

A Framework For Achieving An Industry-Driven, Open-Source Clinical Reporting System

Dante diTommaso, F. Hoffmann-La Roche, Basel, Switzerland

Ann Marie Martin, UCB Pharma, Braine l'Alleud, Belgium

Paul Oldenkamp, Seattle Children's Research Institute, Seattle, WA, USA

Paul D. Hamilton, Amgen Corporation, Seattle, WA, USA

ABSTRACT

How many in-house reporting systems have been presented at recent clinical development conferences? Each has innovative features that spark interest in the development community, but each remains a separate, incompatible system that is unlikely to be replicated within any other organization. We could continue individually culling ideas that are feasible for our separate environments. Or we could combine the best of each into a collaborative project that leverages standards in a flexible framework.

Our purpose is two-fold. First, establish a clear rationale for a global, open-source clinical reporting library that builds on the success of industry standards such as CDISC's Study Data Tabulation Model (SDTM) and MSSO's Standardised MedDRA Queries (SMQs). We outline the benefits of participation for sponsors, regulatory agencies, individual programmers, and ultimately for patients who benefit from development transparency and efficiency. Second, outline a framework for a viable open-source project, minimum requirements to make adoption a realistic option, and an initial development plan.

INTRODUCTION

Each year, colleagues from around the world present clinical reporting systems that they have developed in-house. These typically begin with the traditional objective of standardizing listings, summaries and graphs to describe clinical trial data. Each organization maintains their customized system. Thus the effort is repeated throughout the industry, with each company absorbing the full cost of developing and maintaining remarkably similar, largely redundant applications.

Instead, the expertise gained from each of these isolated projects could be leveraged to pioneer a modularized, open-source system that would create efficiencies and business value. Within a relatively short period, a majority of standard reporting could be performed through a validated, open system. A well-designed system would allow departments to customize algorithms and presentation to suite local preferences. Describing to reviewers these customizations to an open, transparent application would be simpler than describing each novel system.

The authors describe a staged path to an open-source reporting system: establishing the project framework; adopting standards and methodologies; gathering requirements; and coordinating lifecycle activities.

RATIONALE FOR PARTICIPATION

Ann Martin (PhUSE 2008) previously reviewed projects that establish mass collaboration as a business trend, not simply an endeavor of hobbyists, and explored the requirements for a statistical reporting project. This year Paul Oldenkamp (PharmaSUG 2009, companion paper **PO06** by the authors) further establishes the economic rationale for industry's participation in world-wide collaboration, and describes tools that are currently available to support an open, global initiative.

Our proposal is not without precedent within clinical research. Within two years of inception, Akaza Research (from the Sanskrit word for free or open) launched OpenClinica, their web-based electronic data capture (EDC) platform built upon industry standards (see **Referenced Standards And Initiatives**).

This paper examines the potential benefits of applying existing models of open collaboration to the development of a clinical reporting system. We consider the programming community, regulators, and the industry as a whole; core challenges and decisions; and propose a development strategy.

INDUSTRY

Standards continue to emerge throughout clinical research, at an accelerating pace. This trend is coming late to the

PhUSE 2009

pharmaceutical industry. For decades other technology-based industries have enjoyed efficiencies based on standards (see Martin 2008).

Examples include common terminology for chemicals (drugs) and medical conditions (adverse events). The World Health Organization (WHO) first published the Anatomical Therapeutic Chemical Classification System (ATC codes) in 1976. In parallel, standardization (coding) of adverse event terminology has evolved from the early efforts of the FDA (COSTART) and WHO (WHO-ART) to the widely adopted Medical Dictionary for Regulatory Activities (MedDRA).

Momentum increases not only as standards emerge in parallel, but also as early standards provide a foundation for future extensions. Grading schemes for laboratory abnormalities, such as the U.S. National Cancer Institute's Common Terminology Criteria for Adverse Events (NCI-CTC) are at least partly based on the International System of Units (SI Units) for clinical laboratory data. Laboratory SI Units may also remind many people of the pains caused by loose adherence to a "standard".

