

Developing a clinical reporting system collaboratively under an Open Source software license

Ann M. Martin, UCB Pharma, Braine l'Alleud, Belgium

ABSTRACT

The Linux operating system and the Wikipedia are only 2 examples of successful collaborative projects. The proposal is to share our clinical reporting programs on the web under an open source software license. The aim is to create a library consisting of template SAS® programs written in SAS/Base® whilst making limited use of macro language in order to provide simple easily readable and verifiable programs. The library needs to have a series of clinical trial databases to allow quick testing of upgrades of the code versus independently verified analysis datasets and tables, figures and listings. The library should require sharing specifications, user manuals and test results.

INTRODUCTION

In his keynote address at the 'The Premier Business Leadership Series' April 2008 in London Don Tapscott highlighted the reasons why the WEB2.0 technology allowing mass collaboration is turning into a business trend. The initiatives mentioned in his book: LinkedIn, Innocentive, Linux, Project, MyTV, MIT OpenCourseWare, Technorati,...are all examples of mass collaboration where users provide updates and content. The book gives a good picture of why WEB2.0 and open source collaboration allows creating value. Of course everyone is familiar with Wikipedia, Google, youtube, skype and the many popular weblogs. For some time now Wikipedia and Google have become our first source of information whatever it may be: from 'how does the weather work?' to 'how do I remove correction fluid stains from the keyboard of our brand new computer?' Or 'let's look up the mystic term my IT colleague used' to 'what is the kind of cancer my friend has been diagnosed with?' It probably is not a coincidence that the research for this paper has been mainly conducted using the Wikipedia and Google.

The lists of open source and free software published in the Wikipedia¹ have led me to look further into this phenomenon and how it can be used to develop a clinical reporting system thereby sharing work and bringing innovation and cost reduction to the pharmaceutical industry. This can hopefully speed up development of remedies for patients with severe and very rare diseases such as DYT1 dystonia where only a few hundred patients are known to have the disease².

PROBLEM STATEMENT

Of the shelf, commercially available software typically means we get into expensive software and implementation projects which often are devil do-it-all packages. Many major companies and CROs build their own validated program libraries and in the process get exhausted by the validation requirements which are stopping them to introduce regular updates and enhancements. SAS Institute does not seem to undertake anything specific to the Pharmaceutical industry. As an aside it would be a major undertaking for SAS Institute to develop a clinical reporting tool that would satisfy the many possible clinical trial designs and the different requirements for studying a multitude of diseases. This requires the efforts of a large community of users.

FACTS AND RECENT DEVELOPMENTS

The Linux operating system and the Wikipedia are the 2 most famous examples of successful collaborative projects. The directories and lists of open source or free software packages show there is a wealth of open source software around. Whilst researching the subject, it did not take long to come across a list of open source healthcare software³. In the process of researching this subject it is intriguing to see that most of the open-source projects listed are related to the healthcare sector rather than the pharmaceutical industry. A notable application for the industry is OpenClinica⁴ a web-based application that facilitates electronic data capture in clinical trials.

In 2006, Tinazzi⁵ and co-authors noted that the number of projects made available through official Open Source is still limited. They found only 19 SAS related projects on sourceforge.net⁶ (a database of open source projects). Nowadays, in 2008, the number of projects has increased to 27.

SEVERAL WEBSITES

Several websites related to clinical research and the pharmaceutical industry are currently promoting open source projects. A snapshot:

Farmavita.Net – A community of Pharmaceuticals Executives have recently proposed a new business model of Open Source Pharmaceuticals. The project is targeted to development and sharing of know-how for manufacture of essential and life saving medicines. Farmavita.Net is now inviting proposals for 'open source pharmaceutical projects'. According to the website, Web 2.0 based software available at www.farmavita.net is allowing knowledge sharing and management of complex international, multi-center projects⁷.

CRIX International intends to offer an electronic information exchange for everyone involved in clinical research. The CRIX Collaborative Platform across the community aims to create a shared knowledge base, increased opportunities for collaboration, and the efficiencies of scale afforded through centralization. The board of directors includes the presidents of CDISC and HL7. Associate membership is free and allows participation in Special Interest Groups⁸.

SOME CONCEPTS

THE WEB 2.0

The term became notable after the first O'Reilly Media Web 2.0 conference and is a term that describes the trend in the use of the World Wide Web technology and web design that aims to enhance creativity, information sharing and most notably, collaboration among users⁹ It has allowed rapid development of software and content in a collaborative manner such as Linux and Wikipedia.

OPEN SOURCE AND FREE SOFTWARE

On Wikipedia, open-source software is described as computer software 'for which the human-readable source code is made available under a copyright license (or arrangement such as the public domain) that meets the Open Source Definition. This permits users to use, change, and improve the software, and to redistribute it in modified or unmodified form. It is very often developed in a public, collaborative manner. Open source software is the most prominent example of open source development and often compared to user generated content'¹⁰ (as one finds in Wikipedia). The software can be free or not.

