This article introduces some of the elements of existing frameworks and argues that compliance with CSV regulations can and should be achieved as a byproduct of systems engineering practices designed to maximize return on investment.

Computer Systems Validation: A Systems Engineering Approach

by Sameh Uzzaman

Introduction

A compliance-driven approach to Computerized Systems Validation (CSV) is widely prevalent in the pharmaceutical industry. As a consequence, the scientific body of knowledge which promotes validation in the pursuit of broader business objectives is frequently overlooked. In addition, concepts and processes are reinvented, often without reference to best practices outside the industry. This article introduces some of the elements of existing frameworks and argues that compliance with CSV regulations can and should be achieved as a byproduct of systems engineering practices designed to maximize return on investment.

Background

As expressed in its 11 January 2002 Final Guidance for Industry and FDA Staff,[1] the FDA believes that "Information on software validation presented in this document is not new. Validation of software, using the principles and tasks listed in [this document], has been conducted in many segments of the software industry for well over 20 years."

Government agencies such as NASA, DARPA, and DoD, their suppliers, and academic research institutions were among the first to develop and require quality assurance strategies for software-based systems. They shared the current FDA view that computerized systems are fundamentally 'different' from manual, hard-wired, or mechanical systems, even when designed to perform the same basic functions.

The core reason for the difference is the complexity of software-based systems, expressed via large numbers of decision points, intricate interrelationships, multiple levels of abstraction, and the relative ease with which they can be modified. Validation of these systems requires control of the process by which they are developed and utilized because it is not feasible to test them comprehensively after they are already built. This systems engineering objective is typically expressed as "verification and validation" to emphasize the need to ensure not only that "the system gets built right," but also that "the right system gets built."

While the process-oriented quality concept is familiar in the pharmaceutical industry, there is less clarity on the form and sequence of activities required to support the development and use of compliant computerized systems, or how such processes can be implemented with finite resources. Activities which go beyond the writing of detailed specifications and qualification testing are often poorly received and treated simply as compliance overhead.

In fact, significant evidence exists (in the form of standards and studies[1-2, 10, 16-20]) which indicates that commitment to targeted and gradual refinement of selected organizational practices can improve not only compliance, but

Figure 1. IEEE life cycle implementation.
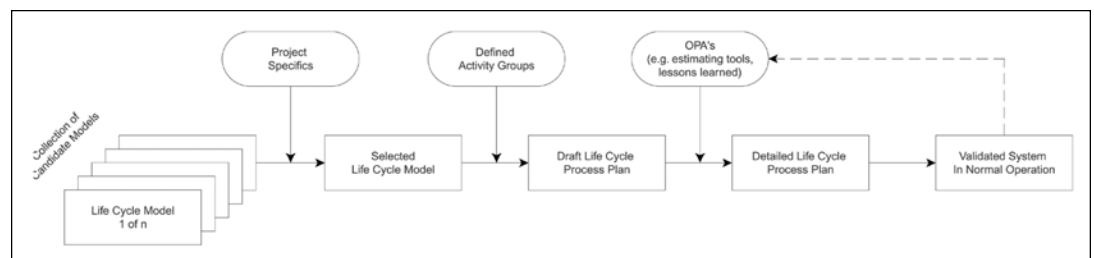
also the quality and reliability of computerized systems. This article introduces examples of three key elements of a strategy for the adoption of systems engineering practices which improve the return on investment in CSV:

1. Alternate life cycle models (in lieu of the familiar waterfall or "V" model) which can be used by organizations to leverage their particular strengths and resources in order to implement cost-effective compliance practices

2. Definitions of the management, systems, and software engineering processes which should be implemented and internalized by organizations seeking to be compliant with best practices as well as regulations

3. An established set of measures which can be used to investigate, quantify, and improve the effectiveness of essential organizational practices related to systems engineering.

What is a Life Cycle Model?

System architects typically construct models in order to learn from and to *control* the functionality of a system according to a defined set of priorities. Since real-world systems are complex, it is normal to introduce approximations which reduce the absolute fidelity of their models. Systems theory generally requires that all such approximations maintain enough accuracy to permit a meaningful investigation of the desirable properties of a system, including stability, controllability, observability, and robustness.

The objectives underlying the use of a Life Cycle Model as a basis for verification and validation are the same as those of a system architect. It permits the assessment and control of a system of defined elements (including business or organizational functions) which can be harnessed to generate a (controlled) work process and create required work products in a favorable (cost-effective and reliable) manner.

What a life cycle model is categorically *not* is a process of creating user specifications, turning them over to a vendor,

and then performing qualification testing on the computerized system shipped back by the supplier. While these activities are important elements of a life cycle approach, they are not likely to generate significant value without an overall framework to support their execution.

How is a Life Cycle Model Implemented?

It is not by accident that organizations from the ISO and IEC to the FDA, ANSI, and IEEE recommend the use of life cycle models for the development of validated computerized systems, but do not actually specify the implementation of a particular model or the development of a prespecified set of deliverables.

This open-ended requirement is entirely consistent with the underlying rationale which is that creation of a supporting framework is the critical first step and must be carried out in a consistent and coherent manner. The specific actions or deliverables which support the framework emerge as a consequence of business opportunities and risks and cannot be specified by an outside agency.

The systems-oriented or model-based approach also supports the traditional role of outside auditors by allowing them to focus on verifying that the existing elements of a framework are consistent with its stated objectives, and that defined activities are carried out as specified. External regulators can also specify additional requirements in a relatively structured manner, including those which an organization might otherwise exclude on the basis of low perceived risk, as in the case of 21 CFR Part 11.

For an introduction to the specific actions which an organization should take, more concrete recommendations are provided by existing engineering standards, such as IEEE 1074.[2] This standard actually specifies a list of 17 activity groups for which it expects compliant organizations to build a supporting infrastructure. It also provides information such as precursors and expected outcomes of specific activities.

The main concept of this standard is that compliant organizations will devise a family of templates - including specific models - which suggest how to sequence, schedule,

Capability Level 1 Initial (Performed Informally)	Base practices of the process area are generally performed but the performance of these base practices may not be rigorously planned and tracked. Identifiable work products for the process exist to demonstrate the performance of the base practices, but the process itself is characterized by frequent crises and a pattern of unforeseen delays, unpredictable outcomes countered by heroic individual effort, and process flow that is overly dependent on the individuals involved
Capability Level 2 Repeatable (Planned & Tracked)	Base practices of the process area are planned and tracked. Performance according to specified procedures is verified. Work products conform to specified standards and requirements. As a result of tracking the process using predefined metrics, planning & control can be based on experience, immediate corrective action is possible if problems are identified, and realistic costs and schedules can be predicted for future projects
Capability Level 3 Defined (Well Defined)	Base practices are performed according to a well-defined process using approved, tailored versions of standard, documented processes. The primary distinction from the planned and tracked level is that the process is planned and managed using an organization-wide standard process which includes continual efforts to improve quality and productivity
Capability Level 4 Managed (Quantitatively Controlled)	Detailed measures of performance are collected and analyzed. This leads to a quantitative understanding of process capability and an improved ability to predict trends which lead to deviations. Statistical quality controls can be used to distinguish between random deviations & meaningful violations
Capability Level 5 Continuously Improving	At this level, an organization establishes quantitative performance goals (targets) for process effectiveness and efficiency, based on its business goals. In other words, at level 2 an attempt is made to find & correct faults but at level 5 defect prevention is practiced in order to ensure there are no faults in the first place

Table A. Description of maturity levels.

and track standard activities according to the requirements of a particular project or business process. An approved template may be selected on the basis of any combination of specific factors such as the novelty of a project, the number of project participants, consumer-provider relationships, contracts, schedules, and the attendant risks - *Figure 1*.

Once a model template has been chosen, it can be associated with the necessary activity groups to develop a baseline project execution plan which identifies, for instance, the need for appropriate specifications, prototyping or builds, review cycles, and training.

The term Organizational Process Asset(s) (OPA) is used in the standard to indicate the body of knowledge or business intelligence that an organization relies on to conduct its activities. Typical OPAs include:

- organizational policies
- defined process descriptions
- estimating and operating procedures
- development plans
- quality assurance plans
- training materials
- process aids, including checklists
- lessons-learned reports

The OPAs are used to quantify and plan the required levels of effort and resources over the life of the project. As reflected by the dotted line in Figure 1, feedback from completed projects should be used in an ongoing effort to improve the accuracy and quality of the OPAs.

It should be noted that the OPA 'library' is a conceptual, not necessarily a physical, entity. However, ongoing improvements in organizational capability and efficiency will be reflected by more streamlined and centralized management of OPAs, including progressively sophisticated documentation through a process of iterative refinement.

In practice, the consistent management of OPAs is usually one of the hardest goals to accomplish when implementing the model-based approach. This heterogeneous body of knowledge, much of which may not be in recorded form, must gradually be captured in a manner which allows:

- design and implementation of an OPA collection or library, including the library structure and support environment
- specification of the criteria for including items in the library
- specification of the procedures for storing and retrieving items
- entry and cataloging of the selected items into the library for easy reference and retrieval
- provision of training and making the items available for use in projects
- periodical review of the use of each item and action on the results to maintain the library contents
- revision and change control to accommodate new items, retire obsolete items, and update revised items
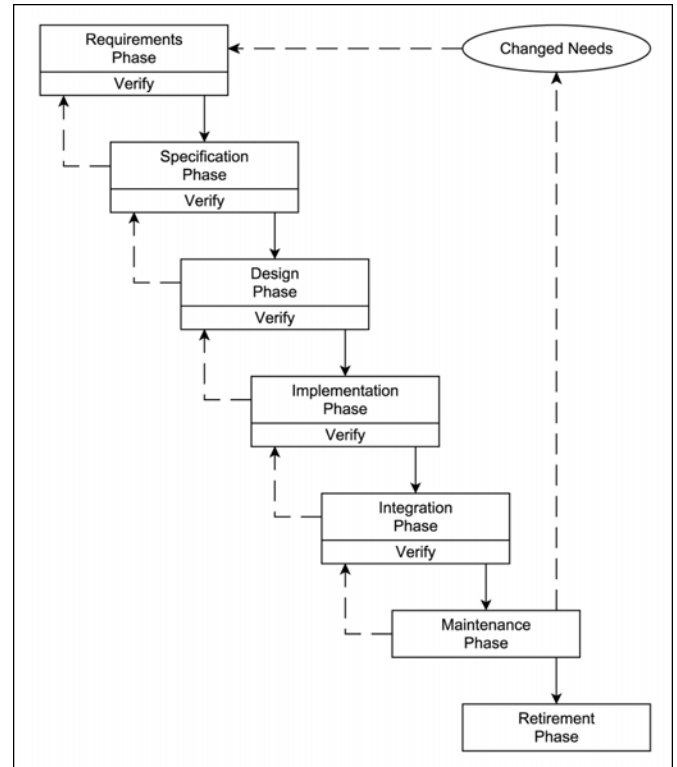

Figure 2. (Enhanced) waterfall model.

Selecting a Model

While business models are generally familiar to upper management, their formal extension to computerized system design and utilization is not widely understood. The biggest single misconception is that there is one 'best,' 'approved,' or 'cheapest-to-implement' model, and that it is required only in order to satisfy compliance requirements.

In fact, each model simply represents one of a family of strategies for system planning and control. The existing models have been formalized partly as a reaction to disorganized attempts at system development which often had unintended consequences such as late delivery, bugs, software ill-suited to user needs, and excessive expenditure on customization after delivery.

These models have overlapping characteristics because they all attempt to specify a well-defined and methodical sequence of steps for achieving the objective of creating the 'best possible' system within the resources available. Their intent is to define the required inputs, resources, and outputs at each step in the life cycle and then ensure that these are provided for and verified at each stage during the execution of life cycle processes.

The appropriateness of a specific model is determined by factors including risk (in terms of both business and compliance) as well as project resource constraints (such as financing, people, and technology). The most significant difference between models is the manner in which the associated process unfolds: rigidly sequential like a schedule with iterative reviews of progress and assumptions, or recursively returning to the drawing board on a regular basis before making the final decision to continue with a project.

Model Types

While a reiteration of all the available models is beyond the scope of this article, several cases which more or less span the range of possibilities are outlined here for illustrative purposes. Other variations include, for example, evolutionary, incremental, and rapid-prototyping models.[11-15]

The **waterfall model** is the basis for the **V-model** that is now fairly well known in the pharmaceutical industry. In its most basic form, it is a once-through document-driven model in which system development "flows downhill" in phases from requirements definition, through implementation, to final testing and use.

While this model fits well into a management schedule, it is often implemented in a manner which overlooks the inherent difficulty of developing systems without in-process reviews, risk-assessment, testing, and document revision. Complications which may arise as a result of using this model include:

a. It is often difficult for end users to state - or even visualize - all requirements explicitly at the start of the life cycle.
b. The process can result in the development of large quantities of documentation in the early stages of a project when it may be hard to comprehend without a point of reference, such as a prototype or simulation. This can reduce the effectiveness of reviews.
c. The first time the end user sees a working product may be after it has been coded, when re-design is problematic. This can result in products that don't meet actual needs.

Project planning based on the waterfall model is best suited to situations where requirements are well known and stable, and supporting technologies are clearly understood.

In practice, organizations attempting to use this model often fail to define and approve specifications until after a system has evolved into its final form. This deprives the development process of the very controls which provide value in the basic waterfall model. As shown in Figure 2, one solution to this problem is to use an iterative waterfall model which makes allowances for a methodical return to an earlier phase on the basis of newly acquired information.

The **spiral model** is a relatively sophisticated alternative approach shown in Figure 3. According to its originator, Boehm,[3] a typical cycle of the spiral begins with the identification of:

- the objectives of the product
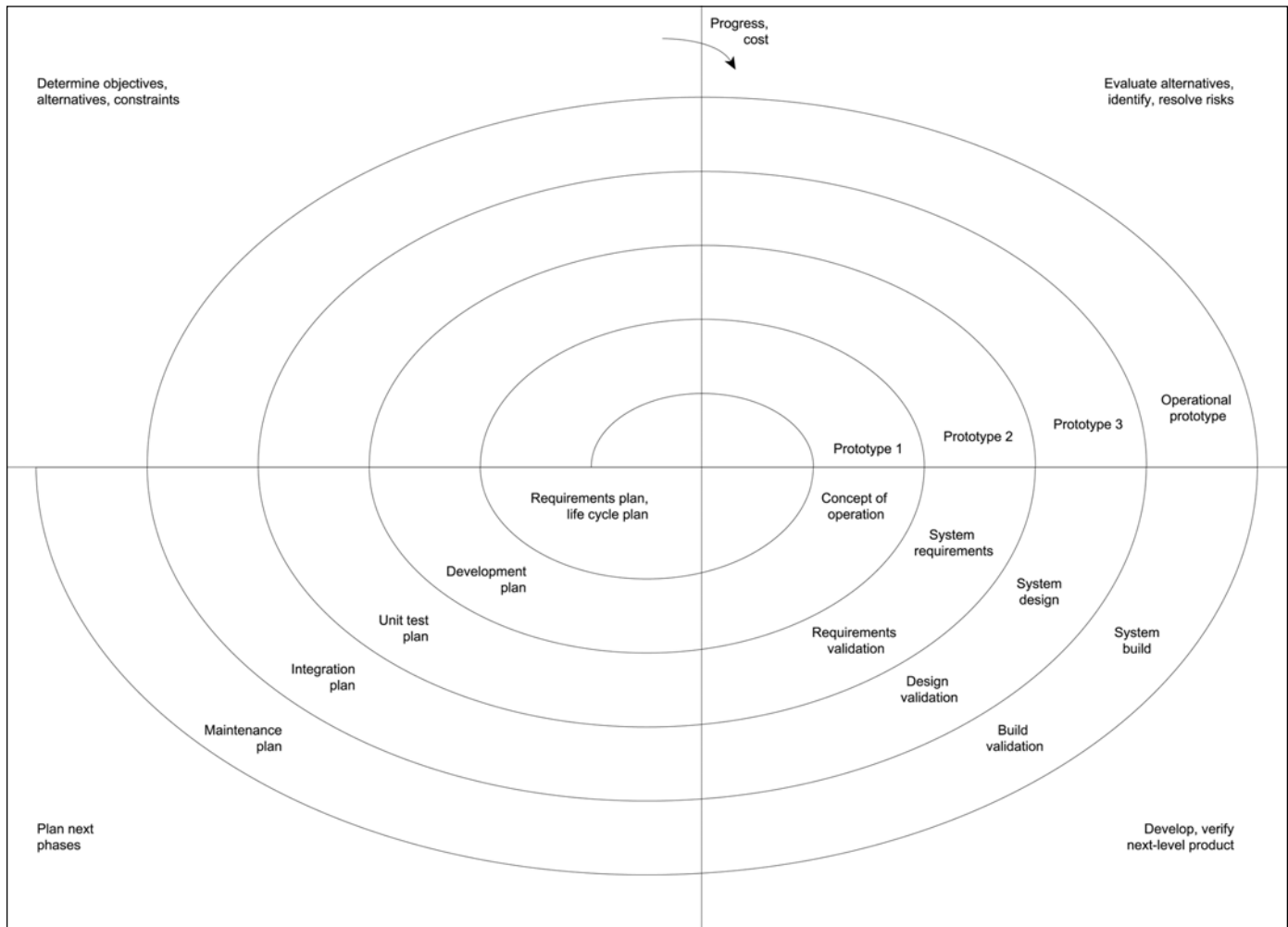- the alternative means of implementing this portion of the product

Figure 3. The spiral model.

- the constraints imposed on the application of the alternatives

The alternatives are evaluated as part of risk management. The evaluation may involve "prototyping, simulation, administering user questionnaires, analytic modelling, or combination of these and other risk-resolution techniques."

The next step is designed to permit "a minimal effort to specify the overall nature of the product, a plan for the next level of prototyping, and a development of a much more detailed prototype..." The next step also may revert to "follow[ing] the basic waterfall approach, modified as appropriate to incorporate incremental development."

Each cycle is completed by a review, with go/no-go plans made for the next cycle. As the spiral lifecycle progresses, the overall cost of the product is determined by the radius of the spiral, and the progress is determined by "angular displacement."

The key idea of the spiral life cycle is to minimize risk using methods such as building prototypes and simulations. Explicit attempts are made to identify potential future problems, not just in initial stages of design, but also later, when more has been learned about the problem and the design. The approach is most useful in common situations where many options, requirements, and constraints are unknown, and the cost of poor decisions is likely to be significant.

Models for Vendor-Supplied Systems

A practical consideration which bedevils discussions of model-based approaches is that of systems developed more or less entirely by vendors or contractors. It is important to remember that:

a. The ultimate responsibility for the quality of a system lies with the purchaser, not the supplier. The costs associated with flaws and reengineering are also borne disproportionately by the purchaser, particularly toward the end of a project.
b. It is possible to structure the procurement process in favor of suppliers with mature quality systems or to select models which grant the purchasing organization greater oversight privileges.
c. The life cycle does not begin with an RFQ, nor does it end with commissioning or PQ. It is in an organization's own best interests to ensure that each purchased system can be integrated into its user environment and maintained in a controlled manner with or without further involvement of the supplier.

For these reasons, when planning projects involving multiple organizations, it is generally advisable to select a model which requires judiciously determined amounts of interaction between participants. Models which allow development in a vacuum until some form of final testing may not constitute a particularly effective approach.

If the purchasing organization does not have the in-house resources to implement meaningful oversight, it is often

AT&T Labs	Boeing	Ernst and Young
General Dynamics	Honeywell	KPMG
Lockheed Martin	Motorola	Northrop Grumman
Raytheon	Reuters	TRW

Table B. Some partners in the development of CMM.

advisable to delegate some of those responsibilities to an independent third party rather than to eliminate them. This relationship may be leveraged to develop additional internal resources.

Life Cycle Activities

In view of the fact that structured processes are necessary to build validatable systems, it is fortunate that considerable effort has already been invested in defining the components of these processes and that IEEE/ANSI 1074 and ISO/IEC 12207 are dedicated to this proposition. IEEE 1074 clearly identifies 17 required activity groups which form the basis of such processes and organizes them into the following five functional areas:

1. Project Management
2. Pre-Development
3. Development
4. Post-Development
5. Integral

By creating the infrastructure and resources to support these activity groups (with gradually increasing sophistication, if so indicated), an organization transforms validation into a documentation of its own internal practices, including prudent and diligent testing of the computerized systems which it uses. Compliance with newer regulations such as 21 CFR Part 11 becomes a simple matter of ensuring that system requirements are defined in a compliant manner and that defined requirements are satisfied by each product.

The aim of the life cycle approach, which is to reduce uncertainty and risk by providing a well-defined control architecture, can then be achieved because this methodology actually *decreases* the overall complexity of the process of creating and utilizing computerized systems.

Project Management Activity Groups

Activity groups listed in the management category reflect not only those included in conventional deliverables, such as a Validation Master Plan. They also require the *controlled documentation* of commonly overlooked elements such as initial assumptions, required supporting resources (including people and procedures), and roles and responsibilities, as well as the metrics to be used to monitor and control progress and compliance throughout the SLC.

Initiation

In general, initiation may refer to the point in time when the need for a system is first formally expressed, or the beginning of a new stage in the development of a system.