We now have new opportunities to build on emerging standards. World-wide adoption of CDISC's SDTM introduces the opportunity to build open-source libraries and applications on a standard data model. As standards gain acceptance and prove their value, it is a natural course to leverage and extend them.

REGULATORY

Dr. Armando Oliva (PhUSE 2007) outlined the FDA's long-term goal of facilitating and streamlining the regulatory review process through the development of their Janus study data repository, a direct extension of SDTM 3.1.1 (see draft Janus specifications, Jan 2008). Already, the ICH's Common Technical Document relegates detailed data displays to appendices of safety and efficacy modules. Increasingly, health authorities conduct their own analyses of sponsors' data.

While sponsors will continue to design studies and advance methodologies, customized data presentation will become less accepted by health authorities and consequently stripped of business value. In our companion paper, **PO06**, the authors discuss this "commoditization" of software products; the natural progression of emergent technologies. When they first arrive, profit margins are high as producers of novel products are able to recover the costs of development. Once established, these technologies become commodities with minimal profit margins. Profits instead go to organizations that use mature technologies in innovative ways. In this context, the costs of maintaining independent, in-house systems becomes unjustifiable. A common application for presenting standard data displays would bring not only cost savings, but efficiency and transparency.

The ultimate beneficiaries are patients, as the industry continues to improve the efficiency, safety and transparency of the clinical development process.

PROGRAMMING COMMUNITY

The benefits to individual contributors are as varied as their motivation to contribute. Clinical programming typically requires extensive on-the-job learning. Especially at the beginning of a programmer's career, exposure can be restricted to specific tasks and techniques. Such focus may even be limited to skills that are not transferable to another study, therapeutic area, department or organization. It can take years for a person to learn the true nature and breadth of clinical development expertise. Participation in a global, collaborative project would provide exposure and expertise that are otherwise unavailable.

Programmers of all backgrounds and aspirations can find through mass collaboration opportunities to share and gain technical industry expertise, and express their creative energies in ways that are not otherwise available. An open-source project would also provide critical experience to programmers looking to break into the industry.

Streamlined and generalizable skill development directly benefits organizations. Furthermore, through direct participation and leadership, organizations can drive the open-source initiative. By contributing relatively few resources, organizations can harness the expertise of a global community.

ESTABLISHING THE FRAMEWORK

An open-source initiative must achieve broad participation to reach the ultimate objective of industry-wide adoption. A truly collaborative process is essential. Only representation of varied perspectives, opinions and requirements can lead the project down a viable path. As with related initiatives such as CDISC and Akaza/OpenClinica, support and sponsorship from within the industry would accelerate progress.

First we must prove our mettle. We need to address several decisions and technical challenges. The authors propose several of these, below, but emphasize that by definition we cannot answer them here; our purpose instead is to challenge and inspire the industry to resolve them collectively.

PROJECT FRAMEWORK

Many clinical programmers post the products of their experience on the web: technical tips, SAS® macros, even

PhUSE 2009

complex systems. But which are reliable? Which contain hidden assumptions based on individual experience? Was independent input, review, testing involved? What about flexibility, transparency, support or responsiveness new requirements? Doubts stemming from such central questions typically preclude the integration of these published tools into our development environments.

Adoption of open-source tools depends on us establishing a central framework based on transparent and rigorous software development principles. In April 2008, Paul Oldenkamp registered the Open Source SAS® Software Applications (OS3A) project on <http://sourceforge.net/> with this goal in mind.

SourceForge is a popular host of open-source projects, providing a version-controlled development environment with common public interfaces such as a project wiki and web portal. It currently supports over 150,000 open-source projects, making it the leader among such service providers (Wikipedia 2009). Other services do exist, each with distinguishing features. The first step is to settle on an environment; SourceForge is an obvious choice.

DEVELOPMENT STANDARDS

Once we decide where to work, the first priority would be to establish standards for participation and governance. Ann Martin (PharmaSUG 2009) has proposed that the clinical research industry agree on Good Programming Practices (GPP). Within OS3A we would further need to establish procedures for each stage of the development process: gather requirements, design, develop, test, validate, document, deliver, and maintain.