The software is usually distributed according to a GNU General Public License, GNU Lesser General Public License. Two initiatives the Open Source Initiative¹¹ and the Free Software Foundation govern the rules and provide guidance. These licenses definitely do set legally binding conditions on the use of copyrighted work, and are enforceable under existing copyright law as recently established in a ruling by the US federal appeals court¹².

These licenses specify that anyone can use the programs for free at own risk and provided the author is acknowledged, proposed changes are made available to the author and other potential users. The programs could not be used as part of a commercially available software package.

SOLUTION?

What if the pharmaceutical community and healthcare organizations shared SAS programs for reporting on clinical data on a dedicated website under an open source software license? After all a program making a laboratory dataset cannot have that many variations. A large number of SAS programs for clinical reporting address the same kinds of statistical summaries and analyses over and over.

This should evolve into a peer-reviewed, collaborative development process in which the statistical programmers working in healthcare or pharmaceutical industry update the program library on an ongoing basis.

The library should consist of template programs written in SAS/Base making limited use of macro language in order to provide minimum requirements to allow a statistical programmer to utilize the programs for the analysis of a different study. Preferably only simple, easily readable programs should be posted. These template programs should address the reporting of data domains following the CDISC-SDTM standard. Output data and tables, figures and listings would have to follow the CDISC-ADM standard insofar the standard provides adequate guidance.

The benefit of doing this in an open collaboration model is that we potentially could get a large library of these programs in a very short time frame. Example is the growth of the Wikipedia

PhUSE 2008

which has known an unprecedented success. A few statistics: the content on Wikipedia is generated by more than 1 million registered users and has over 10 million articles in 253 languages, about a quarter of which are in English. The website started in March 1995, had 200 articles in the first month and reached 18 thousand in the first year. Its content is updated on an ongoing basis. The Wikipedia was updated within 20 minutes of the 2008 earthquake in China happening. Joe Biden was declared the running mate of Barack Obama on 23 August. An article on Joe Biden has since had over 62000 hits per day within 7 days. The 2008 Summer Olympics has had 168 000 hits/day in the month of August alone.

Clearly a SAS programming library will not generate this amount of interest. However there are a few highly active mailing lists Comp.soft-sys.sas (3657 members), Sci.stat.math (2298 members), Sci.stat.edu (1082 members) related to either SAS or Statistics through which an open source project could be spread. Other mailing lists are listed in the paper of Tinazzi and co-authors.

Why a wiki? A search on the keywords Wiki and SAS gave already over 20 000 results. Clearly the mediawiki software is the tool to organize a SAS program library collaboratively. A wiki is fast becoming the standard collaboration tool in use. The following wiki is currently the main collaborative online community for SAS users worldwide www.sasCommunity.org. The wiki statistics show the use of the website is increasing and content seems to gradually get richer. Even the criteria for posting programs on the library could be developed through open collaboration in this community.

BUSINESS CASE

Clinical trials have an immense amount of variety in terms of measurement instruments, designs and protocol design. Whereas many companies have a program library to address the safety data domains, a library of programs to process the efficacy data is often considered out of scope except for a few very large companies. Here are the reasons as to why it makes sense to share the SAS programs for analyzing all types of data:

PROQOLID A DATABASE FOR PATIENT REPORTED OUTCOME AND QUALITY OF LIFE INSTRUMENTS.

ProQolid¹³ has detailed information on over 620 instruments. In order to analyze data generated by each of these 620 different instruments one would need to generate 620 different SAS programs.

STANDARDIZED MEDDRA QUERIES (SMQS)

Implementation of the Standardized MedDRA Queries (SMQs)¹⁴ in safety signal detection is key in drug development. Several companies have implemented their own proprietary system* or implemented new commercially available applications. In principle a merge with the production SMQ worksheet and subsetting the AE tables for the listed adverse events is a fairly simple SAS program. Although MedDRA MSSO website has a lot of useful material such as the Introductory Guide for Standardized MedDRA Queries, the material is not freely accessible to all. Hence the slow adoption of the MedDRA tools.

CDISC-SDTM IMPLEMENTATION

SDTM implementation is rapid. Again each company is developing their own library of macros and tools to map legacy data to SDTM or implements commercially available solutions such as SAS Data Integration® studio. The fancy interfaces of these commercial solutions for a fairly simple task require dedicated specially trained teams to utilize these applications in order to

PhUSE 2008

perform the mappings. As the CDISC SDTM model does allow quite a lot of interpretation, there is quite a bit of variability in these implementations. So we get a SDTM format that is slightly different for each company. If companies were to share programs to get these mappings done, we could quickly expand the library of available programs to map data to an SDTM compliant format and speed up true standardization.

REGULATORY GUIDELINES

Regularly the regulatory agencies request information within the industry on selected adverse events. Recent examples are requests for information from each company on suicide related behavior, requests for information on co-morbidity in attention deficit. Typically each company tries to implement the request for information as best as possible. The results lead to a wide variety of approaches that point to new guidelines and rising costs. The implementation of these guidelines to provide information in a systematic way does not need to be re-invented for each company. The programs needed to provide this type of information could easily be shared on collaborative tools. "Think of how many organizations have programmed up international standards such as the RECIST¹⁵ criteria (Response evaluation criteria in solid tumors). The implementation of this international standard in our data analysis could have been shared.