Project initiation activities include:

1. selection or development of a project execution framework (or adoption of an existing plan template, model, or SLC framework)
2. creation of an estimate of required resources based on available information
3. allocation of resources, including clear delineation of roles and responsibilities
4. selection of metrics for estimating progress and compliance with adopted standards

Planning

Planning activities defined by the standard are expected to address all project requirements as well as contingencies, including:

1. evaluation of progress or deliverables
2. configuration management
3. system transition
4. system installation
5. preparation of documentation
6. provision of training
7. project management strategy
8. integration of subsystems

Monitoring and Control

Monitoring and control activities include:

1. management of risks, including evaluation, resource real-location, facilitation, and expediting
2. management of project scheduling and progress information
3. identification of opportunities for continuous improvement of SLC methodology
4. record retention
5. collection and analysis of supporting data

Pre-Development Activity Groups

Concept Exploration

Documentation of concept exploration is required because it forms the basis of many subsequent activities and decisions. While it may not be necessary in all cases to commission a formal study to refine a concept or to generate a feasibility report, the decision making process should not be arbitrary or unclear. Concept exploration activities should include:

1. identification of ideas or needs
2. formulation of potential approaches
3. investigation of feasibility
4. refinement and finalization of the idea or need

System Analysis and Resource Allocation

The system analysis activity group is the bridge between concept exploration and the definition of functional requirements. It maps the required workflow, business process, or operation to automated, mechanical, and procedural components (i.e. software, hardware, and people). Systems acquired without adequate analysis of the resources needed to maintain and integrate them into their working environment, or to modify the business process that they automate, frequently lead to costly implementation problems later in the life cycle. System analysis activities include:

1. analysis of system requirements
2. development of a system architecture
3. decomposition of system requirements (into components definable in terms of explicit functional requirements)

Importation of Pre-Existing Components

In certain cases, some or all new system requirements may best be satisfied by reusing or acquiring an existing system (or vendor-provided component). In these cases, definition of the manner in which an imported component is to be utilized and managed still requires the use of the life cycle approach, but some design activities may be substituted by importation activities. These include:

1. identification of imported system requirements
2. evaluation of system import sources
3. definition of system import method
4. system import

Development Activity Groups

Requirements

This group includes those activities that are directed toward the development of component requirements. Requirements activities defined by the standard are:

1. definition and development of component functional requirements
2. definition of interface requirements
3. prioritization and integration of component requirements

Design

The objective of the Design activity group is to develop a coherent, well-organized representation of systems that meet their requirements. At the architectural design level, the focus is on the components that comprise the system, and the interfacing of those components. At the detailed design level, the emphasis is on the internal structure and algorithms for each system component. Design activities include:

1. architecture design
2. database design
3. interface design
4. detailed design

Implementation

The Implementation activity group creates a realization of a system from its requirements (subject to controls included in the Integral activity groups). Implementation activities include:

1. creation of executable code or fabrication of hardware
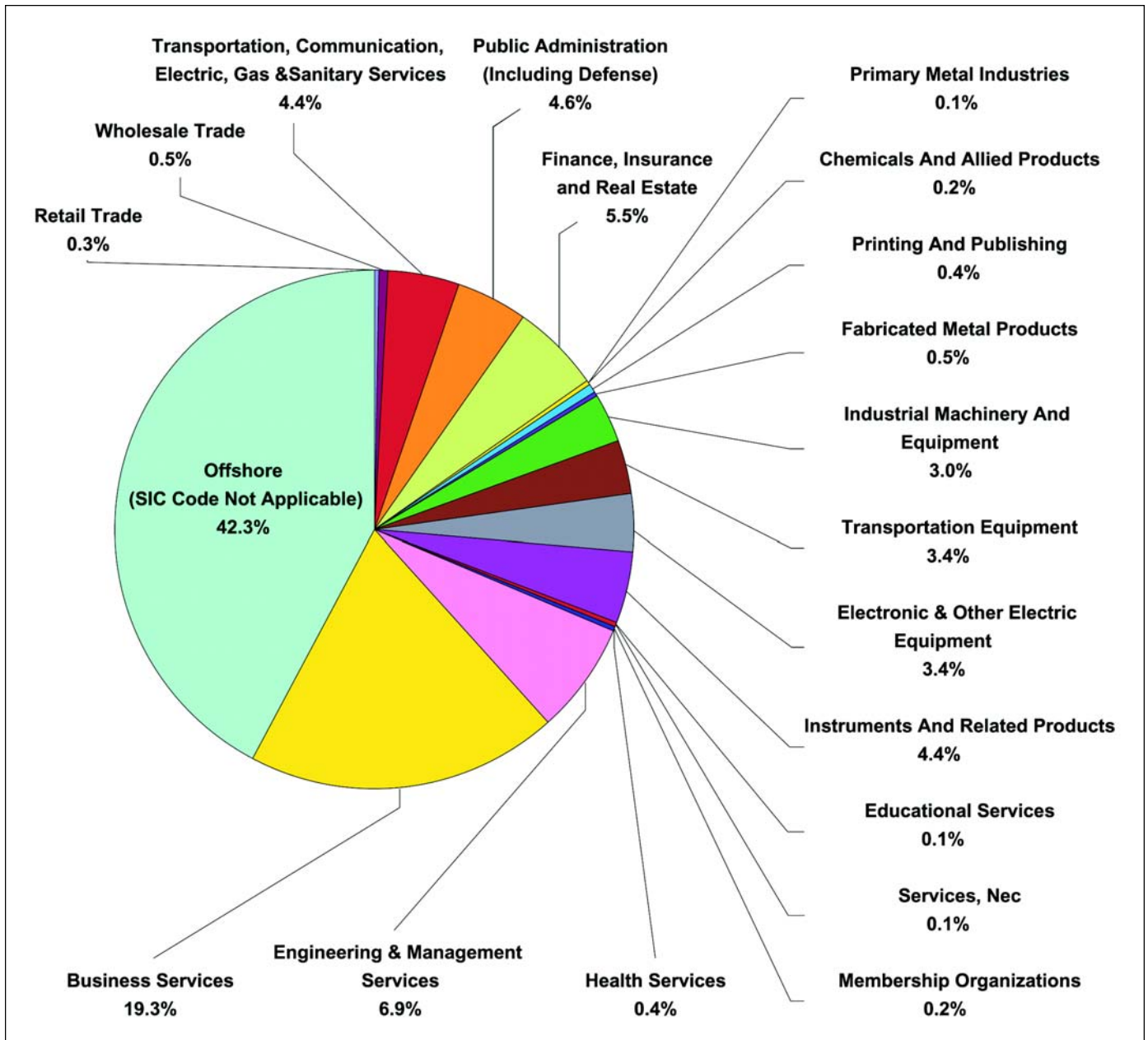2. creation of operating documentation

Figure 4. CMM report respondents by sector.

3. system integration

Post-Development Activity Groups
Installation
Installation activities involve the transportation of a system from its development environment, installation, and checkout in the target environment(s). The standard activities include:

1. system distribution or shipment
2. system installation
3. acceptance of system in operational environment

Operation and Support
The Operation and Support activity group supports organized utilization of a system. It may trigger maintenance activities (via related monitoring and control activities) but is itself intended to include only:

1. system operation (in a controlled manner)
2. provision of technical assistance and consulting
3. maintaining a support request history

Maintenance
The Maintenance activity group includes preventive and corrective activities, as well as enhancements (including iterations of development according to life cycle processes). Maintenance activities include:

1. identification of system improvement needs
2. implementing a problem reporting method
3. reapplication of system life cycle methodology

Retirement

Retirement consists of the actual removal and archiving of a system from regular usage. It could be spread over a period of time and take the form of a phased removal, or it could be the simple removal of the entire system. Retirement activities include:

1. user notification
2. parallel operation
3. system retirement

Integral Activity Groups

These groups include activities which are necessary for the successful completion of projects and the quality of computerized systems but which are not normally limited to one particular phase of project execution.

Evaluation

Evaluation activities are designed to uncover defects in a system or the processes that are used to develop the system. They include:

1. execution of reviews
2. creation of traceability matrices
3. execution of audits
4. development of test procedures

5. creation of test data and resources
6. execution of tests
7. reporting of evaluation results

Configuration Management

Configuration Management activities are required to track and control the evolution of computerized system components and products, both during the initial stages of development and during all stages of maintenance. They include:

1. development of identification conventions
2. implementation of configuration control
3. performance of status accounting and reviews

Documentation

Documentation activities are required to plan, design, implement, edit, produce, distribute, and maintain those documents that are needed to support system development and usage.

Documentation includes product-oriented materials (user manuals, SOPs) and procedure-oriented materials (standards, conventions) for internal and external users. Examples of internal users include those who plan, design, implement, or test systems. External users include those who install, operate, apply, or maintain the system.
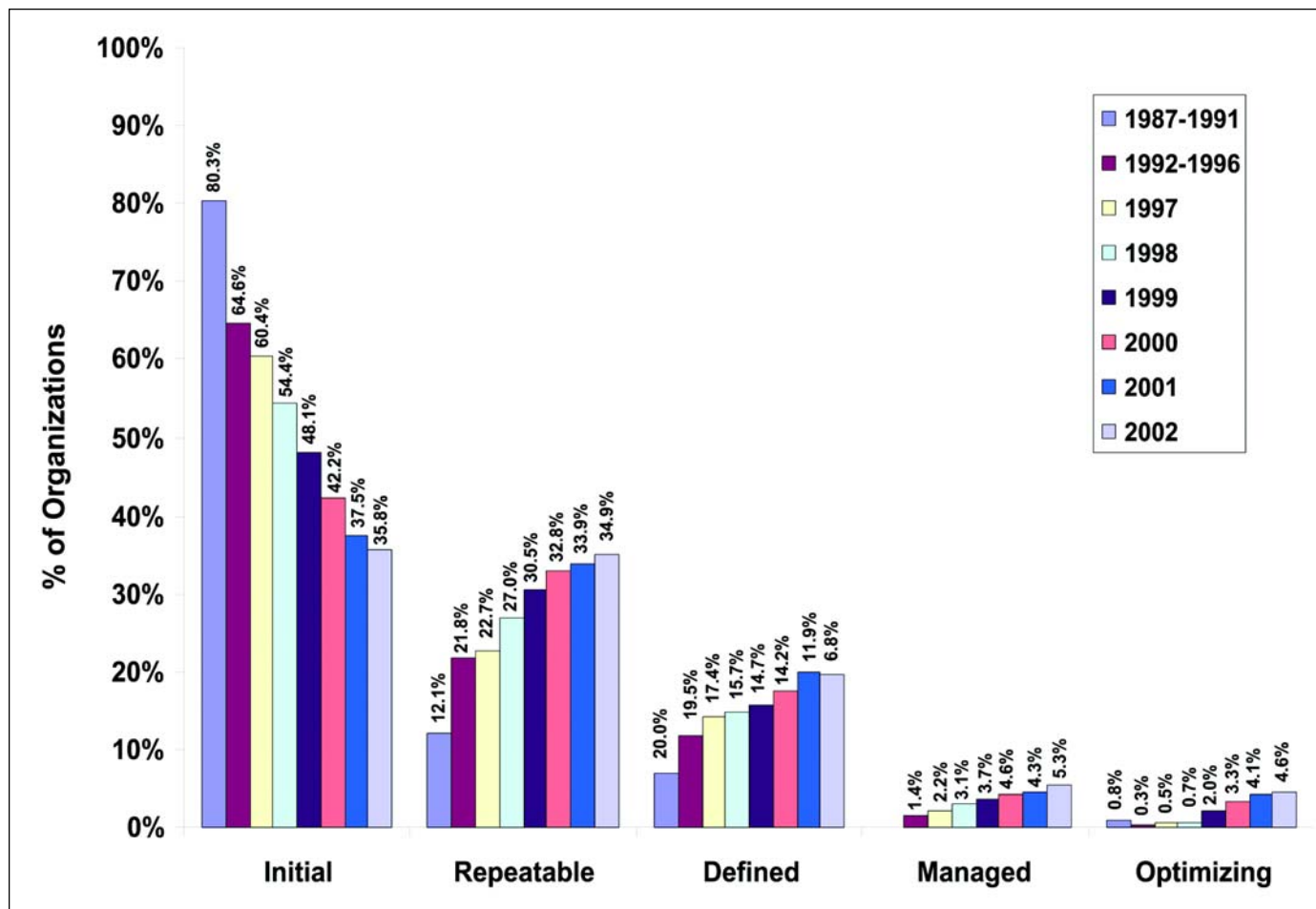

Figure 5. Historical development of maturity profiles.

Documentation activities must provide timely system documentation to those who need it, when they need it. These needs determine the number of documents required, as well as the phases of the life cycle during which they should be developed. They include:

1. preparation of documentation
2. production and distribution of documentation

Training

The development of quality systems is dependent upon knowledgeable and skilled people. These include managers, system developers, the user community, and operations and maintenance staff. It is essential that the planning of training requirements is completed early enough in the life cycle to ensure that personnel can be prepared to apply the required expertise. Training activities include:

1. development of training materials
2. validation of the training program
3. implementation of the training program

Capability Maturity

While it is easy to argue that better organization of work processes is beneficial, it seems harder to justify and fund potentially costly and disruptive organizational change when its chief contribution is measured in terms of *problems that do not occur*. It is also not immediately clear how long the effort may take, or how quickly the benefits may be observed.

In fact, strategies for improving system engineering processes - irrespective of the life cycle model chosen - have been widely studied and documented. Results from one popular approach are provided here because they provide an uncommonly lucid and intuitive introduction to the objectives, mechanics, and benefits of process optimization.

Developed by the Software Engineering Institute (SEI) at Carnegie Mellon University, and sponsored in part by the Department of Defense, the Capability Maturity Model (CMM) family of strategies investigates the premise that the risk of project failure and cost overruns decreases with increasing organizational maturity. It attempts to quantify the benefits of working from the assumption that system or software development techniques are not the problem, their management is, and that improved management leads to improved techniques.

The CMM approach recognizes that, in the majority of organizations, there is room for *gradual* improvement in the 'maturity' with which the systems engineering process is managed or controlled. In order to develop benchmarks, CMM defines five levels of "maturity" designed to induce and assess change incrementally; these are described in Table A.

A thorough discussion of CMM goals and techniques is available from the SEI Web site (sei.cmu.edu) which also provides access to the results of a series of ongoing assessment studies dating back to the mid-1980s; some essential aspects of this research are reproduced here.

The partners in this research and the subjects of the associated studies are a mix of suppliers to the US government as well as government, independent, and foreign organizations seeking to improve quality and performance and to receive recognition for their efforts. As shown in Figure 4, although the participants have included representatives from a wide range of businesses, pharmaceutical sector participation is not yet significant. Some major participants are shown in Table B.

Participants and Benefits

CMM data has been collected through two primary mechanisms, CMM® Based Appraisals for Internal Process Improvement (CBA IPIs) and Software Process Assessments (SPAs). The most recent report is based on data from 1756 organizations and 2325 CBA IPIs and SPAs returned to SEI up to July 2002.

As shown in Figure 5, participants have demonstrated consistent movement toward higher maturity levels by implementing model based frameworks. The commitment to the CMM approach has been sustained both by profitability (reported at the SEI Web site) and by increasing reliance on certification in qualifying for government contracts.

Among the most significant results of this research for the pharmaceutical industry is a quantification of the expectations. For organizations that began their CMM-based SPI effort in 1992 or later, the median time to move between maturity levels was:

Level 1 to 2 - 23 months
Level 2 to 3 - 22 months
Level 3 to 4 - 28 months
Level 4 to 5 - 17 months

This suggests that the implementation of structured processes must be treated as an evolutionary process in itself – there is no quick fix. While an external organization can provide guidance and support, it is in the interests of every business to adopt the necessary practices as an integral part of its own activities.

While the value of a structured approach has been conclusively demonstrated (in terms of defect prevention, failure statistics, productivity, and profitability, for instance), anecdotal evidence also suggests that the following factors may obscure its benefits:

- inadequate preparation
- unrealistic expectations
- undue haste
- use of a 'canned' one-size-fits-all strategy

When these potential problems are addressed early, the structured approach leads to well-integrated processes which consistently deliver high-quality validatable systems.

Conclusions

The systems engineering approach defined in existing standards provides a complete and consistent basis for creating the necessary framework to maximize the benefits of compli-

ance and quality assurance efforts. It ensures not only that "the system gets built right" but also that "the right system gets built."

Even though improvements in the quality of software-based systems have been driven primarily by developments in industries with a more visible risk profile than pharmaceuticals - including defense, aerospace, banking and finance - a general consensus has developed that system quality in any industry can only be assured by implementing a framework and processes to support that objective. [Readers wishing to view case studies and reports can find extensive reading lists related to this and related industry trends at the Web sites of the Software Productivity Consortium and the Software Engineering Institute[19, 20]].

More importantly for the pharmaceutical industry, there also is a developing consensus that the associated investment is both necessary and worthwhile. The only barrier to full participation in the benefits of an improved systems engineering approach is a commitment to the process, followed by the selection of a pragmatic implementation strategy based on the capabilities of an organization and the nature of the systems it depends on.

References

1. Institute of Electrical and Electronics Engineers, IEEE Std 1074-1997, "IEEE Standard for Developing Software Life Cycle Processes," December 1997.
2. International Standards Organization, ISO/IEC 12207, "Standard for Information Technology - Software Life Cycle Processes," March 1998.
3. U.S. Food and Drug Administration CDRH/CBER, "General Principles of Software Validation; Final Guidance for Industry and FDA Staff," January 2002.
4. Institute of Electrical and Electronics Engineers, IEEE Std. 610.12-1990, "IEEE Standard Glossary," September 1990.
5. U.S. Food and Drug Administration, "Compliance Program Guidance Manual on Drug Manufacturing Inspections," Publication no. 7356.002, February 2002.
6. Paulk, Mark C., Editor, The Capability Maturity Model, Guidelines for Improving the Software Process, SEI, Addison Wesley, 1995.
7. Royce, Winston, W., Managing the Development of Large Software Systems: Concepts and techniques. In WESCON Technical Papers, 1970. Reprinted in Proceedings of the Ninth International Conference on Software Engineering, 1987, pp. 328-338.
8. CMU/SEI, A Systems Engineering Capability Maturity Model, Version 1.1, SECMM-95-01, CMU/SEI-95-MM-003, November 1995.
9. CMU/SEI/Raytheon Systems Company, "Hughes Aircraft's Widespread Deployment of a Continuously Improving Software Process", Technical Report CMU/SEI-98-TR-006, May 1998.
10. CMU/SEI, "Benefits of CMM-Based Software Process Improvement: Initial Results," Technical Report CMU/SEI-94-TR-013, August 1994.
11. Belz, F.C., "Applying the Spiral Model: Observations on Developing System Software in Ada," Proceedings of the 1986 Annual Conference on Ada Technology, Atlanta, GA, 1986, pp. 57 - 66.
12. B.W. Boehm, "A Spiral Model of Development and Enhancement," Software Engineering Notes, Vol. 11, No. 4, August, 1986.
13. Boehm, B.W., Gray, T.L., and Seewaldt, T., "Prototyping Versus Specifying: A Multiproject Experiment," IEEE Trans. on Software Engineering, Vol. SE-10, No. 3, May 1984, pp. 290-302.
14. Jensen, R.W., Tonies, C.C., Software Engineering, Prentice-Hall, 1979.
15. Benington, H.D., "Production of Large Computer Programs," Proceedings of the ONR Symposium on Advanced Program Methods for Digital Computers, June 1956, pp. 15 - 27. Also available in the IEEE Annals of the History of Computing, October 1983, pp. 350-361.
16. Lockheed Martin Corp. press release, "Lockheed Martin Management and Data Systems Earns World Class Recognition for Systems Engineering Capabilities," King of Prussia, PA, November, 2000, http://www.lockheedmartin.com/news/articles/111500_1.html.
17. Lockheed Martin Corp. press release, "Lockheed Martin Canada Achieves Carnegie Mellon Software Engineering Institute SW - CMM® Level 3 Rating," Kanata, Ontario, May 2002, http://www.lockheedmartin.com/canada/news/pr02j2.htm.
18. Raytheon press release, El Segundo CA, Feb. 8, 2001, "Raytheon Company joins Software Productivity Consortium," http://www.raytheon.com/newsroom/briefs/spc.html.
19. Software Productivity Consortium, Web Site Reading List, http://www.software.org/pub/externalpapers/.
20. Software Engineering Institute, Publications Web Site, http://www.sei.cmu.edu/publications/publications.html.

About the Author

Sameh Uzzaman is a Project Manager at Millennium Validation Services. He received a control systems PhD from Drexel University where he also led an R&D project team at the Laboratory of Applied Machine Intelligence and Robotics, and shared a patent for a decision support system. His interest in the practical aspects of developing large-scale systems prompted him to spend five years with an international EPCV services company, during which he designed, redesigned, commissioned, and/or qualified a variety of automation and information systems. In addition to his project responsibilities, he is currently involved in the development of practices to support the optimal control of validation and verification processes for computerized systems.

Millennium Validation Services, 220 Continental Dr., Suite 209, Newark, DE 19713.

This article illustrates the risk analysis guidance discussed in GAMP 4.[5] By applying GAMP's risk analysis method to three generic classes of software systems, this article acts as both an introduction to the method and an illustration of its use.