We would need to identify those standards that establish our foundation, ranging from a licensing model to SDTM, MedDRA, and ATC. In the absence of a dominant or stable standard, such as for so-called value-added or analysis data models, we would either identify emerging standards deserving support and refinement, such as CDISC's nascent ADaM specifications, or establish them from our own expertise.

Basic design considerations are also essential, whether establishing programming conventions recommended in the GPP or general principles such as separation of customizable business rules from fixed algorithms.

As with even simple programming tasks, establishing the ground rules presents the greatest challenge and requires the greatest care and discipline. Upfront investment is most critical to future success.

PROJECT SCOPE

Once we establish the mechanisms and conventions for managing our content, we can begin to design and develop software solutions. A planned evolution from basic, stand-alone utilities or template programs to sophisticated applications would provide opportunity to test and refine our approaches to development and governance.

A reporting system is only one candidate for collaborative development. Early participants in the OS3A project have already proposed others, such as a tool to read from and write out SAS data binaries, a set of CDISC/SDTM tools, a data integrity toolset, et cetera.

Any of these seemingly independent efforts could lead to a critical building block of a reporting system: a report registry, job queuing mechanism, report layout interface (cleanly separating analysis from display). Would we continue with the traditional focus on listings, summaries? Or could we leverage the latest hardware and software technologies into a graphical reporting system that also provides supporting listings and summaries? Would we have a graphical interface to facilitate use beyond the programming community? If so, what technology?

TESTING AND VALIDATION

Validation presents specific challenges. As every programmer knows, certain syntax and routines depend on the local environment (hardware and software). We would have to explicitly manage these dependencies during all lifecycle stages. We would have to test software on multiple operating systems, which would require multiple software licenses for products such as SAS.

This suggests that we would need support from within the industry for technical resources; SourceForge, obviously, cannot provide OS3A with the SAS licenses to test our software. We would almost certainly need support from vendors such as SAS, and an interface between SourceForge and our test environment.

SUPPORT AND MAINTENANCE PLAN

Without support or a commitment to responsive maintenance and enhancement, the industry could not adopt even a seemingly perfect system. Again, some commitment of resources by the industry seems essential, a point further developed in our companion paper, **PO06**.

STAGED DEVELOPMENT PLAN

We would have a lot to learn, new ways to work, and not all answers will be available from the start. A staged plan, however, would allow us to experiment with and refine our infrastructure before moving on to complex development projects.

PhUSE 2009

Initially, modest software objectives such as standalone utility macros or easy-to-read template programs for basic listings and summaries would provide experience not only with remote collaboration using new tools, but also with an automated testing environment. Once we become comfortable with these basics, and introducing new collaborators, we could effectively move on to delivering complex systems.

CONCLUSION

Creative energy within clinical programming communities is widely available. It is not difficult to find dozens of developers eager to share their expertise. All that is lacking is a focal point, a central framework. We could channel the contributions that individuals are already making, and draw in more participants by formally incorporating mass collaboration into the normal workday. This is already happening across technology sectors. It seems inevitable even for clinical research. There is no shortage of interesting, engaging, rewarding and valuable work to be done. So why not start now?

REFERENCES

Martin, Ann M. (PhUSE 2008), Developing a clinical reporting system collaboratively under an Open Source software license, <http://www.lexjansen.com/phuse/2008/ra/ra05.pdf>.

Martin, Ann M. (PharmaSUG 2009), Need for a Good Programming Practice for Clinical Trials, PO18, <http://www.lexjansen.com/pharmasug/>.

Oldenkamp, Paul, et al. (PharmaSUG 2009), Open Source SAS® Software Applications: the OS3A Program for Community Programming, PO06, <http://www.lexjansen.com/pharmasug/>.