WHAT ABOUT VALIDATION?

Providing SAS programs on a collaborative platform does not guarantee correctness of the program and there is an inherent risk in utilizing programs downloaded from the web. Ideally when posting a program on the web one should provide the data (if desired one could always blind the data!) and summary results together with some brief details in connection to study design and indication. When using such a program one still needs to validate the program when applying it to new data. Adaptations to the program may be necessary. Testing the revised program versus the original test data and confirming the program still gives the published study results that no intentional or unintentional changes were made to program. The procedure provides evidence of validation on the modified and enhanced program. Testing the revised program on the new data and ensuring its appropriateness ensures that the program does what it purports to do for the new clinical study data and therefore meets its goal.

Validation of this program library could be provided through an automated test framework that runs the programs on many test datasets, several operating systems on, for example, a nightly basis and thus confirm the programs are still providing the correct results. According to Tapscott et al. the company SpikeSource¹⁶ has managed to implement this innovation on a large scale and provides testing services for open source software. Whenever a new open source application comes through on a bulleting board, SpikeSource will test it and integrate it into a complete solution sometimes referred as a 'stack'. These solutions can be downloaded for free. SpikeSource makes its money providing customer service and support.

A similar model can be adapted for the pharmaceutical industry, a few experienced SAS programmers funded by a non-profit organization deriving funds from corporate sponsors can evaluate the posted program changes and decide whether to accept them or not. When accepting, they run the updated program through the test battery of clinical studies and confirm no other intentional or unintentional changes were made other than the documented changes.

CONCLUSION

Many pharmaceutical companies and contract research organizations have their own program libraries for the routine safety data domains: adverse events, laboratory tests, etc. Additionally they have invested in many utility macros. Each and every company went through an arduous, time consuming and expensive effort to validate these program libraries. Often validation gets in the way of providing timely updates to those program libraries. Yet given the validation, when using the program libraries once still has to carefully evaluate the use of these 'validated' programs given the very many different trial designs and data situations.

It is proposed to make the programs we create on a daily basis available through an open source collaboration allowing us to reduce the development costs of programs for all kinds of data. In this procedure statistical programmers would have to carefully evaluate a downloaded program in the light of the data it is being utilized for.

The opinions are those expressed by the author and not necessarily those of my company.

REFERENCES

See end of this text.

ACKNOWLEDGMENTS

With thanks to SAS Institute who invited me to 'The Premier Business Leadership Series' April 2008 in London where I learnt about the work of Don Tapscott and colleagues and which was my inspiration for the research presented in this paper.

Pierre Mayeur and Jean-Michel Bodart for a critical evaluation of the ideas presented in this paper and Eudokimos Hadjisotiriou for a close reading of the text.

RECOMMENDED READING

Tapscott, Don and Anthony D. Williams (2006, 2008) *Wikinomics How Mass Collaboration Changes Everything*. Published by Portfolio.

For a summary see YouTube - Authors@Google:
<http://www.youtube.com/watch?v=zF0k6dEm0zQ>

A. Tinazzi, M. Scott and A. Cattaneo 'SAS Programmer Resources on the web. Experiences Acquired Surfing the Net' PhUSE 2006

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Ann M Martin
UCB Pharma
Chemin du Foriest
B-1420 Braine l'Alleud (Belgium)
Work Phone: +32-2-386 35 57
Email: annmarie.martin@ucb-group.com

Brand and product names are trademarks of their respective companies.

¹ List of Open Source Software Packages

http://en.wikipedia.org/wiki/List_of_open_source_software_packages

² Tyler's Hope for a Dystonia Cure <http://www.tylershope.org/>

³ List of Open Source Healthcare Software

http://en.wikipedia.org/wiki/List_of_open_source_healthcare_software

⁴ Open Clinica website <http://www.openclinica.org/>

⁵ A. Tinazzi, M. Scott and A. Cattaneo 'SAS Programmer Resources on the web. Experiences Acquired Surfing the Net' PhUSE 2006

⁶ Sourceforge.net <http://sourceforge.net/>

⁷ Open Source http://en.wikipedia.org/wiki/Open_source

⁸ CRIX International <http://www.crixinternational.org/>

⁹ Web2.0 http://en.wikipedia.org/wiki/Web_2.0 Retrieved on 14Aug2008

⁹ Open Source Software http://en.wikipedia.org/wiki/Open-source_software citing Verts, William T. (2008-01-13). "Open source software". *World Book Online Reference Center*.

¹¹ Open Source Initiative http://en.wikipedia.org/wiki/Open_source_initiative

¹² Free Software foundation http://en.wikipedia.org/wiki/Free_software_foundation

¹³ ProQolid <http://www.proqolid.org/>

¹⁴ MedDRA standardized search queries
<http://www.meddrasso.com/MSSOWeb/SMQ/index.htm>

¹⁵ RECIST <http://www.recist.com/>

¹⁶ SpikeSource <http://www.spikesource.com/>