# Risk Assessment for Use of Automated Systems Supporting Manufacturing Processes
Part 1 - Functional Risk

by the ISPE GAMP Forum

The FDA recently announced a significant new initiative to enhance US regulation of pharmaceutical manufacturing and product quality.[1,2] The initiative is based on the FDA's current Good Manufacturing Practice (cGMP) program and covers veterinary and human drugs, including human biological drug products, such as vaccines. The aim is to enhance the established 'quality systems' approach with risk management. Other regulatory authorities have already embraced science-based risk management as a key operating principle.[3,4] With this in mind, this article endeavors to develop a common understanding of the relative risks posed by different types of automated system used to support manufacturing processes. An underlying assumption is that the rigor of validation for an automated system should be commensurate to risk. The significance of any compliance deficiency then

Figure 1. Use of automated systems.
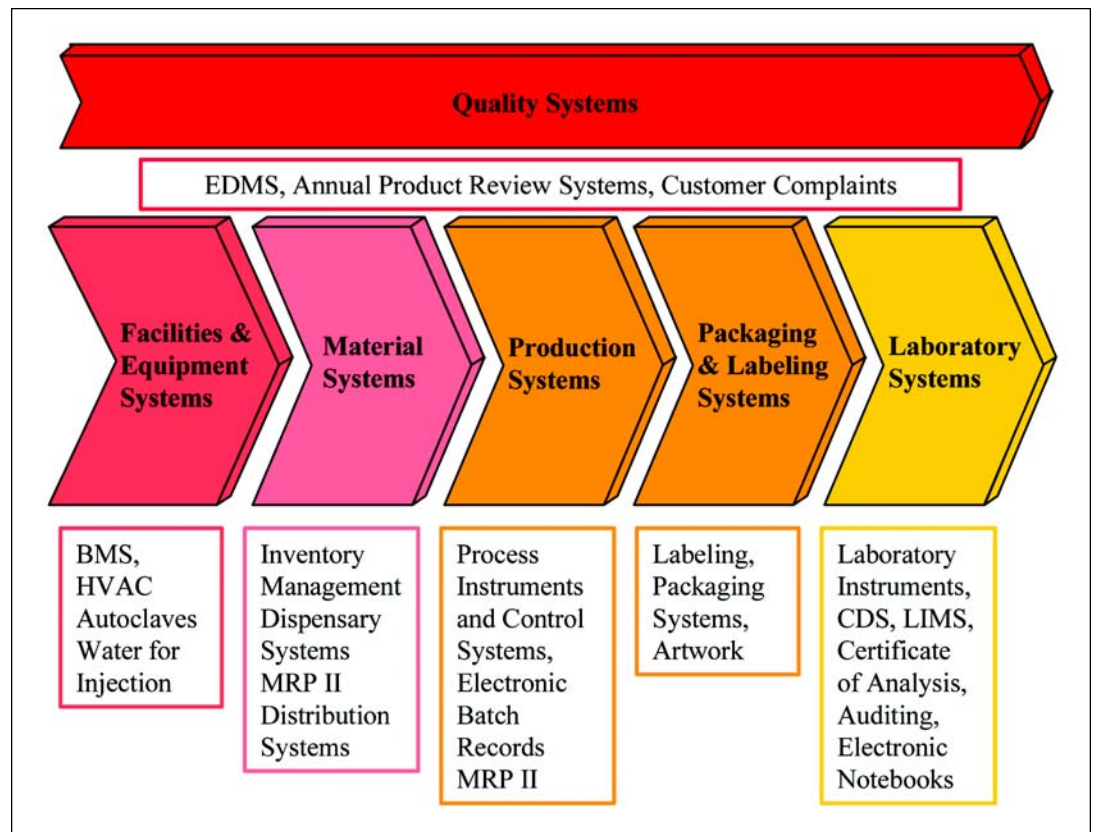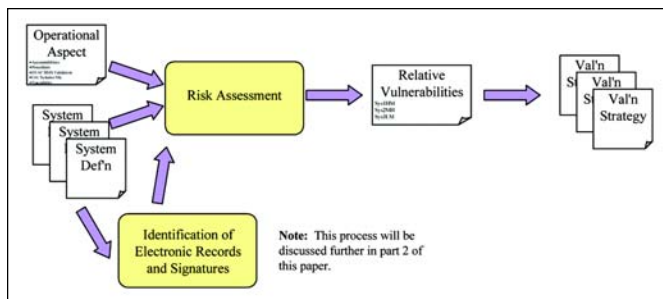
Risk Assessment

Figure 2. Risk assessment process.

needs to take account of the use of that system in supporting a manufacturing process.

This analysis of relative risks is split into two parts:

- The first part concentrates on functional risks associated with different classes of software solution.

- The second part, to be published later this year, will address the relative risks associated with electronic records.

This article illustrates the risk analysis guidance discussed in GAMP 4.[5] By applying GAMP's risk analysis method to three generic classes of software systems, this article acts as both an introduction to the method and an illustration of its use.

Use of Automated Systems

Automated systems are widely used in support of pharmaceutical manufacturing. A workflow analysis of the manufacturing process (based on the FDA's Systems Approach to inspection[6]) identifies six main operational aspects where computer systems are used:

- Quality Systems - dealing with roles and procedural controls
- Facilities and Equipment Systems - dealing with the physical environment used in the production of drug products
- Materials Systems - dealing with drug product components, inventory control processes, and drug storage
- Production Systems - dealing with manufacturing controls
- Packaging and Labeling Systems - dealing with packaging and labeling
- Laboratory Systems - dealing with analytical testing

Figure 1 illustrates where various automated systems might be used. It is important to appreciate that some automated systems support multiple aspects of the manufacturing process such as MRP II systems, while other automated systems are dedicated to specific aspects of the process such as HPLC systems.

Risk Assessment Process

1. The first step of the risk assessment process used here uses the six operational aspects of the manufacturing process to identify the functional criticality of an automated system.

2. The second step is an analysis of the automated system's vulnerability to deficient operation.

3. The third step is the determination of a validation strategy. Differing levels of system vulnerability require different levels of rigor of validation activity.

Equally, validation must address any electronic record/signature requirements. The three-step risk assessment process is illustrated in Figure 2.

Functional Criticality

Determining which operational aspects of the manufacturing process that are most critical requires an understanding of the potential impact that these aspects have on drug product safety, quality, and efficacy. The Canadian Health Products and Food Branch Inspectorate have already identified a number of high risk issues that are likely to result in non-compliant drug product and present an immediate or latent public health risk.[4] These high-risk issues are applied here to automated systems and aligned to the six operational areas identified previously.

Quality Systems
- Document Management
- SOP Administration
- Security Access Controls (e.g., User Profiles and Password Management)
- Change Control Records
- Customer Complaints
- Adverse Event Reporting
- Review/Audit/Corrective Actions Management
- Training Records

Facilities and Equipment Systems
- HVAC Controls and Alarm Handling
- Critical Equipment and Instrumentation (Calibration and Maintenance)
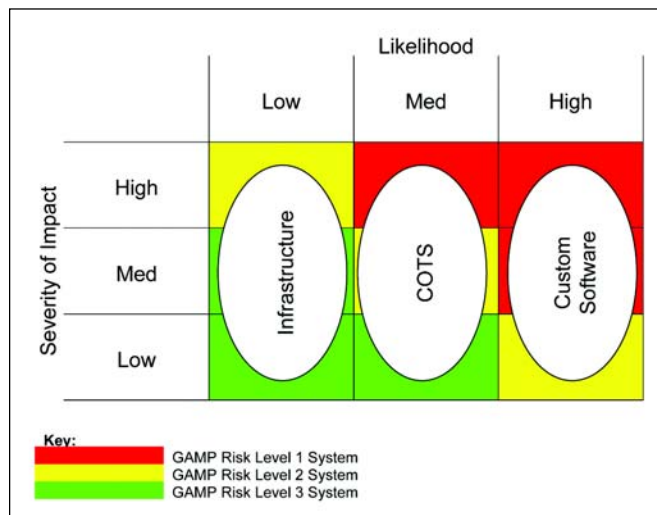- Change Control Records
- Validation Records

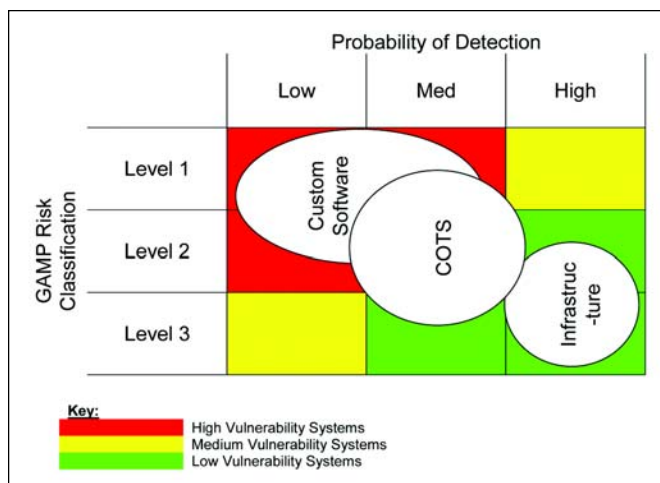
Figure 3. GAMP risk classifications.

Figure 4. Relative system vulnerabilities.

Materials Systems
- Traceability of Material Handling
- Raw Material Inspection/Testing/Status Management
- Storage conditions
- Containers Usage and Cleaning Management
- Distribution Records and Recall Management

Production Systems
- Recipe/Formulation Management
- Batch Manufacturing Instruction and Records
- In-Process Testing
- Yield Calculation
- Purified Water
- Aseptic Filling

Packaging and Labeling Systems
- Labeling Information

Laboratory Systems
- QC Raw Data
- Stability Testing
- Sterility Testing
- QC Analytical Results
- Quality Disposition
- Out of Specification Investigations

The rigor of validation for automated systems supporting these critical operational aspects of the manufacturing process should take account of their composite custom (bespoke) software, commercial Off-The-Shelf (COTS) software, and supporting computer network infrastructure.

System Vulnerability
GAMP's Risk Assessment methodology[5] is used here to analyze the relative vulnerabilities of three typical classes of software system:

- *Custom Software* refers to a software solution that has been specifically developed for application within a pharmaceutical manufacturing set of requirements (see

GAMP 4 glossary of terms). It reflects GAMP Software Category 5 - 'Custom (bespoke) Software' or the application specific configuration code of a GAMP Software Category 4 - 'Configurable Software Packages' system.

- *Commercial Off-The-Shelf Software (COTS)* refers to existing (i.e., not developed specifically for an application) standard software products used across many applications within the pharmaceutical operations and potentially other industries. It reflects GAMP Software Category 3 - 'Standard Software Packages' or GAMP Software Category 1 - 'Operating Systems' or the standard product component of a GAMP Software Category 4 - 'Configurable Software Packages' system.

- *Infrastructure* refers to the typical infrastructure consisting of physical network components, switches, hubs, routers, servers, firewalls, network operating systems, and their configuration.

Initially, the three classes of automated system are analyzed, based on how significant a threat arising from the system might be, both in terms of system function, and system data - *Figure 3*. With all three classes of system, the severity of impact that may arise from the system will depend on its application (i.e., number of critical operational aspects of the manufacturing process the system supports, what breadth of business operations it impacts, and to what extent the system might fail). Each class of system may therefore represent a threat with low, medium, or high severity. However, the likelihood of failure will vary with class of system.

Custom Software
These systems have been developed specifically for this application. This application will, therefore, be the first use of the software so it will not have been proven through an installed base. This class of system will, therefore, tend to present a relatively high *Likelihood* of failure. Applying a high *Likelihood* to the GAMP Risk Classification grid therefore classifies *Custom Software* as predominantly a Level 1 or Level 2 risk.

COTS
These systems typically have an existing significant installed base. The software will, therefore, be in part proven by previous validation exercises and by use. However, the likelihood of failure is not insignificant, as these are often highly complex systems that are highly configurable so that parts of the code might be unproven. This class of system will, therefore, tend to present a medium *Likelihood* of failure. Applying a medium *Likelihood* classifies *COTS* as a Level 1, Level 2, or Level 3 risk.

Infrastructure
The infrastructure is typically built from industry standard network components. These components are proven across all industries as highly robust and also self-correcting (e.g.,
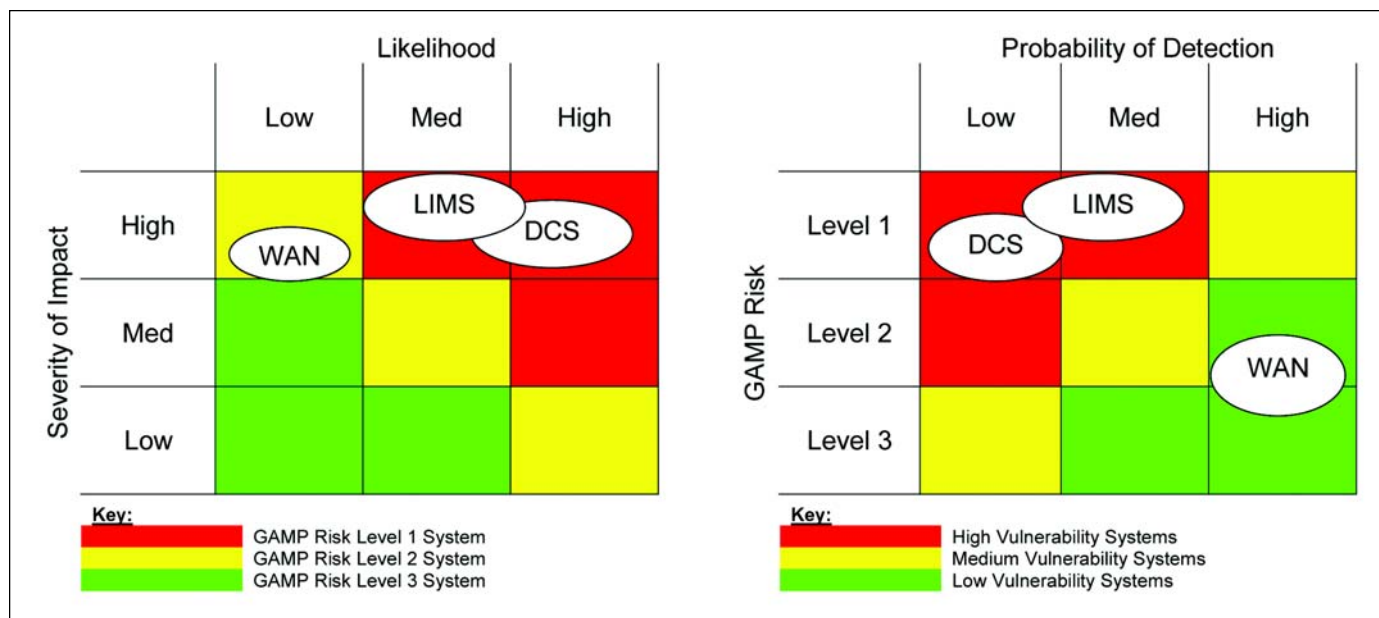
Figure 5. Illustrative systems' vulnerabilities.

TCP/IP protocol). Component failure can often be tolerated without significant impact on infrastructure function or performance. This class of system will, therefore, tend to present a relatively low *Likelihood* of failure. Applying a low *Likelihood* classifies *Infrastructure* as a Level 2, or predominantly, a Level 3 risk.

The relative vulnerability of a system is then deduced by comparing the system's risk classification (Level 1, 2, or 3) with the probability of detecting failure arising from the system - Figure 4. The *Probability of Detection* of failures arising from a system depends on a number of factors, such as:

• error detection function built-into the software function itself

• use of separate and independent systems to duplicate certain functions (redundancy) or monitor the output of the system and report deviations

• use of manual inspections or testing to monitor the correct behavior of the system

Clearly, these last two items will depend on the application, rather than the class of software. However, these different classes of software do tend to have different levels of error detection capabilities:

Custom Software
Error detection is often fairly complex and expensive to develop. It is, therefore, relatively unlikely that a *Custom Software* solution will have good error detection support. These systems will, therefore, tend to have low or medium *Probability of Detection*, yielding a system with a predominantly high vulnerability.

COTS
As COTS have a larger installed base; and therefore, a larger development budget than a *Custom Software* solution, the probability of a COTS product featuring some form of error detection mechanism is higher than with *Custom Software*. These systems will tend to have a mainly medium *Probability of Detection,* yielding a high, medium, <u>or</u> low system vulnerability.

Infrastructure
Most standard network components now have some form of error detection mechanism (e.g., - collision detection at the ethernet level, datagram checksums on TCP/IP). While the correct function of an infrastructure will be largely undetectable to human eyes, these built-in detection mechanisms make it extremely unlikely that an error will be propagated by the infrastructure without detection by the infrastructure itself. In the event of significant infrastructure failure, the applications that employ the infrastructure typically will either report the fault or completely fail, i.e., crash so the failure cannot go undetected. This yields a low system vulnerability.

Rigor of Validation
Broadly, the three classes of software system from *Infrastructures* to *Custom Software* represent increasing vulnerability for public health from drug safety, quality, and efficacy. With increasing vulnerability goes the demand for greater rigor in system validation. Table A lists these classes of risk with suggestions of appropriate levels of compliance activity required to validate that system.

Illustrative Examples
As an illustration, the severity of risk (GAMP Risk Analysis Method step 1) is considered for three typical systems that between them include aspects of each of the classes of software system discussed above.

Distributed Control System (DCS)

While almost certainly based around a proven software DCS product or suite of products, the engineering of DCS installation that controls batch manufacture of a pharmaceutical API is dominated by the application specific configuration and coding. This 'control application' within the DCS will, therefore, fit into the category of Custom Application.

Laboratory Information System (LIMS)

There are now well-established LIMS products on the market that provide the full breadth of function required for information management in most GMP laboratories. As a large part of a typical installation's required functionality is met by standard function, a LIMS can usually be considered as a GAMP Category 4 solution, i.e., a composite of COTS and application specific configuration.

Company Wide Area Network (WAN)

Almost all multisite organizations have some form of WAN. WANs are clearly infrastructure systems, and may include standard hardware and software components such as domain servers, bridges, routers, and firewalls.

Class of System	Vulnerability/ Validation Rigor	Emphasis of User Validation Activities		
		Plan/Report	**Design Phases**	**Qualification Phases**
Custom Software Application	High	• Validation Plan and Report • Development SOPs • Supplier Audit with closure on significant deficiencies • Project Audit(s) • Periodic Review • Change Control	• URS (business and regulatory needs) • FS (full functionality of the system) • Design down to the level of module specifications • Design Review process • Source Code Review (general coding practices and detailed walk-through of highest risk code) • Traceability Matrix (comprehensive)	• Detailed risk assessment against the operational aspects of the manufacturing process identified in this article • Comprehensive positive functional testing (it does what it should do) • Risk-focused negative functional testing (it does not do what it should not do where the risk assessment identified vulnerability)
COTS Application	Medium	• Validation Plan and Report • Development SOPs • Supplier Audit with compensating actions for significant deficiencies • Periodic Review • Change Control	• URS (business and regulatory needs) • FS (full functionality for application specific requirements, points to standard product documentation for standard functions) • Design documents application configuration aspects only • Design Review process • Traceability Matrix (user documents to standard product documents).	• High level risk assessment against the operational aspects of the manufacturing process identified in this article • Positive functional testing of the defined user operation for this specific application (it does what it should do) • Risk-focused negative functional testing (it does not do what it should not do where the risk assessment identified vulnerability)
Infrastructure	Low	• SLA • Quality and Compliance Plan • Work SOPs • Periodic Review • Change Control	• Network topology diagram • Network definition (list of supported applications, network performance and security requirements only) • Design (network configuration)	• High level risk assessment against the operational aspects of the manufacturing process identified in this article • Risk-focused functional testing (e.g., security controls, data integrity, backup and recovery)

Increasing vulnerability, increasing need for rigor in validation

Table A. Summary of vulnerabilities and required validation rigor.

Risk Assessment

Risk Area	High Risk Issues		
	Illustrative DCS	Illustrative LIMS	Illustrative WAN
Quality Systems	-	-	• Security Access Control
Facility and Equipment Systems	-	-	-
Materials Systems	-	• Raw Materials Testing and Status Management	-
Production Systems	• Recipe Formulation and Management • Batch Manufacturing	• In-process testing	-
Packaging and Labeling Systems	-	-	-
Laboratory Systems	-	• QC raw data • QC Analytical results	-

Table B. Illustrative high risk functions for the illustrative systems.

Step 1 - Severity of Risk

The precise role and related risks of DCS, LIMS, and WAN installations will vary from installation to installation. For the purpose of this illustration, Table B suggests some typical functions that each system may provide and can be identified as high-risk issues.

Table B shows that all three of our example systems include high-risk function, and should therefore, be considered high-risk systems. However, this table also helps clarify the severity of the risks relative to each other. LIMS, impacting five different high-risk issues across three of the FDA's inspection systems clearly represents the most severe potential risk to public health.

Steps 2 and 3 - Overall Vulnerability

Assuming that the arguments around the Likelihood and Probability of detection discussed for Custom Software, COTS, and Infrastructure discussed above stand for these three illustrative systems, then application of GAMP's Risk Analysis method steps 2 and 3 will yield relative vulnerabilities as depicted in Figure 5.

The combined steps 1, 2, and 3 of GAMP's functional risk analysis method indicates that both the DCS and the LIMS are high vulnerability systems, and therefore, should be subjected to the full validation rigor proposed in Table A. On the other hand, WAN is a relatively low vulnerability system, and need therefore, only be subjected to validation rigor commensurate with its vulnerability.