Oliva, Dr. Armando (PhUSE 2007), "The Use of Standardized Clinical Trial Data during the Clinical Review of a Pre-marketing Application", Keynote speech, <https://www.phuse.eu/keynotes.aspx>

Wikipedia (2009), "Comparison of open source software hosting facilities", http://en.wikipedia.org/wiki/Comparison_of_open_source_software_hosting_facilities.

REFERENCED STANDARDS AND INITIATIVES

ATC/DDD (Anatomical Therapeutic Chemical/Defined Daily Dose) System: <http://www.whocc.no/atcddd/>.

CDISC (Clinical Data Interchange Standards Consortium): <http://www.cdisc.org/standards/index.html>.

CTD (Common Technical Document): http://www.ich.org/Main.jsp?@_ID=520.

GPP (Good Programming Practices for clinical trials), hosted by SASCommunity.org, http://www.sascommunity.org/wiki/Good_Programming_Practice_for_Clinical_Trials.

Janus specifications for STDM validation (Draft, Jan 2008), U.S. Department of Health and Human Services, Food and Drug Administration (FDA) and the National Cancer Institute (NCI), http://www.fda.gov/oc/datacouncil/janus_operational_pilot.html.

MedDRA (Medical Dictionary for Regulatory Activities), and Standardised MedDRA Queries (SMQ): <http://www.meddrasso.com/MSSOWeb/SMQ/index.htm>.

Note: The FDA switched from their COSTART dictionary to MedDRA in 1997 (<http://www.fda.gov/Medwatch/report/meddra.htm>). The WHO continue to maintain their WHO-ART dictionary for adverse reactions (<http://www.unc-products.com/DynPage.aspx?id=4918>).

NCI-CTC v3.0 (National Cancer Institute Common Terminology Criteria for Adverse Events), <http://safetyprofiler-ctep.nci.nih.gov/>.

OpenClinica, <http://www.akazaresearch.com/OpenClinica/features.html>, an open-source, web-based clinical trial software platform for electronic data capture (EDC) built on industry standards by Akaza Research, <http://www.akazaresearch.com/>.

PROQoLID (Patient Reported Outcome and Quality of Life Instruments Database), <http://www.proqolid.org/>.

SI Units for Clinical Data (International System of Units):

- JAMA instructions to authors (Feb 2009), http://jama.ama-assn.org/content/vol295/issue1/images/data/103/DC6/JAMA_auinst_si.dtl.
- Russ Rowlett, Russ (2005), A Dictionary of Units of Measurement, The International System - SI Units for Clinical Data in Medicine, University of North Carolina at Chapel Hill, http://www.unc.edu/~rowlett/units/scales/clinical_data.html.

PhUSE 2009

- Young DS (1987), *Implementation of SI Units for Clinical Laboratory Data*. Ann Intern Med.; 106:114-129, <http://www.ajcn.org/cgi/reprint/50/1/190>.

RECOMMENDED READING

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

CONTACT INFORMATION

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



Dante diTommaso

Statistical Applications Manager, Team Lead

F. Hoffmann-La Roche Ltd

Methodology and Innovation
Malzgasse 30 (B. 670 / R. 506)
CH-4052, Basel, Switzerland

Tel: +41 61 68 74314

dante.di_tommaso@roche.com



Ann Marie Martin

Director, Global Statistical Programming

UCB Pharma s.a.

Chemin du Foriest
B-1420 Braine l'Alleud, Belgium

Tel: +32-2-386 35 57

annmarie.martin@ucb.com



Paul Oldenkamp

Clinical Trials Systems Analyst/Prog.

Seattle Children's Res. Inst.

8321 Twenty-ninth Avenue NW
Seattle, WA 98117, USA

Tel: 206.884.7539

paul.oldenkamp@seattlechildrens.org

OS3A Project Administrator

Open Source SAS® Software Applications.

<https://sourceforge.net/projects/os3a/>

OS3A email:

<https://sourceforge.net/sendmessage.php?touser=2044988>



Paul D. Hamilton

Biostatistical Programming Sr. Mgr.

Amgen Inc.

1201 Amgen Court West
Seattle, WA 98119

Tel: 206.265.7689

hamiltop@amgen.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.