can and does exist cohesively alongside proprietary commercial software. When looking at open source software vs. commercial software, it is interesting to look at the example of R and SAS.

SAS is a well-established statistical programming language and statistical modeling system that is widely used in many industries. The system was initially developed by Jim Goodnight and Anthony James Barr in the 1970s as a North Carolina State University (NCSU) research project to analyze agricultural data. By 1976, they had over 100 customers and so they decided to create the SAS Institute. Headquartered in North Carolina, the SAS Institute has since grown into the world's largest private software company. They provide comprehensive sets of user guides, study manuals and certification exams for their products. With over 4 million users, 60,000 customer sites, 13,000 employees, and $2.7 billion in revenue, SAS has become the worldwide leader in business analytics and statistical modeling software.[2]

R is a functional programming language that is relatively easy to learn, incredibly fast and highly adaptable. R was originally created by Ross Ihaka and Robert Gentleman as an implementation of the S programming language. Currently, the *R Core Development Team* maintains write access to the source of the R language. John Chambers, who had developed the S language while at Bell Labs, is now part of the *R Core Development Team*. The team also creates many manuals for the R language. The open source community has made several integrated development environments (IDEs) and graphical user interfaces (GUIs) that work with R for various purposes (e.g., RStudio, Rattle and the R Commander). Anyone can freely download and install the R programming language as well as the various IDEs and GUIs that provide additional features. Users can also create R packages that add to the base functionality of R and they can make their packages freely available on the CRAN.

In the book *SAS and R: Data Management, Statistical Analysis, and Graphics*, Ken Kleinman and Nicholas J. Horton mention advantages and disadvantages of both SAS and R. They mention how SAS maintains backwards compatibility and provides customer support to users, which is an advantage over R. Meanwhile, methodologists tend to release R functions to assist them in their work alongside their publication, so R tends to be more up to speed with recent developments in mathematics, statistics and other fields, which is an advantage over SAS. Another advantage of R over SAS is that you can always add more packages to R. In other words, you can always add more functions, algorithms, methods and models to the Base R system. Both R and SAS have a purpose, and I think that they should be used alongside each other.

Some studies have shown that open source software is growing at an exponential rate.[3] There are now over 50 free and open-software licenses that have been approved by the Free Software Foundation (FSF). Notable licenses include the GNU GPL, the Affero General Public License, the Mozilla Public License and the Apache License. Open source software hosting sites like SourceForge, GitHub and Launchpad claim to have a user base in the millions. From Web browsers such as Mozilla Firefox, to operating systems such as Ubuntu, to multi-paradigm programming languages such as Python, to Office suites such as Apache OpenOffice and software framework platforms such as Apache Hadoop, the open source software revolution is spreading around the world at an incredible rate.

END NOTES

[1] Sourced from Stallman, R., *The GNU Manifesto,* 1985, The Free Software Foundation. Available online at: *http://www.gnu.org/gnu/manifesto.html.*

[2] *http://www.sas.com/company/about/statistics.html* (sourced January 2013).

[3] Sourced from Deshpand, A., and D. Riehle, *The Total Growth of Open Source*, in Proceedings of the Fourth Conference on Open Source Systems (OSS 2008).

Available online at: *http://dirkriehle.com/publications/2008-2/the-total-growth-of-open-source/*