Conclusion

This article has applied a functional risk assessment method to the use of automated systems supporting manufacturing processes. It has been shown that functional risk assessment provides a mechanism for assessing and ranking the risks arising from computerized systems. By linking degree of rigor of validation to the overall vulnerability of a system, a process for developing risk-appropriate validation strategies has been demonstrated. High-risk operational aspects of the manufacturing process relative to the use of automated systems have been identified based on previous work by regulatory authorities. The relative risk posed by custom applications, COTS applications and infrastructure also has been analyzed to show the lower vulnerability of infrastructure to erroneous operation impacting drug product quality, efficacy, and safety.

Care must be taken when applying the general risk assessment presented in this article to individual automated systems. It is acknowledged that each system is different. Nevertheless, the general approach is well founded and should help pharmaceutical manufacturers and regulatory authorities alike appreciate the relative rigor of validation appropriate to specific automated systems.

A second part to the article considering the relative risks of electronic records will be published later this year.

References

1. U.S. FDA (2002), Pharmaceutical cGMPs for the 21st Century: A Risk Based Approach, FDA News, 21 August, www.fda.gov.
2. European Union Guide to Directive 91/356/EEC (1991), European Commission Directive Laying Down the Principles of Good Manufacturing Practice for Medicinal Products for Human Use.
3. Trill, A. J., Computerised Systems and GMP - Current Issues, Presentation UK Medicines Control Agency Seminar 'Top 10 GMP Inspection Issues' 24 September 2002.
4. Canadian Health Products and Food Branch Inspectorate (2000), Good Manufacturing Practices - Risk Classification for GMP Observations.
5. ISPE (2001), GAMP Guide for Validation of Automated Systems (known as GAMP 4), International Society for Pharmaceutical Engineering (www.ispe.org).
6. FDA (2002), CPG 7356.002 Drug Manufacturing Inspections: Systems Based Approach.

Acknowledgements

J&J's VP of Worldwide Engineering talks about helping the FDA forge better cooperation with the industry, maintaining and streamlining the drug manufacturing process, and successfully implementing strategic alliances.

PHARMACEUTICAL ENGINEERING Interviews
Ulrich Rudow, Vice President, Worldwide Engineering, Johnson & Johnson

Ulrich Rudow started his career with Philip Morris in Switzerland and Canada as an engineer and later as production and packaging supervisor. In 1971, he joined Schering-Plough in Montreal as manager of engineering and moved on to become plant manager, director, managing director, and vice president of operations and engineering. This career path led to Portugal, Brazil, Ireland, and finally, the US. In 1996, Rudow joined Johnson & Johnson as Vice President, Worldwide Engineering, where he is now responsible for project management, engineering and design, strategic planning, energy and fire policy management, real estate, and the newly initiated engineering leadership development program. These groups serve the global needs of all of the Johnson & Johnson companies worldwide. Rudow has a BS in mechanical engineering from the Engineering College, Hamburg and an MBA from McGill University in Canada. He also attended Harvard and MIT for executive programs.

Johnson & Johnson, with approximately 108,300 employees, is the world's most comprehensive and broadly based manufacturer of healthcare products, as well as a provider of related services for the consumer, pharmaceutical, and medical devices and diagnostics markets. Johnson & Johnson has 198 operating companies in 54 countries around the world, selling products in more than 175 countries.

Q Tell us about yourself and your responsibilities at J&J.

A I now have been with Johnson & Johnson for more than six wonderful years. I am responsible for Worldwide Engineering and Real Estate (WWERE). We get involved with projects more than $3 million and handle all real estate transactions. Right now, we are working on projects with a value of $1.9 billion plus and real estate transactions in 2003 that will exceed $350 million. In addition to our US office, we now have engineering offices in Puerto Rico, California, Shanghai for Asia Pacific, and Belgium for Europe.

WWERE is also responsible for J&J's fire and energy policy, and we have started a new program for high-caliber engineering graduates with great potential—the Engineering Leadership Development Program (ELDP). In the past, our companies hired independent of each other and only for their specific operating company. We now hire for J&J, rotate these young engineers through several of our companies, train them together several weeks each year, and assign a coach for everyone. This makes us much more attractive to high potential engineers. We try to do more than the traditional work of engineers; we want to do our part in promoting excellence in engineering and encourage more engineers to become future leaders. After all, we are a technology company.

Q You serve on a committee that talks regularly with the FDA about fostering better cooperation with the pharmaceutical industry. Could you tell us more about that?

A The committee you are talking about is really a consortium, and I happen to be the Chairman of the Consortium for the Advancement of Manufacturing in Pharmaceuticals (CAMP). It is a group of seven companies that work with academia (at this point MIT and Purdue) to bring new technologies into manufacturing.

We talk regularly to the FDA because changes we see in our industry must have FDA approval, but more importantly, they like what

we are doing and trying to achieve. Process Analytical Technology (PAT), for example, is now very much on FDA's agenda. While the group started out focused on improving manufacturing, we now have substantial involvement by our regulatory managements and R&D. Better products start with R&D!

PAT has not only a potentially great impact on reduced time-to-market and product costs, it will give us a far superior knowledge about our processes on a continuous basis. This is what I call a win-win situation.

Q The pharmaceutical industry is under pressure to innovate its manufacturing operations. Some are saying that we continue to make drugs the same way we did for the past 25 years. Is manufacturing innovation addressed at J&J?

A Yes, I agree. That is one reason for CAMP. We are a very regulated industry and that has led to sometimes not being innovative in the way we manufacture. Change, even for the better, always raises many eyebrows. Working with the FDA will make progress much more efficient; they understand that they are part of the solution to better and more efficient ways to make drugs.

Q We hear alot about the FDA and compliance, but not much is being said in the pharmaceutical industry about streamlining the drug manufacturing operations. To what extent does J&J use "lean-manufacturing" and "six-sigma" methods to reduce manufacturing costs?

A J&J employs what we call "Process Excellence." It does encompass six sigma and lean manufacturing; however, we do not limit the application to manufacturing only. The entire corporation, down to the smallest unit has green-belt or black-belt projects, a dashboard, and documented key business processes. I think that with all the mergers, consolidations, and supply chain management, it is easy to forget that our customers see only a tablet, vial, or other product. The big picture is important, but we cannot forget that we make small medications one by one. Even the smallest improvements are important to us and that is why our strategy is to use "Process Excellence" at every level.

Q Today's popular prescription drugs were developed through traditional biochemistry and are enzyme-based. Opportunities to develop new breakthrough drugs through this route have pretty much been exhausted, while significant new drugs based on the human genome are believed to be seven to 10 years away. What is J&J doing to address this challenge?

A I am not in a position to predict what is going to happen in the next seven to 10 years, but the biotech industry is growing at an accelerated rate right now. What I do know, however, is that meeting the demand for R&D facilities and manufacturing capacity presents a real challenge. We have to make huge capital investments without any payback for five to six years. Do we know what regulations will be in place six

years from now? Is new technology just around the corner; can we predict sales six years hence? We are not used to this kind of risk.

Q These days, any hint of corporate malfeasance or wrong doing by a corporation, and its stock is punished severely by jittery investors on Wall Street. What is J&J doing to continue its excellent standing in the industry?

A Every employee at J&J knows the Credo and is expected to live by it. We continuously emphasize the importance of ethical behavior. It takes proactive management to protect a good reputation. As you know, we are decentralized, which seems to make consistent ethical behavior more difficult. I think that our employees are very proud of our reputation and go the extra length to keep it that way.

Q With J&J's numerous businesses competing for their share of capital, how do you decide which area of the world to produce a product, which campus receives what capital, which project to fund?

A A key goal at J&J is Capital Efficient Profitable Growth. We don't want to grow at any price so we are very prudent with our expenditures. We do put limits on our capital spending and projects have to make their case, but it is not only ROI that counts. We do not compromise on safety. We also make sure we meet all regulatory requirements. These investments might not look like a good payback, but I beg to differ. These are investments that give us the right to do business; they are inherently efficient.

Q Recently, the world has become a dangerous place. Do you give your managers who serve on international assignments any tips regarding their personal safety while traveling on business to one of the world hot spots?

A We have a travel advisory for every place in the world highlighting precautions and dangers. We travel a fair amount of time, but if an employee does not feel comfortable traveling to a specific place, I will accommodate him/her.

Q As an organization with 198 operating companies in 54 countries and capital projects throughout the world, how much does your group get involved with and how is your staff organized to handle the work?

A We get involved in all projects more than $3 million (unless it is just equipment). If it is a small project and engineering is on site, we act as consultants. On large projects, we actively manage with our clients. Since we are a relatively small group working on many projects, we rely heavily on our systems and best practices, and by no means in small measure on A&E firms and constructors.

Q J&J has successfully implemented and maintained Strategic Alliance relationships with designers, construction managers, equipment suppliers and other key vendors associated with your capital projects. What has been the key to J&J's success?

A Because we are a small group, we can't afford to start anew with every project. We must work with teams that know us, our processes, and expectations. Our project delivery process is well defined, accessible to all our partners on our Web site. We also manage the projects with PrimeContract® (a Web system developed by Primavera Systems, Inc). Part of our success is that our partners know what to expect from us. We treat everybody fairly and we just don't have any major issues with any of our projects.

I believe designers, CMs, and contractors know very well how differently all pharmaceutical companies operate and they will adjust accordingly. Some companies use designers like "body shops." They only look at the hourly rate and they employ an army of engineers to check and verify the hours spent. We believe that we get better results if we build in a fun factor for everybody—fun to be creative, fun to do an outstanding job, and fun to be trusted as the expert. In the end, we get better value, including lower overall costs.

Q You and many of your direct reports have been involved with ISPE for quite a long time. What can ISPE do to become a better resource to J&J?

A I think ISPE has done a great job working with the FDA developing industry documents. There is credibility and trust between the two organizations. The training sessions also are very important. Nothing is more valuable than sharing information. I would just like to caution that we should be careful not to be seen as a lobbying group in any way.

ISPE has become so successful there will soon be no conference center large enough for the annual meetings.

FDA's new draft guidance lowers some of the significant hurdles that industry has been struggling with since Part 11's inception in 1997.

FDA's New Draft 21 CFR Part 11 Guidance - What Does It Mean to Industry?

by Arthur D. Perez, PhD

Introduction

Since the introduction of 21 CFR Part 11 to the regulatory landscape in 1997, the life science industries have struggled with questions of how to interpret and implement the requirements of the regulation. Several factors contributed to the lethargy with which industry seemed to embrace the regulation. Not the least of these was the need to address the far more pressing concerns associated with fixing the millennium bug. For the first two years the rule was in effect, virtually the same resources that would be needed to address Part 11 were totally committed to Y2K remediation. Also a factor was the FDA's apparent lack of understanding of the complexity and cost associated with remediating legacy software to achieve compliance with Part 11. In point of fact, such remediations are significantly more difficult and costly than typical Y2K fixes were, upon which the industry was already spending millions. A final piece that didn't fit into the puzzle was the fact that most life science companies had quite logically been moving for several years away from designing and implementing their own software, and few software suppliers had compliant software solutions. To the Agency's credit, they were slow to enforce Part 11 as long as there were not significant other reasons to question the integrity of a firm's electronic records. Many of the warning letters issued during the first five years of the regulation's existence related to problems that could have been cited in relation to validation requirements associated with predicate rules, pertaining to concerns like inadequate security, loss of data due to ineffective or non-existent system management procedures,[1] etc.

Still, the pharmaceutical industry worried about large expenses that were planned for Part 11-driven remediations that offered little or no business benefits beyond compliance with the regulation. Further, considerations around expectation for management of electronic records, especially archives, promised to become much costlier as the records aged. Firms struggled with questions of how to keep a record processible without retaining obsolete hardware and software.

Companies also found themselves making decisions not to implement new technologies with clear business benefit because the perceived complexities and expense of Part 11 compliance seemed to outweigh the benefits. The Agency found this trend especially disturbing in regard to Process Analytical Technology (PAT), which it sees as a major avenue for improvement of process control.[2]

The paradigm shift undertaken by FDA in the new initiative "cGMPs for the 21st Century," in which a shift to a risk/science driven approach to enforcing the Food, Drug, and Cosmetic Act was embraced,[3] seemed an appropriate place to reconsider Part 11. Accordingly, when asked by ISPE how they could help with the new initiative, Dr. Janet Woodcock, Director of the Center for Drug Evaluation and Research, suggested that some ideas on how to make Part 11 more risk and science-based would be appreciated. Within a month, a White Paper was produced by GAMP (ISPE's technical committee that deals most intimately with computer systems compliance). This document was so well received by the Agency that many of its proposals were adopted in the new Part 11 Draft Guidance issued two months later.

FDA's Draft Guidance - Update

Old Guidance Withdrawn, Old Standards Suspended

With the issuance of the draft guidance, FDA announced they are "embarking on a re-examination of Part 11 as it applies to all FDA regulated products."[4] While they may revise the regulation, that process can be expected to take a very long time. Much of the current interpretation of the regulation comes from the Agency-issued guidance documents. The FDA therefore formally withdrew all existing guidance documents, including the Compliance Policy Guide referenced by investigators. However, in addition, Part 11 has been especially susceptible to "regulation by podium," wherein a variety of pundits, including Agency personnel, exercise a large influence through conference presentations or in responses to questions at such conferences and seminars or by e-mail. In a general information vacuum, many firms tend to take such statements and opinions more heavily into account than FDA leadership would like when developing their corporate strategies. As a result, the guidance document includes an acknowledgement that "some statements by Agency staff may have been misunderstood as statements of official Agency policy."[5] Finally, FDA announced the intent "to exercise enforcement discretion with respect to certain Part 11 requirements,"[6] implying that if there were a combination of low risk and other adequate measures to protect data integrity, the Agency does not intend to hold firms to the ultimate letter of the law.

However, individuals must not make the mistake of assuming that Part 11 is dead. To the contrary, the regulation still stands, and despite the narrowing of scope (see below) it still covers a large proportion of the records managed electronically by life science companies. If an FDA investigator feels that data integrity within a particular GxP-relevant computer system is compromised, Part 11 is still there to back up his observations and possible regulatory action. It is important to realize that "enforcement discretion" means just that: Agency personnel will exercise their judgment in deciding where and how to proceed.

Interpretation of the Guidance - Narrowing the Scope of Part 11

The new guidance document specifically states that FDA intends to interpret the scope of Part 11 in a narrower sense so that fewer records will be required to meet the full requirements of the regulation. This should not be interpreted as a statement that the Agency does not care about data created by or stored in computer systems that are outside of the narrowed scope, but rather that they are willing to accept evidence of data integrity that does not meet *all* of the specified requirements in Part 11. The most obvious narrowing is in relation to systems that were already in existence when the regulation became law on 20 August 1997. The Agency intends to exercise enforcement discretion in inspecting such systems, which should be understood to mean that if there is adequate protection of the records (through measures such as system access control, role-based logical security for critical functions, tested and implemented back-up

and archiving processes, change control, and other good system management practices), then it is unlikely that the firm will be cited for not having audit trails on the system. Of course, defining such legacy systems six years later may not be so simple, as it is likely that the system may have gone through one or more upgrades since 1997.[7] Firms who made conscious decisions not to implement available audit trails during such an upgrade may be at risk, at least until the Agency clarifies expectations around the definition of what comprises a legacy system vis à vis Part 11. However, the Agency is to be commended for recognizing the limited value and high difficulty and expense for retrofitting audit trail functionality to old systems.

Another area where FDA recognized an unintended impact of Part 11 is manifested in the reluctance of some firms to implement new technological solutions that would introduce tangible process and/or record-keeping improvements, but at the cost of substantial Part 11 overhead, especially in relation to long-term electronic records management. This has been addressed in two ways: management of archives (discussed later in this article) and the narrowing of the definition of which records are subject to Part 11. Concerning the latter, the Agency clarified that the manner in which a record is used should define whether a record can exist only on paper (or an alternative medium or format such as microfilm or PDF) as opposed to being managed electronically in full compliance with Part 11. It states that "the *merely incidental* use of computers in those instances would not trigger Part 11,"[8] which clarifies that there is no intent to try to extend coverage by the regulation to draft documents generated using word processing software.

Clarification of the importance of how a record is used also should allow a firm to define *where* e-records are to be managed. For example, if a laboratory instrument accumulates data which is automatically sent to a Laboratory Information Management System (LIMS), and all future access to that data is through the LIMS, then it should be acceptable to have Part 11 controls on the LIMS and not on the instrument.

The regulation specifically states cases wherein Part 11 compliance is expected:

- "Records that are required to be maintained by predicate rules and that are maintained in electronic format *in place of paper format*. On the other hand, records (and any associated signatures) that are not required to be retained by predicate rules, but that are nonetheless maintained in electronic format, are not Part 11 records."[9]

  This means that records not required by a specific FDA regulation should not be interpreted as needing to comply with Part 11, and repudiates the widely held concept that any electronic signature in a system subject to Part 11 needs to comply with Part 11, regardless of whether the signature is required by the any predicate rule.[10] The new interpretation thus focuses on *records* rather than *systems*.

- "Records that are required to be maintained by predicate rules are maintained in electronic format *in addition to paper format*, and *are relied on to perform regulated activities*."[11]

  This reiterates that the key factor is how the records are used. A company cannot declare a paper record to be the "official" copy, and then allow employees to reference, copy, or otherwise use an electronic copy of the record to carry out GxP activities. In practice, this is likely very difficult to control so firms should exercise restraint in trying to substitute paper for electronic records if the regulated business process depends heavily on computer-generated records. Ideally, it should be impossible for users to retrieve an electronic record if they should indeed be using paper.[12] The guidance appropriately recommends that practices substituting paper for electronic records be well documented.

- Any records submitted electronically to the FDA as part of a predicate rule requirement fall within the scope of Part 11.

- Any electronic signature executed as a substitute for a handwritten signature required for the purpose of compliance with a predicate rule falls within the scope of Part 11.

Validation

The new guidance document includes a statement that the Agency intends to exercise enforcement discretion regarding the validation requirements put forth in §11.10(a) of the regulation. Unfortunately, some overzealous system owners may choose to interpret this as relief from the general requirement to validate their systems. This bit of selective reading could not be any more wrong. All expectations to validate computerized systems that carry out GxP activities regulated by a predicate rule remain fully in force. Further, the guidance states "Even if there is no predicate rule requirement to validate a system in a particular instance, it may nonetheless be important to validate the system to ensure the accuracy and reliability of the Part 11 records contained in the system."[13] Frankly, it is difficult to imagine a system collecting or managing Part 11-controlled records that would not be expected to be validated in compliance with some predicate rule.

Audit Trails

"The Agency intends to exercise enforcement discretion regarding the specific Part 11 requirements related to computer-generated, time-stamped audit trails (§ 11.10 (e), (k)(2), and any corresponding requirement in §11.30). Persons must still comply with all applicable predicate rule requirements related to documentation of, for example, date (e.g., § 58.130(e)), time, or sequencing of events. Even if there are no predicate rule requirements to document, for example, date, time, or sequence of events in a particular instance, it may nonetheless be important to have audit trails or other physical, logical, or procedural security measures to ensure the

trustworthiness and reliability of the records."[14] These statements recognize that there may be other methods than an audit trail that will give adequate assurance of data integrity based on the scenario for managing and using the record.

In general, it is probably a wise idea to implement audit trails in situations where data may be changed as part of the normal business process, e.g., in a system used in monitoring a clinical study. In situations where data will not likely be changed, such as temperature/humidity measurements in a building management system, an audit trail has distinctly less value than does security around the recorded data. Any decisions not to implement an audit trail, even in such an obvious case as this, should be supported by a documented risk assessment.

Copies of Records

Under the original description of Agency requirements for provision of e-records during inspections, much was made of the expectation that investigators would have the identical ability to process data off-site as the firm being inspected has on-site. This was a bit unrealistic, as no company was going to provide the FDA with licenses for their major computer applications. However, it should be noted that, in general, investigators seemed satisfied to get their data in a spreadsheet or some other standardized format. This guidance validates this approach to providing electronic records for regulatory inspection. Further, it states that in situations where providing e-records to investigators in a format they find satisfactory is infeasible, provision should be made to "allow inspection, review, and copying of records in a human readable form, on your site, using your hardware and software, following your established procedures and techniques for accessing those records."[15] This implies that working on-site with the investigator on the firm's equipment can satisfy the requirement to allow analysis of the data with the same capabilities as is available to the company. While less convenient for investigators, it is a more reasonable expectation.

Record Retention

Perhaps one of the largest long-range benefits life science companies will accrue from the FDA's rethinking of Part 11 will stem from the new interpretation of the archiving requirement for electronic records. While the predicate rules for record retention are not draconian in themselves, the way in which firms use the data has a major impact. For example, data from a clinical study must be maintained for two years following approval to market a drug.[16] However, firms typically reference the same clinical studies in subsequent applications for new therapeutic effects, combination therapies, etc., which has the effect of extending the retention time to up to, or possibly exceeding, two decades.

The initial interpretation of Part 11 required that a firm retain a processible record (including all metadata) for as long as the record is required to be retained. This presented enormous potential technological ramifications: when a company wanted to retire a system containing such records, it either had to migrate the data to a new system or retain the

old one. Migration can be very complex and expensive, and may not be a perfect solution since functions in the new system, (e.g., calculation algorithms) may not process raw data in a manner identical to the old one. Retention of the old system is never a viable long term solution since once vendor support is no longer available expertise in the software becomes scarcer and more costly. An even bigger complication can arise when the hardware supporting the old application becomes obsolete, and both spare parts and people who know what to do with them disappear from the market. While the Agency avowed that they did not expect firms to maintain a hardware museum, no alternative was suggested other than migration.

The new guidance alleviates these concerns. While reaffirming that "Persons must still comply with all applicable predicate rule requirements for record retention and availability,"[17] the Agency suggests that the "decision on how to maintain records be based on predicate rule requirements and that you base your decision on a justified and documented risk assessment and a determination of the value of the records over time." Within these strictures, "FDA normally does not intend to object if you decide to archive required records in electronic format to nonelectronic media such as microfilm, microfiche, and paper, or to a standard electronic file format, such as PDF."[18] This new flexibility with record retention methodology will allow companies to make the decision on how to retain GxP-relevant records based upon a legitimate assessment of risk vs. business need, without significantly compromising the value of the information. For example, batch records are a form of electronic record that does not generally change once they are approved. A firm may choose to retain such records in a processible form for a period of time to allow trending for such business-related purposes as statistical process optimization, but when it is no longer needed for this purpose, it would be reasonable to retain an archive record as a PDF for the remainder of the record's lifetime. Another example might be the clinical study scenario noted above. Often the important data derived from the clinical records resides in a more portable format like SAS, while the initial records remain in the clinical reporting application. If in the firm's judgment the need to be able to manipulate the raw data from a particular study is negligible after ten years, but they still need to retain it, the firm might choose not to migrate the data to a new database, but rather to write it to tables that will be stored on paper, microfilm, or PDF. Since the SAS application still contains the processible data, there is minimal risk to this approach.

Conclusions

While there are still some open questions regarding the true impact of this guidance document, and the exact direction of FDA's thinking in relation to Part 11 is not crystallized, the life science industry should recognize the significant step taken by the Agency in attempting to rectify problems that were either not recognized when 21 CFR Part 11 was written, or which arose as people in industry and the Agency struggled

to interpret it. FDA should be commended for this, and also for the approach they have taken in consultation with professional organizations such as ISPE/GAMP while developing further guidance.

References

1. For examples, see warning letters on the FDA Web site: http://www.fda.gov/foi/warning.htm
2. See *Process Analytical Technologies (PAT) Initiatives* on the FDA Web site: http://www.fda.gov/cder/OPS/PAT.htm
3. See *Summary Progress Report, Pharmaceutical cGMPs for the 21st Century: A Risk-Based Approach* on the FDA Website: http://www.fda.gov/cder/gmp/21stcenturysummary.htm
4. *Draft Guidance for Industry: Part 11, Electronic Records; Electronic Signatures - Scope and Application,* http://www.fda.gov/cber/gdlns/prt11elect.htm
5. *Ibid.*
6. *Ibid.*
7. Many systems underwent Y2K remediation. It is probably a safe assumption that this would not be considered a system upgrade unless the solution was to implement a new vendor package (or major version upgrade). It is also a reasonable assumption that unavailability of a vendor solution for Part 11 compliance for commercially available software packages when the upgrade was implemented will be part of the equation, although as of this writing, there is no precedent for this hypothesis.
8. *Op. cit. Draft Guidance*
9. *Ibid.*
10. Comment 100 in the preamble to 21 CFR 11 states "The agency intends that this section [11.50] apply to all signed electronic records regardless of whether other regulations require them to be signed." While this was in a discussion regarding manifestation of signatures in output and perhaps intended to be more limited in scope, the ultimate effect was the perceived position that all signatures, whether required by a predicate rule or not, had to be handled the same way in systems subject to Part 11 controls.
11. *Op. cit. Draft Guidance*
12. Retaining the electronic copy for the purpose of "back-up" of the paper copy introduces the complication that if it is ever "recovered," the electronic record will then have been used for a GxP purpose. The best solution is probably "in for a penny, in for a pound." Delete the electronic copy once the paper record becomes the master copy. Since the common understanding of archiving typically includes removing the record from the production system, the conversion should be handled as an archiving event.
13. *Op. cit. Draft Guidance*
14. *Ibid.*
15. *Ibid.*
16. See 21 CFR 312.57(c)
17. *Op. cit. Draft Guidance*
18. *Ibid.*

About the Author

Arthur (Randy) Perez is the Executive Expert of IT Quality Assurance for Novartis Pharmaceuticals Corporation and is responsible for quality management within the framework of Novartis' "Crystal" system development methodology. Perez was trained as an organic chemist at the University of Michigan and first joined Ciba-Geigy's process research department in Summit, NJ. After nine years, Perez moved to chemical manufacturing where he led the effort to validate all chemical processes. When manufacturing in New Jersey was discontinued, he moved to quality assurance where he assumed the newly created role of head of computer validation. Following the merger with Sandoz, Perez was appointed associate director of computer validation and GMP training. In addition, Perez also served as head of Novartis' Global Working Group for Computer Validation, which led to the creation of their global CSV policies. Perez is a member of the GAMP Americas Steering Committee, GAMP Council, and GAMP Industry Board.

Novartis Pharmaceuticals Corp., 1 Health Plaza, East Hanover, NJ 07936.

This article covers the US FDA regulatory requirements for computer infrastructures and provides recommendations on how to comply with these requirements.

Regulatory Requirements for Computer Infrastructures

by Orlando López

Introduction

Computer infrastructure comprise all of the computer systems with their associated hardware, software, and networks used to run the business other than the software applications, including:

- servers housing applications
- servers providing a specific service (e.g. file server, print server, database server)
- Wide Area Network/Local Area Network (WAN/LAN) components (e.g. networking equipment such as junction devices, bridges, gateways)
- WAN/LAN systems
- miscellaneous equipment (e.g. network cabling, patch panels and cable drops)
- desktop computers
- data/network centers

Computer infrastructures optimize resources, enable sharing of data, allocate resources to the different users, locate intended receivers of messages, handle messages routing through the network, and make available essential data communication services.

This article covers the US FDA regulatory requirements for computer infrastructures and provides recommendations on how to comply with these requirements. A comprehensive description of computer networks[1] and LAN security[2] qualifications are not covered in this article.

This article covers the infrastructure enabling the transmission of regulated data and access to computing resources across networks. The same principles can be applicable to the infrastructure not transmitting regulated data.

Infrastructure Basics
Servers

A server is a computer or device on a network that manages network resources. For example, a *file server* is a computer and storage device dedicated to storing files. Any user on the network can store files on the server. A *print server* is a computer that manages one or more

Table A. cGMPs regulations applicable to computer systems.

US Drugs GMP	Description
211.22	Responsibilities of QC Unit
211.25	Personnel Qualifications
211.42	Design and Construction
211.63	Equipment design, size, and location
211.67	Cleaning and Maintenance
211.68	Maintenance and Calibration
211.68	Written Procedures
211.68(b)	Record Controls
211.68(b)	Validation of computer systems (implicit requirement)
211.100	Written Procedures, Deviations
211.101(d)	Double Check on Computer
211.105(b)	Equipment identification
211.180	General (Records and Reports)
211.180(a)	Records retention
211.180(c)	Storage and record access
211.180(d)	Records medium
211.182	Use of log(s)
211.188(a)	Reproduction accuracy
211.188(b)	Documentation and operational checks
211.189(e)	Records review
211.192	QC record review
211.220(a)[9]	Validation of computer systems (explicit requirement)

Computer Infrastructures

printers, and a *network server* is a computer that manages network traffic. A database *server* is a computer system that processes database queries. Servers are often dedicated, meaning that they perform no other tasks besides their server tasks. However, on multiprocessing operating systems, a single computer can execute several programs at once. A server in this case could refer to the program that is managing resources rather than the entire computer.

Networks
A network is a system consisting of transmission channels and supporting hardware and software that connects several remotely located computers via telecommunications. Typical hardware components that can be found in a network system are servers, WAN/LAN components, and other equipment. System-level software and firmware completes the infrastructures. Data/network centers host all components, provide appropriate environmental conditions, and provide the appropriate utilities in support of the correct operation.

Local Area Networks (LANs)
A network is any collection of independent computers that communicate with one another over a shared network medium. LANs are networks usually confined to a geographic area, such as a single building or a college campus. LANs can be small, linking as few as three computers, but often link hundreds of computers used by thousands of people. The development of standard networking protocols and media has resulted in worldwide proliferation of LANs throughout business and educational organizations.

Wide Area Networks (WANs)
The WAN can be a routed network supporting multiple network protocols. The network may employ both leased lines and shared commercial telecommunication services, such as frame relay, to provide internetworking service to client companies. WAN can be as simple as a modem and remote access server for employees to dial into, or it can be as complex as hundreds of branch offices globally linked using special routing protocols and filters to minimize the expense of sending data sent over vast distances.

Internet
The Internet is a system of linked networks that are worldwide in scope and facilitate data communication services such as remote login, file transfer, electronic mail, the World Wide Web, and newsgroups.

With the meteoric rise in demand for connectivity, the Internet has become a communications highway for millions of users and LANs.

Corporation's proprietary and confidential information can be sent over public networks, such as the Internet. If this is the case, technological controls can be implemented to ensure that data travels safely, unseen, unchanged, uncopied, and intact.

Three of the areas for conducting business communications over the Internet are:

- Intranets - (B2A = to/from Corporate Operating Companies)

- Extranets - (B2B = to/from Corporate business partners)

- Remote Access by both Corporate employees (B2E) and business partners

Intranet
With the advancements made in browser-based software for the Internet, many private organizations are implementing intranets. An intranet is a private network utilizing Internet-type tools, but available only within that organization. For large organizations, an intranet provides an easy access mode to corporate information for employees.

Regulatory Requirements
The FDA derives from the FD&C Act[3] the authority to regulate the use of computer systems performing functions in drug and device manufacturing. Similar sections in the FD&C Act govern the use of computer systems in other regulated operations, e.g., food,[4] blood.

The FDA compliance policy guideline[5] (CPG) 7132a.11 confirms that when a computer system is performing a function covered by the cGMP regulations, hardware[6] will be viewed as equipment.[7]

Table A lists a few sections in the cGMP regulations applicable to computer systems performing manufacturing-related regulated functions. Equivalent sections can be found in the other FDA regulations.

The main cGMP regulations impacting the computer infrastructures are:

- 21 CFR 211.25 establishes that the personnel involved in the installation, maintenance, and management of the computer infrastructures must have the training and experience to perform the assigned functions.

- 21 CFRs 211.42 and .63 establish the suitability of the design, construction, and performance of the computer infrastructures.
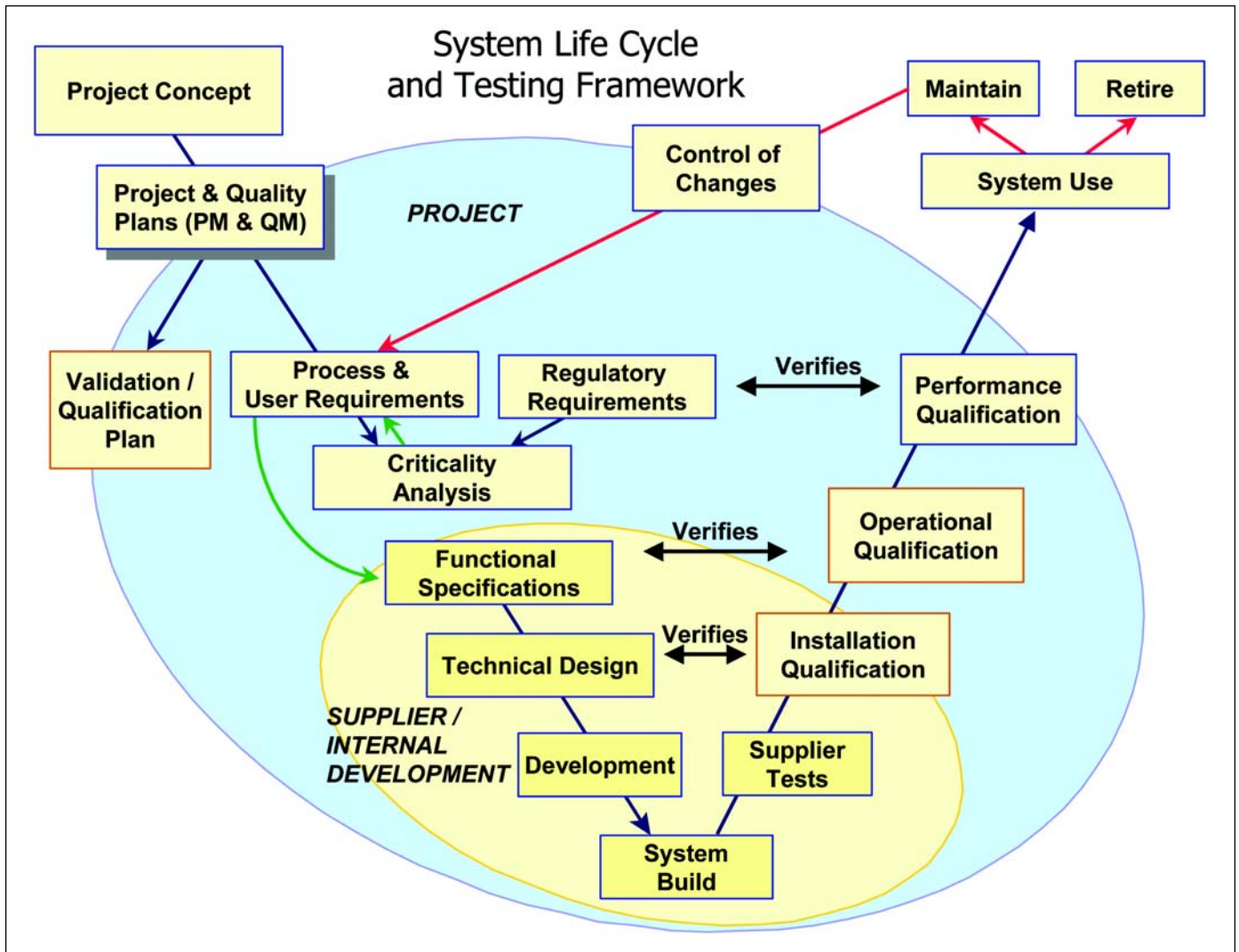
Figure 1. SLC and testing "V" framework.

- 21 CFR 211.68 establishes that there must be documented verification of the inputs and outputs (I/Os) for accuracy and that computer infrastructures must be qualified.[8]

Supporting 21 CFR 211.68 is CPG 7132a.07, which focuses on the need to qualify computer infrastructures and to assure that the data going in and data coming out of the network system are accurate.

Computer systems validation, as introduced in 21 CFR Part 11.10(a) and further defined in the recent draft FDA guideline,[10] is one of the most important requirements applicable to computer systems performing regulated operations. Computer systems validation is the confirmation by examination and the provision of objective evidence that computer system specifications conform to the users' needs and intended uses, and that all requirements can be consistently fulfilled. It involves establishing that the computer system conforms to the user, regulatory, corporate, safety, and the intended use.

A typical computer system-based validation approach is depicted in Figure 1. The Good Automated Manufacturing Practice (GAMP[11]) guideline identifies the importance of the

links established by the "V" framework between the design documentation[12] and the qualification testing.

A computer system-based validation approach applicable to the computer infrastructures would be inefficient and impractical given the nature of the computer infrastructures. A computer system-based validation approach would result in the computer infrastructure being qualified repeatedly because these may be part of multiple software applications.

By approaching the network systems and network-related hardware as equipment, equipment needs to be qualified once, and when applicable, as part of any modification.

The application software should identify in the requirements definition document the required infrastructure technologies and services supporting the application. The verification and qualification testing to the infrastructure services are performed as part of the qualification.

On the context of infrastructures installation and qualification testing,[13] Figure 2 depicts the relationship between the application and infrastructure.

The qualification of the computer infrastructures is an element of the System Life Cycle (SLC). After the installation of the infrastructure-related technologies and services, the

Requirement	Implementation
The following features must be implemented: • Integrity controls • Message authentication One of the following features must be implemented: • Access controls • Encryption All of the following features must be implemented: • Alarm • Audit trail • Entity authentication • Event reporting	• Access controls • Alarm • Audit trail • Encryption • Entity authentication • Event reporting • Integrity controls • Message authentication

Table B. Communication/network controls.

qualification testing comprises verification of the installation related with the proposed design, and executing the functional testing aligned with the functions performed by the installed technologies/services.

In addition to the computer infrastructures qualification testing, other verification activities include design reviews and traceability analysis.

To clarify the intent of 21 CFR Part 211.68, CPG[14] 7132a.07, "I/O Checking" was published in 1982. According to this CPG, computers' I/Os are to be tested for data accuracy as part of the computer infrastructures qualification, and after the qualification, as part of the ongoing monitoring program.

CPG 7132a.07 is based on the realistic anticipation that computer I/O errors can occur on qualified infrastructures. A hardware component (servers, switches, routers, logic circuits, memory, microprocessors), like mechanical parts, can fail after it has been tested. An on-going monitoring program shall be established and followed to verify hardware I/Os during the operation of the network system.

The level, frequency, and extent of the I/O checking were suggested in the Federal Register of 20 January 1995 (60 FR 4091). The level and frequency of the I/O verifications shall be guided by written procedure and shall be based on the complexity and reliability of the computer system.

The introduction in 1997 of Part 11 provided the formal codification applicable to computer systems performing regulated operations. The network systems enable sharing data that may be part of electronic records regulated by the FDA, and therefore subject to compliance with Part 11. As such, the responsibility of the network for the in-transit records includes:

• Reliability
• Authentication
• Confidentiality
• Integrity
• Usability

The above feature required for quality records can be achieved by following security across the infrastructures. In order to guard against unauthorized access to the records transmitted over a communications link, technical controls are requirements as delineated in Table B.[15]

Key Elements that should be in Place to Ensure FDA Compliance

As stated in the Sidebar 1, the essential elements to ensure FDA compliance to the computer infrastructures are the design and/or selection of equipment, installation of equipment, configuration of data communication equipment, evaluation and testing the equipment. These activities should be documented, verified, and tested with the objective to establish the equipment capability within the integrated environment. Additionally, any changes to the equipment and/or the operational environment should be subject to change control.[16]

During the life cycle of the computer infrastructures and associated components, the FDA expects companies to have adequate documentation to demonstrate that sound procedures are used to establish intended use of the computer infrastructures and associated components. The infrastructure documentation should consist of design documentation, diagrams, equipment configuration worksheets, qualification testing, summary report(s), change control records, maintenance records, and training records.

Based on current regulatory requirements, the following documentation practices are essential to support the computer infrastructures:

1. Define a process for organizing, controlling, and disseminating computer infrastructure qualification documentation to include the following:

   • computer infrastructure design documentation and configuration worksheets

   • computer infrastructure qualification test procedures and protocols

   • computer infrastructure qualification related technical operating procedures

   The diagrams and worksheets are organized in a way that both high level overview and equipment specific information is maintained up-to-date.
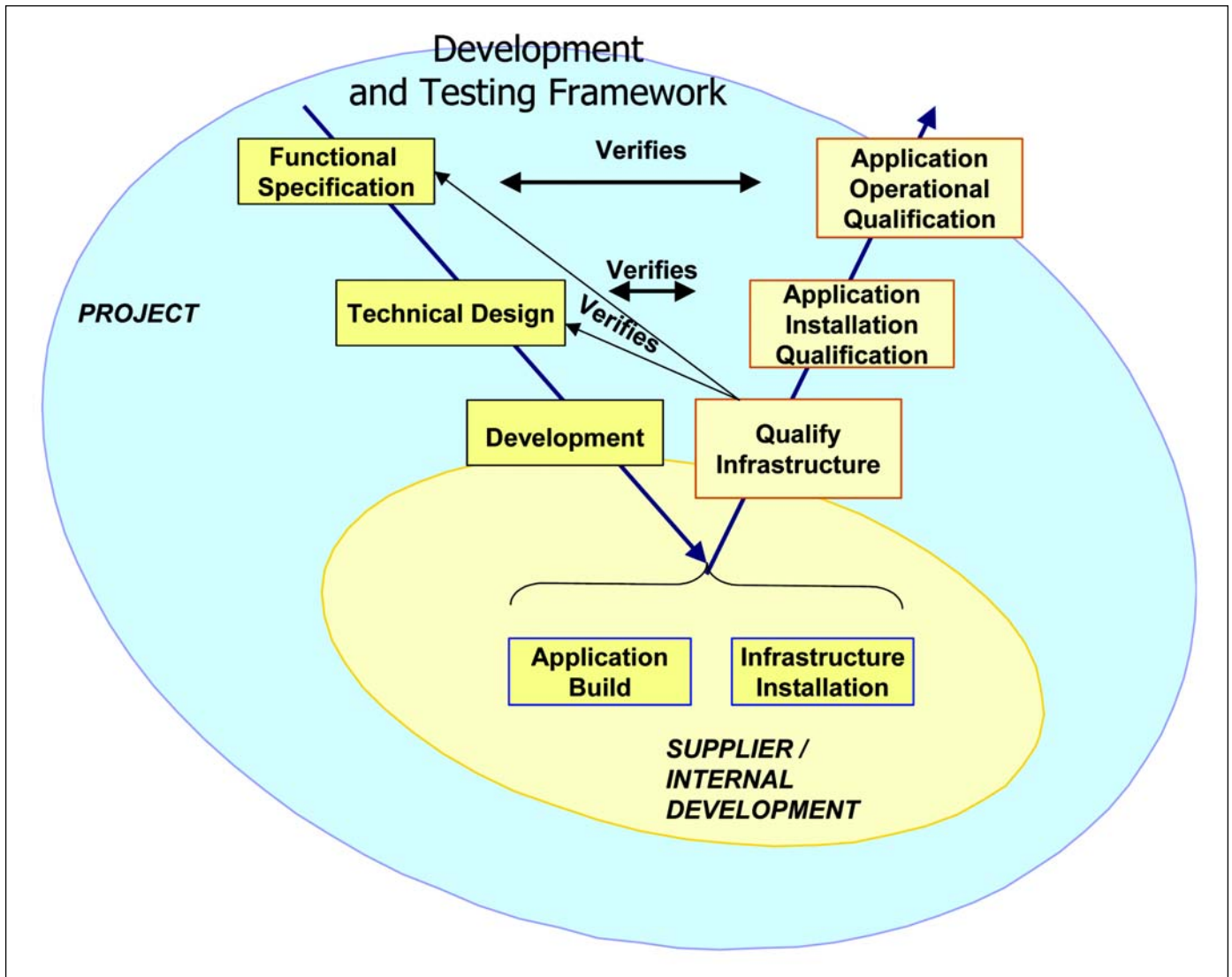
Figure 2. Application/infrastructure development and installation correlation.

2. Written equipment design documentation should describe what the computer infrastructure intended functions are and how they will work. The equipment design documentation to consider are:

• Description of the hardware and communication design in broad terms, including the design and functional requirements to support associated connectivity to networks; system backup and restore time frames, availability window, maintenance window, number of clients (total and concurrent) that will use the system, and database growth rates.

• Creation/maintenance of high-level diagram that identifies the data/network centers and associated diagrams depicting all data communications equipment deployed within the data/network centers.

• Define the characteristics of each section of the infrastructure. The characteristics should be in sufficient detail as regards performance and capacities for use in the procurement of the hardware.

• Equipment Configuration Worksheet: an equipment configuration worksheet shall be completed for each device. The equipment configuration worksheet shall provide the basis for IQ testing.

• List all applicable constraints of the proposed design.

• List all assumptions of the proposed design.

• List all dependencies of the proposed design.

If a redundant and fault tolerant network is required, then the design must include outbound and inbound redundancy, failover, route aggregation, and automatic loop prevention.

3. Equipment configuration worksheets are completed for equipment, including changes from out of the box installations. The equipment configuration worksheet should provide the basis for Installation Qualification (IQ) verification.

4. Qualified staff shall deploy the infrastructure and its implementation shall be performed in accordance with defined processes. The qualified staff should be trained in established procedures through formal procedure training prior to an SOP's effective date. SOP training also should be part of new staff orientation.

5. A written qualification protocol(s) should be developed based on the design specification, which is (are) specific and meaningful in relation to the attribute being tested. The qualification protocols include both structural and functional analysis, as applicable, including checks of I/Os to ensure the accuracy and security of computer inputs, outputs, and data.

6. The IQ protocols shall be maintained for configurable infrastructure hardware, system-level software and associated data communications equipment to be deployed. The objective of each installation qualification is to provide documented evidence that the equipment and system-level software have been installed in accordance with the manufacturer's specifications, and have been properly configured in accordance with relevant design guidelines and the equipment configuration worksheet.

7. The Operational Qualifications (OQ) shall be performed after the installation testing (Refer to Sidebar 2) to provide evidence that the infrastructure hardware and the network provide secure and reliable data communications. Protocols should be developed to verify that session

authentication and data encryption occurs when data is broadcast across a public medium such as the Internet or the wireless spectrum.

The OQ is also performed to demonstrate conformance with functionality to be provided by any service and/or network component.

8. Any deviations and/or discrepancies encountered during the execution of the infrastructure qualification must be investigated and resolved.

9. Addition, replacement, and retirement of an infrastructure component or system-level software, including configuration changes, should be subject to the review and approval process similar to a change control program.

10. Test results and an evaluation of how these results demonstrate that the predetermined design specification has been met (e.g. requirements traceability analysis).

11. Procedures need to be developed to maintain the operating environment and computer infrastructure. An audit trail documenting time-sequenced development and modifications or revisions to the documentation is required as well.

12. During the operation, computer electronic records must be controlled, and this includes record backup, security, and retention.

13. Ongoing monitoring of the network program is established to respond to conditions where the network does not conform to the defined specifications.

Summary

The cGMP regulations provide the regulatory requirements applicable to develop, maintain, and retire computer related infrastructures. Since computer related infrastructure components could be considered as equipment, the validation system-based approach is unreasonable and very expensive. Computer infrastructure needs to be qualified once, and when applicable, as part of any addition, replacement, and retirement.

The application software should establish the required infrastructure technologies and services in the requirements definition document. Based on these requirements, the infrastructure services are implemented, qualified, and maintained.

As in the application software, the key elements to ensure FDA compliance are: design, implementation, testing, maintenance, and retirement practices, and the associated documentation.

References

1. López, O., Qualification of Computer Networks, 1st ed., Storrington, West Sussex, UK Sue Horwood Publishing Limited, 2000. ISBN 1-904282-00-8.

2. López, O., Computer Technologies Security Part 1, 1st ed., Storrington, West Sussex, UK Sue Horwood Publishing Limited, 2002. ISBN 1-904282-17-2.

3. Food, Drug and Cosmetic (FD&C) Act Section 501(a) (2)(B).

4. Food, Drug and Cosmetic (FD&C) Act Section 402 (a) (3).

5. Note that a CPG, as described in 21 CFR 10.85, is considered an Advisory Opinion directed to FDA inspectors. These guidelines are the mechanisms the FDA utilizes to spread policy statements within the Agency and to the public.

6. Physical electronic circuitry and associated equipment.

7. Edmund M. Fry, "FDA Regulation of Computer Systems in Drug Manufacturing," Pharmaceutical Engineering, September/October 1988.

8. Qualification: the process of determining whether a system or component is suitable for operational use (IEEE).

9. 1996 CGMP proposed regulations, FR, Vol. 61, No. 87, May 3, 1996.

10. FDA, draft *"Guidance for Industry: 21 CFR Part 11; Electronic Records; Electronic Signatures Glossary of Terms,"* August 2001.

11. GAMP Suppliers Guide Rev 4, December 2001.

12. In this context, a design documentation may include a set of specifications containing the project requirements, configuration, functional, and/or technical design.

13. Qualification testing: testing conducted to determine whether a system or component is suitable for operational use (IEEE).

14. López, O., *"A Guide to Computer Compliance Guides,"* Journal of cGMP Compliance, January 1997.

15. Health Care Financing Administration, 45 CFR 142, Security and Electronic Signature Standards, Proposed Rule, August 1999.

16. The process of identifying, evaluating, coordinating, implementing, documenting, and obtaining formal approval of changes in a computerized system or database.

About the Author

Orlando López is an Executive Consultant for Networking and Computing Services (NCS), a Johnson & Johnson Company. He manages the development and implementation of the infrastructure qualification strategies in NCS. As part of his contribution to the GAMP 4 Guide, he led the development of the Operation Appendix 06, "Guidelines for Record Retention, Archiving, and Retrieval." He is a member of the Pharmaceutical Engineering Editorial Review Board and the author of the book "21 CFR Part 11 - A Complete Guide to International Compliance," published by Sue Horwood Publishing Limited (www.suehorwoodpubltd.com).

He can be reached at 1-908/685-3685 or by e-mail at olopez6102@msn.com.

Disclaimer

Good Batch Practice

by Tiny Heesen, Rudy Kuijer, and Henk Man

Introduction

The S88 standard is applied in a number of different industries as a design methodology for batch processes. The methodology and standard also can be applied regardless of the level of automation in a factory. It can be used for entirely manual plants, fully automated plants, or the more common semi-automated plants. There have been many successful applications of the standard in fully automated operations as well as partially automated operations in pharmaceutical and consumer product industries.

The basic models of S88 and its recipe types and structure are described only in general terms in this article. Furthermore, S88 is placed in the context of GAMP, ISO 9001, and GMP (the context of the pharmaceutical industry and its suppliers). Finally, this article discusses the benefits of using S88 in the pharmaceutical industry, and demonstrates how these benefits surpass the general benefits by making it easier to comply with the GMP requirements.

GAMP

GAMP is an interpretation of a part of FDA's GMP legislation developed by the GAMP Forum with regard to automated systems. The GAMP Guide lays down a certain validation life cycle that is illustrated in the so-called V-model - *Figure 1*. GAMP specifies the required documents and the responsibilities of users and supplier(s) for each life cycle stage.

To find out more about GAMP, visit www.gamp.org; to find more about the FDA, visit www.fda.gov.

S88: Models and Terminology

The Instrument, Systems, and Automation Society (ISA) was founded in 1945 as the Instrument Society of America and changed its name in 2000. In 1995, ISA published ISA – S88.01, Batch Control, Part 1: Models and Terminology. It is often called S88 for short. S88 is generally considered the international standard for batch control. Therefore, it is frequently applied in designing and automating batch processes.

To find out more about the ISA, visit www.isa.org.

S88 is set up to accomplish two goals:

- definition of standard terminology
  - everyone speaks the same language

- definition of models for realizing batch processes with the following aspects in mind
  - flexibility
  - modularity
  - maintainability

The three basic S88 models are:

1. the process model
2. the physical model
3. the procedural control model

The Process Model

Here, the following question is answered: what should be manufactured? A hierarchical breakdown of the process required to manufacture the end product is outlined in this model.

The process model has four levels (starting at the top): process, process stage, process operation, and process action.

Process

A process is a sequence of chemical, physical, or biological activities for the conversion, transport, or storage of material or energy. Industrial manufacturing processes can generally be classified as continuous, discrete parts manufacturing, or batch.

Process Stage

The process consists of one or more process stages that are organized as an ordered set

which can be serial, parallel, or both. A process stage is part of a process that usually operates independently from other process stages. It usually results in a planned sequence of chemical or physical changes in the material being processed.

Process Operation

Each process stage consists of an ordered set of one or more process operations. Process operations represent major processing activities. A process operation usually results in a chemical or physical change in the material being processed.

Process Action

Each process operation can be subdivided into an ordered set of one or more process actions that carry out the processing required by the process operation.

The Physical Model

Here, the following question is answered: with what should it be manufactured? A hierarchical breakdown of the equipment required to produce a batch is outlined in this model. In this case, the word 'equipment' refers not only to the valves and motors, but also to the reactors and raw material vessels that are involved in the batch process.

The physical model has seven levels (starting at the top): enterprise, site, area, process cell, unit, equipment module, and control module. The top three levels (enterprise, site, and area) are frequently defined by business considerations and are not modeled further in S88. The three higher levels are part of the model to properly identify the relationship of the lower level equipment to the manufacturing enterprise. The four lower levels (process cell, unit, equipment module, and control module) are defined by engineering activities.

Process Cell

A process cell contains all of the units, equipment modules, and control modules required to make one or more batches. An example of a process cell is a facility that is used for the production of an ingredient for a pharmaceutical product. In this facility, more than one unit is required to manufacture the end product. There are units for fermentation, enzymatic conversion, purification, and drying.

Unit

A unit is made up of equipment modules and control modules. One or more major processing activities can be conducted in a unit. It is usually centered on a major piece of processing equipment, such as a mixing tank or reactor. Units operate relatively independent of each other. This standard presumes that the unit does not operate on more than one batch at the same time.

Equipment Module

Physically, the equipment module may be made up of control modules and subordinate equipment modules. An equipment module can carry out a finite number of specific minor processing activities, such as dosing and weighing. An example of an equipment module is a collection of devices that control the reactor temperature by either letting in hot or cold media in the jacket of a reactor.

Control Module

A control module is a typical collection of sensors, actuators, other control modules, and other associated processing equipment that, from the point of view of control, is operated as a single entity. An example of a control module is a state-oriented device that consists of an on/off automatic block valve with position feedback switches that is operated via the set point of the device.

The Procedural Control Model

Here, the following question is answered: how should it be manufactured? A hierarchical breakdown of the order, and the way the required equipment is used to produce a batch is outlined in this model.

The procedural control model has four levels (starting at the top): procedure, unit procedure, operation, and phase.

Procedure

The procedure is the highest level in the hierarchy and defines the strategy for carrying out a major processing action such as making a batch. It is defined in terms of an ordered set of unit procedures. An example of a procedure is the procedure to make plastic. There are several unit procedures needed to manufacture the end product. For instance, there are unit procedures for polymerization, recovery, and drying.

Unit Procedure

A unit procedure consists of an ordered set of operations that causes a contiguous production sequence to take place within a unit. Only one operation is presumed to be active in a unit at any time. An example of a unit procedure is a unit procedure for polymerization. In this unit procedure, several operations are needed to carry out the objective of this unit. There are operations for the preparation of the ingredients, charging of the ingredients into the reactor, and the polymerization reaction itself.

Operation

An operation is an ordered set of phases that define a major processing sequence that takes the material being processed from one state to another, usually involving a chemical or physical change. An example of an operation is an operation that carries out a reaction. In this operation, there are several phases. Each phase takes care of a small task like adding the catalyst, heating, and maintaining the correct pressure level.

Phase

The smallest element of procedural control that can accomplish a process-oriented task is a phase.

These three models do not stand alone, but are coupled on several hierarchical levels - *Figure 2*. Although the control module is part of the physical model, it is not directly combined with a phase. Therefore, it does not appear in Figure 2.

In addition to the aforementioned models, S88 defines four types of recipes. Each type has a different purpose in a company, and each type generally is created and maintained
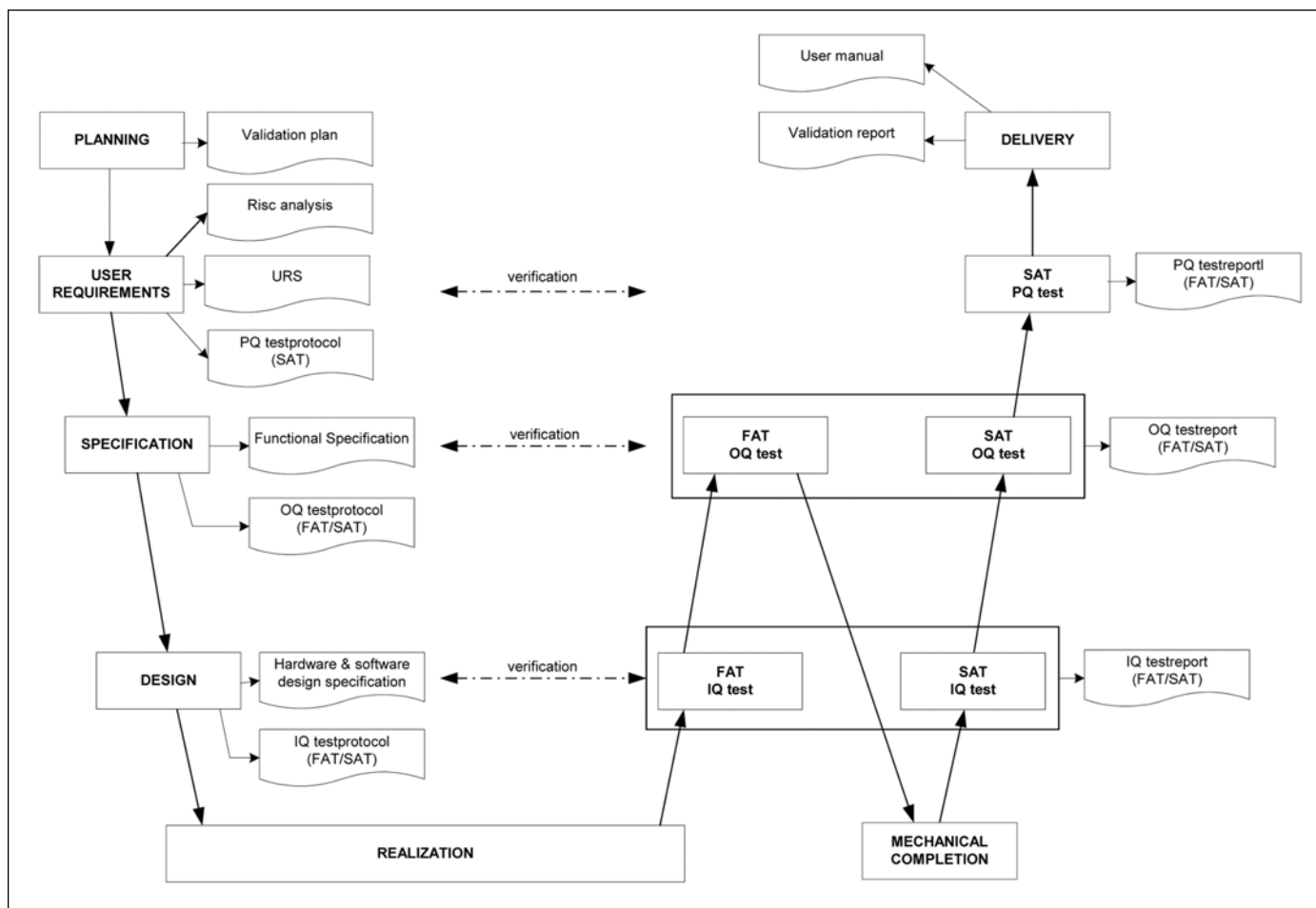
Figure 1. A modified GAMP V-model.

by different people. The types of recipes defined by S88 are: general recipes, site recipes, master recipes, and control recipes.

There is one general recipe for each specific product variation made by a company. It defines, in an equipment independent manner, the material and process dependencies required to make a product. The general recipe is usually created during or after the pilot plant scale-up of an R&D recipe.

There is one site recipe for each site that will make the product, or some portion of the product. A site recipe has the same structure as a general recipe, but may be modified for the local language and unit of measure. It also may be modified to take into account local material availability, or it may only define a part of the general recipe that is actually performed on the site. Site recipes also define the "bill of materials" for production.

Master recipes are the process cell-specific recipes that define exactly how a product is to be made in a specific process cell, based on the units in the cell, material flows between units, and the equipment phases available in the units. There is typically one master recipe per product. Master recipes are usually constructed using a graphical notation format that is similar to sequential function charts. The lowest level of a recipe's procedure is a recipe phase and is a reference to an equipment phase. A recipe phase may reference a specific

equipment phase in a specific piece of equipment, or it may reference a class of equipment phases with the specific equipment element selected at runtime.

A master recipe is a template only. This means that a master recipe itself is not executed. A copy of the master recipe is made for each batch produced, and is called a control recipe. There is one control recipe per batch and it starts as the copy of the master recipe, but it may be modified before it is executed. Some of these modifications include specifying the exact equipment the recipe should run against and specifying values for recipe parameters.

All recipes are made up of five elements: a header, a procedure, a formula, equipment requirements, and other information. The formula contains a definition of the process inputs (materials to be used in making the batch), process outputs (materials generated as a result of making the batch), and process parameters (process or product values which can be specified for making a batch). The recipe procedure defines the procedural logic to be followed to make the product. The master and control recipe procedures have a specific hierarchical organization. The top level of the hierarchy is formed by the unit procedures as defined in the S88 procedural control model. Each recipe is made up of unit procedures that define the contiguous operations that occur within a unit. Unit procedures are made up of operations, and operations are made up of recipe phases.
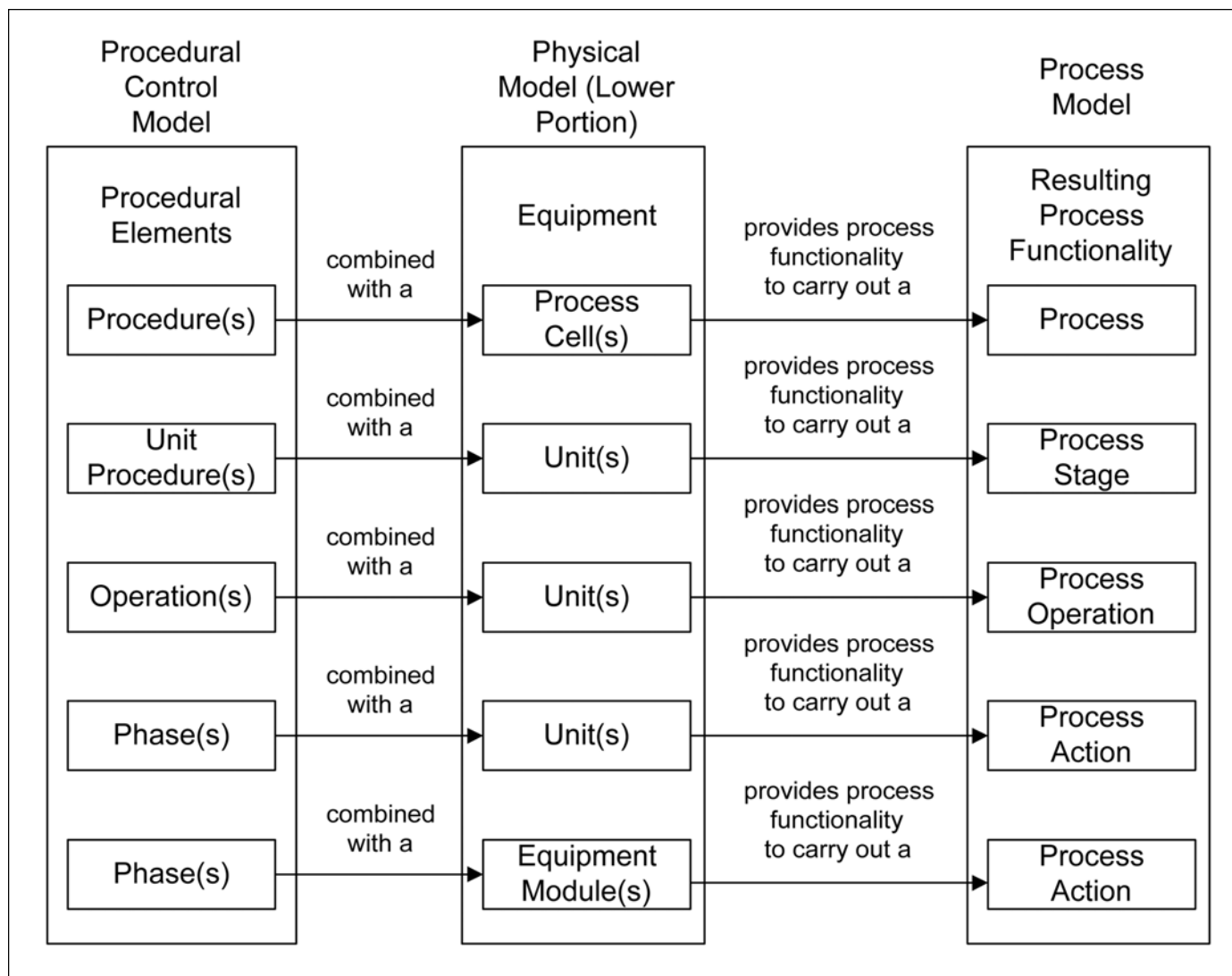
Figure 2. S88 basic models and couplings on various levels.

S88 Related to Quality Systems (ISO 9001)/GMP/GAMP

GAMP and S88 are two items that cannot be compared to each other because they exist on different levels of abstractness. Figure 3 shows both the different levels of abstractness and the relations between GAMP, S88, GMP, and ISO 9001.

S88 is a design methodology for batch processes. GAMP does not lay down the use of S88 as the design methodology; GAMP does not lay down any specific design methodology. GAMP only requires that design methodologies should be so-called 'appropriate methods.' S88 is such an 'appropriate method' for the design of batch processes. The additional benefits of applying S88 in the regulated pharmaceutical industries will be discussed later in this article.

General Benefits of Applying S88 in Batch Control

Realizing a batch application that is in accordance with S88 results in modular batch control software. Modularity is a great asset when testing an application because it makes locating software faults a lot easier. The consequences of

correcting software faults remain confined in one module, thus limiting the chances of introducing new faults in (the rest of) the software as a result of correcting the original fault. In addition to this benefit, the retest effort of a changed or corrected software module is greatly reduced when compared to non-modular software. Imagine retesting the complete batch application instead of retesting just one module of that batch application. When all these benefits are added up, it becomes obvious that the effectiveness of testing modular software is much greater, and this is demonstrated by the relative low numbers of fault occurrences during start-up and operational life of the batch facility.

Because of its inherent flexibility and modularity that a batch facility has when built according to S88, its maintainability is greater and changes can be incorporated with relative ease. This all leads to a lower Total Cost of Ownership (TCO).

Another aspect of the S88 standard that contributes to a lower TCO is that it makes recipe development straightforward enough to be accomplished without the services of a control systems engineer. This ease of recipe development
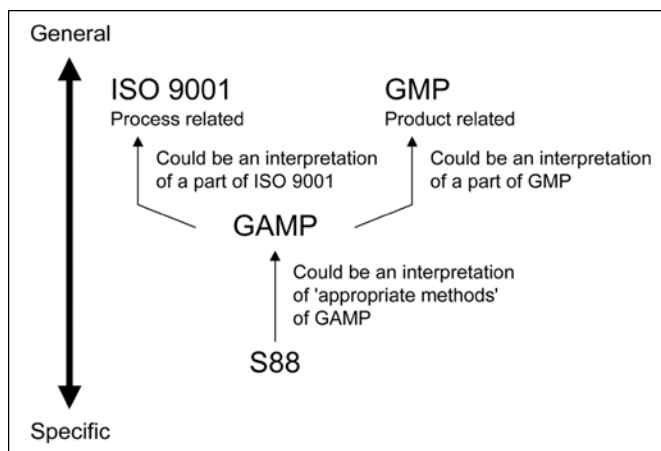
Figure 3. Levels of abstractness of ISO 9001/GMP, GAMP, and S88.

lies in the fact that S88 defines a separation of product information from production equipment capability. This separation allows the same equipment to be used in different ways to make multiple products or to perform different operations on the same product. As mentioned before, recipes in S88 are made up of five elements. Two of those elements (formula and procedure) are important in the simplification of the development of recipes. Changing the formula, while keeping the procedure the same, is an easy way to make different grades of the same product. A recipe for a new product can be made by changing the procedure (and other recipe elements). Since a procedure in S88 consists of phases (at the lowest hierarchical level), this simply means rearranging these phases into different operations and arranging these new operations into new unit procedures. This comes down to reusing recipe phases or even operations when rearranging operations alone is enough to yield the desired new unit procedure. In this way, recipe development can be done at SCADA level without changing the batch control software.

Figure 4 clearly shows the differences in cost, in practice, between a typical automated batch facility that has been built with S88 and without using S88. The cost in Figure 4 relates to the cost development during the design and realization stages and a short period of time after the system (batch facility) has become operational.

The cost in the design and realization stages for batch projects that have been completed with S88 is, in practice, higher than without S88. One of the causes for the higher cost is the lack of experience with S88 or incorrect usage of the S88 terminology. In the beginning, the whole concept of S88 is not easy to comprehend because of its high level of complexity and abstractness. However, in the end, S88 often delivers a lot of added value if one becomes more knowledgeable in this matter. By applying S88 more often (in batch projects) and thus increasing the understanding of S88, the cost due to inexperience will decrease. Another cause of cost increase is the introduction of extra devices (particularly valves) to establish units in such a way that so-called common equipment is eliminated. This means that each device has been clearly assigned to one unit and that other units cannot use it. This results in the increase of the independence of units.

The latter cause contributes a relatively small amount to the increase in cost when compared to the first cause. However, despite the increase in cost during the design and realization stages, the total cost of ownership (i.e. the sum of the cost during the design and realization stages and the operational cost) is much lower. As shown in Figure 4, the surge in cost (due to fault occurrences and implementation of changes necessary to keep the facility running properly) after the system has become operational is non-existent in batch facilities that have been designed and automated in compliance with S88. This is not only attributed to the greater effectiveness of testing of a modular application and lower maintenance cost, but also to higher availability of that facility because maintenance takes up less time. Changes as a result of better (practical) understanding of a process will, of course, always exist, but due to the modular set up (i.e. no 'spaghetti'-like software structures), the cost is manageable.

Another reason for using S88 is that all parties involved in a batch project (this means both users and vendors) should speak the same language. S88 is that language for batch projects. It defines the models and terminology that are relevant for realizing the automation of a batch facility. So, using S88 promotes the unambiguous communication between all parties involved in a batch project.

Within S88, there are some things that are not specified and thus left up to the S88 user. This may lead to different interpretations of S88 in different companies, even when the models and terminology are used correctly. S88, namely, offers some degree of freedom (with 'freedom' is, in this case, not meant the room for ambiguity, but valid choices that are S88 compliant) and does not force the S88 user in a straitjacket. There is not only one way to correctly apply S88. It is in the interest of the user company that the supplier's interpretation of S88 concurs with the user's interpretation of S88, especially when a lot of batch facilities in this company have been designed and built in compliance with S88. In that case,
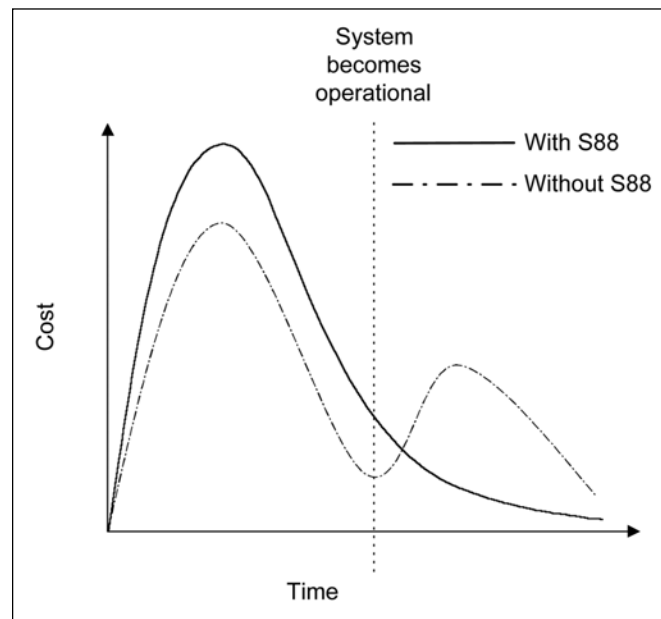

Figure 4. Cost development with and without S88.

maintenance personnel are often very well acquainted with the user's interpretation of S88. Maintaining a new, S88 compliant, batch facility then takes little extra effort unless it is built according to the user's interpretation of S88.

Additional Benefits of Applying S88 in the Regulated Pharmaceutical Industries

GAMP and S88 are two items that, at first sight, do not have a lot to do with each other. However, under certain circumstances, an association between these two items may very well exist. The advantages of applying S88 in batch projects also benefit the qualification stages of GAMP.

In a pharmaceutical (batch) project, a number of parties may be involved. These parties could be the user (the pharmaceutical company), an engineering contractor, and a system integrator as subcontractor of the engineering contractor. GAMP does not lay down S88 as the design methodology. In fact, GAMP does not lay down any specific design methodology. However, a user can demand from the engineering contractor and the system integrator that they work in an S88 compliant manner. If the engineering contractor and system integrator lack the required S88 knowledge, the user may opt to educate them in this matter. The advantage in this case is that the engineering contractor's and system integrator's interpretation of S88 is the same as the user's interpretation.

As mentioned before, S88 is a design methodology. For the GAMP V-model (Figure 1), this would mean that S88 is only important to the specification and design stages. This perception is quite limited because application of S88 results in a facility that has been set up in a modular fashion (both process equipment and control software). Modularity is also of great importance during the qualification stages because locating and correcting faults in a modular system is easier and requires a smaller retest effort than in non-modular systems. Built-in modularity is also an asset in the operational life of a system. The FDA requires that once a system has been validated, it must remain validated for the rest of its life cycle. Therefore, implementing changes is something that should be done cautiously. Implementing and validating changes is easier in a modular system than in a non-modular system.

The effort of adapting recipes to make different grades of the same product or even developing new recipes to make new products has been greatly reduced thanks to S88. Adapting an existing recipe or developing a new recipe can be done on the (higher) SCADA level without the need for changes of the software in the (lower) control layer. Reuse of software in the control layer plays a key role in this concept. Reuse of validated software causes a huge drop in validation effort of new recipes. This is especially important in the current marketplace where it is of paramount importance that the time to market of new products should be kept as short as possible.

Conclusion

S88 is generally considered as the international standard for designing and automating batch processes. The advantages of applying S88 are:

- Use of standard terminology. Everyone speaks the same language.
- Modular set up of both process equipment and batch control software. Modularity increases maintainability.
- Flexibility and ease of recipe development for both grades of existing products and new products.

These advantages lead to a lower TCO for batch facilities that have been designed and built in compliance with S88.

GAMP and S88 are two items that cannot be compared to each other because they exist on different levels of abstractness. Under certain circumstances, these two items may very well be associated with each other. This association exists when a batch project is to be realized in the regulated pharmaceutical industries. The user (the pharmaceutical company) can demand from the supplier (engineering contractor and/or system integrator) to execute batch projects in an S88 compliant manner. However, GAMP does not lay down the usage of S88 as the design methodology.

S88 results in modular batch control software. Modularity is a great advantage when it comes to the qualification (part of validation) of an automated system as prescribed by GAMP. If a software fault should be corrected during the qualification stages, retesting is limited to just a small part of the application. A modular set up is not only a great asset in the qualification stages of the system, but also in maintaining the validated state of that system. A change in a modular application will lead to a much smaller (validation) impact as compared to a non-modular application.

Another great advantage of S88 is the flexibility and ease of recipe development. Previously, development of new recipes required changes of the software in the (lower) control layer. In S88 compliant batch facilities, development of new recipes can take place on the (higher) SCADA level without the services of a control systems engineer. New unit procedures (an important element in a recipe) can be developed simply by rearranging existing software modules. Reuse of validated software reduces the validation effort. Because validation normally takes up a great amount of time, reduction of the validation effort for new recipes greatly shortens the time to market of new products.

References

1. Good Automated Manufacturing Practice, (GAMP 4) Guide for Validation of Automated Systems, Version 4.0, December 2001, ISPE.

2. ANSI/ISA S88.01-1995: Batch Control Part 1: Models and Terminology, 1995: Instrument Society of America.

3. Brandl, D., S88.01 - The Standard for Flexible Manufacturing and Batch Control, The Batch Knowledge Centre (www.batchcentre.tudelft.nl).

4. Man, H.L., and Heesen, M.P., De relatie tussen GAMP en S88 (The relationship between GAMP and S88), Process control, 2002, October: p. 17-21.

About the Authors

Tiny Heesen is a Design Engineer Process Control for the Engineering Department of Diosynth. He has been working in hardware design and in industrial IT for 32 years mainly in the pharmaceutical manufacturing. He is an expert in definition and design of automated systems. He has extensive experience in the application of the quality systems GMP/GAMP and in implementing the S88.01 standard for batch controlled pharmaceutical manufacturing. Currently, he is involved in implementing the requirements concerning 21 CFR Part 11. He can be contacted at: m.heesen@diosynth.com.

Engineering Department Diosynth B.V., P.O. Box 20, 5340 BH, Oss, The Netherlands.

Rudy Kuijer is a consultant for Ordina Technical Automation. He was educated at the University of Twente in the Netherlands, graduating with an MSc in electrical engineering. He has been working in industrial IT for 15 years. He is an expert in definition and design of automated systems and the application of quality systems (ISO 9001, GMP / GAMP) in different industries: food, pharmaceutical, ICT, department of public works. He is a (part time) quality manager for Ordina and member of the Pharma Innovation Team of Ordina Technical Automation. He can be contacted at: rudy.kuijer@ordina.nl or farma.tti@ordina.nl.

Ordina Technical Automation B.V., Science Park 5602, 5692 EN, Son, The Netherlands.

Henk Man is a consultant for Ordina Technical Automation. He was educated at the University of Twente in the Netherlands, graduating with an MSc in chemical engineering. He has been working in industrial IT for eight years, mainly in a great number of DCS projects for several vendors. Among those projects were projects in pharmaceutical batch manufacturing that were realized using S88. As a member of the Pharma Innovation Team of Ordina Technical Automation, he trains customers and fellow Ordina employees in GAMP and 21 CFR Part 11. He can be contacted at: henk.man@ordina.nl or farma.tti@ordina.nl.

Ordina Technical Automation B.V., Science Park 5602, 5692 EN, Son, The Netherlands.

This article discusses some of the more significant developments in API manufacture and supply and explores implications for pharmaceutical engineering.

Authors' Note The authors are developing articles that have in-depth information about API reactor systems, heat transfer, material transfer, ancillary systems, and safety. Readers are asked to submit their anecdotes and opinions on these topics directly to the authors by e-mail or telephone.

APIs: Current Trends

by Stephen Hall and Andy Stoker

Introduction

The Active Pharmaceutical Ingredient (API) sector is facing pressures and changes that few had fully anticipated only a decade ago. Increasingly potent compounds are being developed. Computer technology is enabling more complex processes. Regulatory requirements are being globally harmonized. New chemical synthesis strategies are emerging. Manufacturing costs are under continuous downward pressure. Site security concerns have never been higher. An industry that was once simply international, is now truly global – and this brings with it both opportunities and challenges to current operating models.

This article discusses some of the more significant developments in API supply and explores implications for pharmaceutical engineering. Future articles will focus on practical aspects of plant design, including reactors, heat transfer systems, material handling, ancillary equipment, and safety.

The most successful API manufacturers will be those that are able to adapt quickly to industry needs and global pressures – be they large or small pharmaceutical companies or contract manufacturers. This article will attempt to identify plant design features and strategies that contribute to a nimble facility.

Overview

There continues to be significant growth for API manufacturers, primarily fueled by generic drugs. More than 25 patents for blockbuster drugs are scheduled to expire between 2003 and 2007. As drug patents expire, the generic companies will seek APIs at competitive prices, increasing the demand for third party API manufacturers. In addition, production at major pharmaceutical plants will come under pressure. At the same time, R&D productivity is down and pharmaceutical sales representatives are getting less face time with physicians. Competition between alternative brands of similar products is increasing. There is increasing pressure to contain – and reduce – both operating and capital costs.

Business drivers, such as life-cycle cost and time-to-market, are leading manufacturers to new models. API facilities are being constructed in developing countries as production shifts from Europe and the US to Asia. The number of plants that are pre-engineered and constructed in a factory (modular facilities) is increasing.

Chemistry is changing with chiral chemistry becoming increasingly important, leading to new technologies (such as simulated moving bed reactors) and specialist catalyst usage. Complex chemistries can present new challenges, such as very low temperature processing, high corrosivity, and difficult analytical methods.

Drugs are more potent today with many implications: batch sizes can be smaller, cleaning is more difficult, isolation/containment technologies must be implemented. Regulatory and legislative pressures can be conflicting and difficult to address. GMP regulations are not prescriptive, and standards are subject to the interpretation of individual inspectors. ICH Q7A should be used as the definitive cGMP reference for API facilities. The ISPE *Baseline® Guide for Bulk Pharmaceutical Chemicals* is currently under revision and scheduled for publication in early 2004. The updated version will interpret the GMP requirements and provide designers with a baseline for achieving regulatory compliance.

Other regulatory obligations include building codes, environmental protection, occupational health and safety, and intellectual property product protection rules. Environmental rules and building codes sometimes dictate features that are problematic for GMP compliance. Environmental legislation in Europe and the US is initially passed as overarching requirements; implementation is left to constituent nations and the states with shaded interpretations.

Business Drivers

The pharmaceutical industry has rapidly moved beyond being merely international to a truly global operation. Big pharmaceutical companies plan their operations to form a synchronized manufacturing and supply network. Smaller companies may not have the same reach, but the wholesale dismantling of tariff barriers means that they compete against comparable companies across the world. API manufacturers face similar pressures wherever they are located.

API manufacturers must address the following issues: demand for shareholder value, API production relocation, and technology transfer. These items will be addressed n the following paragraphs.

Demand for Shareholder Value

In the last few years of the 20th century, the pharmaceutical industry saw an unprecedented rate of structural change. The mergers and acquisitions among big pharmaceutical companies are well known, but the changes were much more profound than simple consolidation. Most companies – big or small – now operate within a complex global network of organizations, carrying out research, development, manufacture, marketing, distribution, and a host of other activities. One factor is common – shareholders demand value from these changes in the form of continuing growth and higher margins.

As mergers and acquisitions unfold, there are significant integration problems. New plant construction seems to stop as the existing plants and people are evaluated and reorganized. Some positions are made redundant so corporate experience is lost. Policies, procedures, and attitudes are realigned before significant project activity resumes.

Cost Pressure

The innovators are struggling with decreased R&D productivity. Their sales staffs must work harder to stay in front of busy physicians. At the same time, therapeutic competition from similar branded drugs and generics is fierce. The health insurance industry and government healthcare organizations are constantly striving to reduce spending on drugs.

The response is to implement strategies that get new products to market as quickly as possible, raise revenues for existing products, and reduce costs. Pharmaceutical engineers are most impacted by the cost cutting trends toward "operational effectiveness" and asset utilization.

Operational effectiveness attributes include: clear and controlled management processes that make effective use of resources, the appropriate use of technology, and continual improvement through innovation. Modular plant design and construction is an example of this. Reactor system design can be standardized through a global pharmaceutical company, utilizing nearly identical units in multiple locations. Don Hall, President of Engineered Technologies, a manufacturer of pre-engineered modular reactor systems sees operational effectiveness as a competitive advantage. "Too much capital is wasted in 'border disputes' as a plant design progresses from concept through detailed engineering. There should be a standard design that conforms to global codes, cGMPs, etc.

Of course, there are barriers: technology is constantly changing, each client has his own preferences, and engineers don't like to use another engineer's work."[1]

For an API plant, asset utilization implies much more than smart production scheduling. Facilities must be adaptable to accommodate the range of chemistries, potencies, and batch sizes that could be demanded. Turnaround time between batches and campaigns must be fast with effective cleaning and smooth start-up.

API plant turnaround has long been a problem area for operating companies. Normally, such companies run a campaign of one product or intermediate before cleaning the plant and reconfiguring it for another. This can be costly. Lost production time on a reaction train may easily equate to a cost of several tens of thousands of US $ or € per day. Moreover, plants that are difficult to turnaround encourage extended campaigns and lengthy overall production cycles. Turnarounds lasting months rather than weeks often lead to campaigns taking over a year to move from starting material to the final stage with the consequent costs of work-in-progress material and lack of flexibility to meet changes in demand.

Thus, operating companies increasingly focus on design to facilitate cleaning (*e.g.*, long-radius bends, elimination of dead-legs and convoluted passages) and designate turnaround leaders, together with the resources necessary to reduce lost time. One such company has reported turnaround durations being slashed by 50% through appropriate staffing levels at critical periods, better documentation, and thorough planning. In addition, optimized cleaning methods, integration of re-rigging activities, and structured start-ups have enabled the rapid onset of trouble-free production.[2]

Time-to-Market

Just as big pharmaceutical companies struggle to shorten product development time so they can maximize effective patent life, the generic firms compete intensely to be the first on the market with their product, when products come off patent. This is reinforced by the 1984 Hatch-Waxman Act in the US which provides 180 days of exclusivity for the first successful generic patent challenge. The first to market a substitute is often the big winner in the competitive generic business.

Hatch-Waxman also permits patent owners to recover time lost during a protracted drug approval from the FDA. Patent protection is normally 20 years; up to an additional five years protection is granted. Congress included the extra time to assuage patent holders and balance the provisions that permit generic companies to develop manufacturing processes and submit bioequivalence studies *prior* to the drug going off patent. The generic versions can be manufactured and be ready for market immediately upon patent expiration. New provisions in 2003 close loopholes that, in the past, enabled innovators to delay entry of the generics indefinitely, or that enabled an innovator to enter into a private agreement with a generic firm that would effectively delay the start of the 180-day exclusivity period.

Europe has a much more restrictive law. Supplementary

Protection Certificates (SPC) prohibit plants in EU countries from supplying any drug for qualifying and registration purposes before the drug has lost patent protection. This effectively extends patent protection for an additional six years.

APIs Production Relocation

These circumstances (cost pressures and Europe's SPCs), along with global politics, have fueled the development of the API industry in India and China. The FDA maintains a list of Drug Master Files (DMF) submitted by manufacturers that want to obtain approval for generic versions of drugs (using the Abbreviated New Drug Application process). While DMF details are confidential, their existence is public record. We analyzed the DMF filings and found a steady increase in Asian submittals – *Figure 1*. Moreover, within the Asian segment there has been a shift from Japan toward India and China.

API manufacture in China is increasing and it is reportedly due to their cost advantage.[4] However, chemical manufacturing is technology and/or capital intensive, resources that are scarce in China compared to their biggest advantage, the large labor pool. The competitive advantage for Chinese chemicals is perhaps unfounded except for the intensity with which at least certain Chinese provinces are pursuing the endeavor.[3] Western innovators are wary of transferring technology to China because it may not remain secure. Chinese products have therefore tended to be chemicals with well-known synthesis paths.

India is known for its low-cost base combined with strong R&D capability. Indian pharmaceutical companies have complex synthesis capabilities and experience with cGMP compliance. Their large local dose market gives them the ability to test their products. There are many strong chemists with PhDs from the US or Europe.

A recent development is that partnerships have developed between advanced R&D companies in India and API manufacturing firms in China. "An API manufacturer that combines the low R&D costs of India with the low manufacturing costs of China would produce active ingredients at an exceptional price."[4] The inevitable result is that Western API manufacturers must either try to compete with Indian-Chinese partnerships on the basis of price or provide added value – which includes an extremely high level of custom synthesis capabilities.

Historically, three factors have restricted the extent to which large pharmaceutical companies choose to manufacture APIs in India or China:

- Western manufacturers have had the advantage of being located near to the discovery and development sites, where synthetic routes are first developed. This has enabled them to develop a high level of synthesis capability, enhanced by the ease of technology transfer this provides.

- Large pharmaceutical companies are very reluctant to transfer technology where this incurs commercial risk. The slow progress on product patent law enforcement,
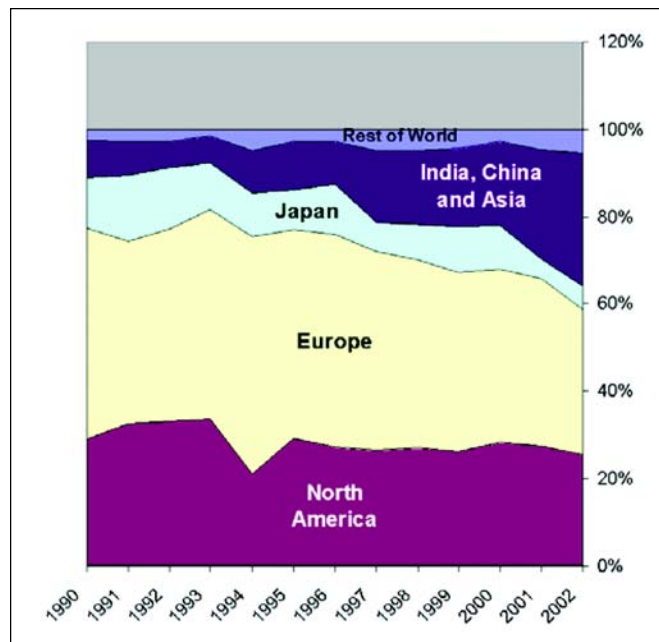

Figure 1. Drug Master Files. Since 1990, the number of type II DMF submittals has held steady at about 300 per year, but the percentage of submittals from Asia has increased from 20% to 37% with all of the increase coming from India and China. (Source: AMEC analysis of FDA data).

coupled with a high level of product counterfeiting, are major deterrents to technology transfer.

- Industry mergers and acquisitions, improvements in operating performance, and the decreased volume demand through greater product potency have led to western API facilities having excess capacity that they have used in preference to manufacturing elsewhere.

For reasons such as these, there is a very limited migration of API manufacture from the West to India or China. We can expect continued growth from indigenous manufacturers, and in the medium to long-term, this may encourage western companies to transfer production.

In addition, the current tense global political situation may delay or inhibit movement of the API industry to Asia. Consider that a flexible API plant would be capable of synthesizing exactly those chemical materials that could be used in weapons of mass destruction. The Australia Group (AG), an informal forum of 33 nations, agrees that export controls are required for technology or equipment that could be used to manufacture chemical weapons.[5] The United States requires export licenses for such technology, in accordance with the AG and also in fulfillment of obligations under the Chemical Weapons Convention (CWC). In addition, the US licenses so-called "dual-use" technologies, which are designed for a commercial application but could be turned to military purposes. Although export of such technology has not, apparently, been inhibited to date, the legal tools for doing so are in place.

Technology Transfer

One result of these various business forces is the growing importance of technology transfer. Processes developed in the laboratory of drug discovery organizations may be transferred to the site of a major pharmaceutical company for clinical trials manufacture and then elsewhere for commercial launch. Subsequently it may move – at a different scale – to a contract manufacturer for mature phase production. Successful technology transfer demands careful analysis and management of technical, regulatory, and organizational issues in all areas of owner and supply chain operations.[6]

Chemical Manufacturing

New Chemical Entities (NCEs) are usually invented, developed, and clinically tested by the major pharmaceutical companies and their collaborators. The innovator typically obtains patent protection, and often produces APIs in their own facilities. Many of these plants are old.

On the other hand, off-patent generics are usually formulated and packaged by one company that purchases the APIs from a contract manufacturer. The API manufacturing process can be developed by the contract firm, provided by the generic company, or developed by the two companies working together. APIs for generics are usually tested for *bioequivalence* to the branded version; clinical trials are not necessary. While the basic chemistry may be similar, the pressures faced by manufacturers of branded drugs differ from generic API manufacture in several important ways.

In 2000, more than 80% of the API facilities inspected by the FDA were comprised of batch organic chemical synthesis technology. Although an increase in biotechnology is occurring, chemical synthesis in the ubiquitous glass-lined reactor continues to be the workhorse of the industry, and most API plants still utilize stirred tank reactors.

Challenging Performance Standards

The great majority of synthetic API processes are focused on stirred tank reactors – commonly made of glass-lined mild steel or a high-grade steel alloy. Such reactors have many advantages, not least their flexibility to accommodate a wide range of processes and the familiarity with them built up by operating companies. Moreover, countless such plants already exist and many are under-utilized. However, an ever-growing body of evidence proves that such equipment is not always the best solution.

Britest Ltd, a "not-for-profit" company, is based in the UK.[7] This organization comprises industrial members, such as GlaxoSmithKline, Avecia, and Rhodia, and academic partners (University of Manchester Institute of Science and Technology and London's Imperial College). Britest seeks to set new standards in performance by step change improvement in competitiveness. They do this by challenging existing process concepts with structured analyses of business, chemistry, and process driving forces. This results in first principle specifications of equipment performance and plant configurations. For example, charting heat transfer and mass transfer performance of various reactor configurations shows that the traditional stirred tank model has the worst performance compared to alternatives such as loop reactors, microreactors, and spinning disc reactors. Early results have been highly encouraging with operating companies reporting significant savings in capital and operating costs, together with improvements to product quality.[8]

Chiral Compounds

Chiral chemistry is the "rising star" in pharmaceutical intermediates technology.[9] Most reported developments in API production relate to chiral chemistry. Single enantiomer compounds – those that are one mirror image form of a chemical that can exist as two distinct structures – comprise 36% of the API market and are growing approximately 10% annually.[10] According to Degussa sales literature, 80% of all pharmaceutical products currently under development are enantiomerically pure.

Pure enantiomers are produced for at least two reasons: 1) by developing single-enantiomer forms of drugs that had previously been approved as racemates (a mixture of two enantiomers), innovators can obtain new patent life (a key defensive move against generic competition) and 2) pure enantiomers may have a better therapeutic effect or the unwanted form may actually be toxic. A drug that works perfectly well as a racemate, even though one of the stereo forms is non-therapeutic (but not harmful, either), can be reformulated as a single-enantiomer to gain new life. But NCEs being developed may only be useful in the single- form.

It is difficult to produce the pure compounds. In one scheme, asymmetric catalysts are developed that are specific to a particular reaction. The catalyst blocks reaction at the unwanted stereo sites. However, the catalysts may have harmful components that can contaminate the batch (*e.g.*, metals), are often restricted to single use, and may require extreme operating conditions (such as cryogenic temperature).

Another scheme utilizes biocatalysts which are enzymes or enzyme-containing microorganisms. When available, a biocatalyst may shorten the total reaction time by accomplishing the chemical transformation in fewer steps compared with other methods.

Pure enantiomers also can be isolated from racemates (two distinct structures) using chiral liquid chromatographic separation. Many big pharmaceutical companies are testing Simulated Moving Bed (SMB) chromatography to separate two or more fractions semi-continuously from racemate mixtures.

The trend to chiral chemistry impacts pharmaceutical engineers who must design for wider temperature ranges (especially low temperature reaction conditions), accommodate new catalysts, specify new equipment (such as SMB), and develop new analytical techniques (both in-situ and off line).

Other Chemistry Trends

There are many developments in the chemistry field that are well beyond the scope of this article. For example, at least one

company is developing synthesis routes that are undertaken at room temperature (instead of cryogenic) with the major drawback being the explosive nature of the catalytic chemistry. Diazomethane, a toxic and explosive gas with the heat of explosion being similar to the RDX used for military purposes, presents a very unusual handling problem.[11]

Other difficult reactions involve hydrogen cyanide, air-sensitive and water-sensitive reactions, radiosynthesis, n-butyl lithium, and triphosgene. Separative reactors which allow a chemical reaction and a separation to occur simultaneously have attracted much academic interest because of their potential to improve the economics of key processes. Reactive distillation is one such example. However, this interest is not yet matched by the commercial uptake. Technological barriers include inadequate scale-up and simulation capability, lack of validated thermodynamic and kinetic data, and a shortage of suitable materials for some types of catalyst.[12]

Some researchers are using smart polymers with properties that can be varied in a controlled manner. For example, some polymers can be given temporary magnetic properties under certain conditions, enabling selective reactions and separations - *Figure 2a and 2b.*

Highly Potent Compounds

According to a report by SRI Consulting, some 25% of the 7,000 drugs currently under development involve highly potent API compounds, compared with 5-10% of the drugs currently on the market.[13] This is due to several factors. New chemical methods permit the development of very refined molecules; instead of having a large molecule with a small active site, the active portion can be isolated which makes the compound highly potent. Therapies against tumors often require extremely toxic chemicals, and this is an area of intense research. Also, drug delivery methods, such as transdermal patches, demand chemical forms that are more potent because of their structure or particle size.

The Occupational Exposure Limit (OEL) for such compounds is typically less than 10 µg (8-hr time weighted average). An implication of being more potent is that lower production volumes are required; this leads to synthesis in smaller equipment.

API plants contain many sources for airborne dusts that would adversely affect the health of employees, including dispensing, sampling, centrifugation, drying, milling and packing. New plants are designed to contain these operations with secondary containment provided by filtering air exhausted from the process rooms. However, it can be problematical to retrofit existing facilities to achieve the same containment level.

Designing for containment requires a holistic approach. All potential sources of operator exposure are identified and characterized. Ways to reduce exposure to acceptable limits are evaluated. The design should account for the instantaneous and time-weighted quantities of harmful dusts and vapors that reach an operator's breathing zone. Additional considerations include: cGMPs, the chemical process (can it

Figure 2. Potent compounds are dispensed into a reactor using a glove box. The reactor and associated components (2a) are separated from dispensing (2b), thus protected from external contamination, by the wall. (Source: AMEC).

be altered?), safety, training, instrumentation and controls, ergonomics, and cost. Supplementary activities such as cleaning and maintenance must be factored.

ICH Q7A has important guidance for facilities processing highly potent compounds. "The use of dedicated production areas should be considered when material of high pharmacological activity or toxicity is involved (*e.g.*, certain steroids or cytotoxic anti-cancer agents) unless validated inactivation and/or cleaning procedures are established and maintained. Any production activities of highly toxic non-pharmaceutical materials, such as herbicides and pesticides, should not be conducted using the buildings and/or equipment being used for the production of APIs."[14]

Regulatory Issues

International pharmaceutical regulations are gradually being harmonized – but there is still some considerable way to go. In the meantime, manufacture – like all other aspects of the supply chain – must take account of differing regulations in both the host country and the markets where the products are sold. This is just one factor that must be considered by organizations that expect to transfer technology and reorganize their global supply networks.

Good Manufacturing Practices

"cGMPs are not regulatory or statutory requirements for APIs, therefore not legally binding, but utilizing cGMP guidance is important in order to produce products that meet the purity standards claimed by the manufacturer." (Edwin Rivera Martinez, FDA).

The two best sources for cGMP guidance are ISPE's Bulk Pharmaceutical Chemicals *Baseline® Guide*,[15] and ICH Q7A, *GMP Guidance for Active Pharmaceutical Ingredients*.[14] ISPE's *Baseline® Guide* was written by pharmaceutical engineers and represents a consensus (with the FDA) as to what design, commissioning and qualification concepts constitute regulatory compliance. ICH Q7A is an internationally accepted guidance document regarding good manufacturing practice for the manufacturing of active pharmaceutical ingredients. Prepared and published by the International Conference on Harmonization, it was adopted by the FDA in August 2001, who will "use it exclusively as guidance for measuring compliance practices."[16]

Active Pharmaceutical Ingredients

ICH Q7A applies to chemical manufacturing, beginning with the step when API starting material is introduced into the process. Starting material is defined as material used in production of an API that is incorporated as a *significant structural fragment* into the structure of the API. Starting materials normally have defined chemical properties and structure. It is up to the manufacturer to determine and justify the step where starting materials are introduced.

Environmental Protection

Compliance with environmental laws tends to be reactive because the cost and effort required are high. Laws vary widely around the world. This complicates technology transfer, and can provide significant cost advantages to plants in less regulated countries.

Air emissions are often controlled by collecting vapors in vent pipes and using end-of-pipe treatment technology such as incineration, carbon adsorption, or cryogenic condensation. There is a trend away from using site-wide "envelope" analysis; environmental laws in the US and UK require that each vent be characterized and appropriately treated. In the US, Maximum Achievable Control Technologies (MACT) are required for major pollution sources. The UK issued the Integrated Pollution and Prevention Control Directive (IPPC) that requires written forecasts of production technologies and raw materials.

Secondary sources of air pollution also must be addressed in the US and Europe. These include fugitive emissions from leaking equipment and evaporation from wastewater and solid waste effluents.

Wastewater effluents may be the next big challenge. In addition to Volatile Organic Chemicals (VOCs) that evaporate from the water, there is growing concern that active pharmaceutical ingredients, in minute quantities, are migrating into the environment by way of wastewater. APIs may not be destroyed or removed by conventional treatment technology. Studies are underway to determine their ultimate fate, including the possible accumulation in plants and animals.

Solid wastes, such as used filter cloths, drum liners and residual chemicals, should be collected and isolated for proper disposal. Defining "proper" is the key question and can have a major impact on the plant's bottom line. Options range from simple landfilling to incineration. Consideration should be given to the potential long-term effects of disposal, but the feasibility of the chosen solution may change with time. UK regulations now prohibit disposal routes that some companies used earlier. For example, all liquid hazardous waste and all corrosive, flammable or infectious waste is now banned from landfill, while all other wastes have to be clearly identified and labeled.

Emergency planning is a crucial component of environmental protection. A spill, leak, or deflagration could cause significant harm if highly potent compounds were released. Expensive cleanup would result from the release of standard solvents. Secondary containment has long been the answer to this problem. Exhaust air from building ventilation systems is often filtered to capture fugitive emissions and toxic dusts.

Equipment that requires an emergency vent (for overpressure protection in the event internal deflagration) can be built to withstand and contain the pressure wave, thus eliminating the need for the vent.

Security

Global terrorism has raised concern over the safety of chemical plants. A typical API facility may not look like a high-profile target, but due to the potential consequences of a plant-destroying event, site security warrants more than a cursory glance.

Security measures should consider many factors including:

- *Physical security*: lighting, intrusion detection, access controls, grounds landscaping, physical barriers, projectile shields, guard force
- *Computer and utility protection:* cyber barriers, failsafe computer backup, redundant utilities
- *Emergency planning, training and exercises:* coordinate response planning, provide for certified training, consider blast and fire-safe control rooms, evaluate potential crime impact, establish testing and maintenance schedules, inspect emergency exits, appoint on-site response teams
- *Process control:* establish safe shutdown procedures, investigate add-on safety equipment, plan for product transportation
- *Design:* prioritize safety in design, reduce/eliminate inherent hazards, make architects aware of safety concerns, establish construction materials standards, evaluate setbacks from property boundaries to create buffer zone
- *Auditing procedures:* establish materials accounting procedures, establish theft prevention guidelines, audit internal security, provide for certified third-party audits
- *Administrative controls:* establish policy statements, line item security budgeting, security record-keeping systems, labor dialogue.

Conclusion

Many changes in chemistry, the business environment, and world affairs are described in this article. The API industry is profoundly affected; pharmaceutical engineers must be aware so that new plants are designed to compete effectively in the global market and that the service life of existing plants can be maximized.

Specific API plant design features will be discussed in future articles, emphasizing practical solutions and design procedures.

References

1. Personal communication with Don Hall, Engineered Technology Corporation, Providence, Rhode Island.

2. Battrum, Mike, Dr., BASF plc, Presentation to IChemE, Newcastle, UK, November 2002.

3. Yue, Changjun, "Comparative Advantage, Exchange Rate, and Exports in China," paper delivered at CERD, conference tilted, "Has China Become a Market Economy?"CERDI, Clermon-Ferrand, France, May 17-18, 2001.

4. Karris, George, "API Manufacturing: How Will Changes in India and China Affect the Outsourcing of APIs,?" *Contract Pharma*, September 2002.

5. Bureau of Export Administration, Chapter 6, "Toxic Chemicals, Chemical Precursors, and Associated Equipment, Technology and Software," *The Foreign Policy Report to Congress,* January 18, 2002.

6. Greally, Declan, and Edwards, Roger, "Transferring a Genetically Engineered Biopharmaceutical from Research to Clinical Development - Impact on Facility Design and Build Projects," *Pharmaceutical Engineering,* 22 (4), July/August 2002, p35-46.

7. BRITEST stands for "Best Route Innovative Technology Evaluation and Selection Techniques," www.britest.co.uk.

8. Personal communication with Professor Paul Sharatt, UMIST.

9. Frost & Sullivan, "European Pharmaceutical Intermediates Market," Report Number B027, November 8, 2002.

10. Rouhi, A. Maureen, "Chiral Roundup," *Chemical Engineering News,* Volume 80, Number 23, June 10, 2002.

11. Archibald, Thomas, "Safe Reactions," supplement to *Chemical Specialties Magazine,* October 2001.

12. "Vision 2020:2000 - Separations Roadmap," published by the US Center for Waste Reduction Technologies of the AIChE in cooperation with the US Department of Energy.

]13. Commetta, Chiara, et.al., "Active Pharmaceutical Ingredients," published by SRI Consulting, July 2001. http://scup.sric.sri.com/Public/Reports/MEDIC000/.

14. Food and Drug Administration, "Guidance for Industry: Q7A Good Manufacturing Practice Guidance for Active Pharmaceutical Ingredients," August 2001. Also available from the European Agency for Evaluation of Medicinal Products (EMEA), the International Federation of Pharmaceutical Manufacturers Association (IFPMA), the Pharmaceutical Inspection Convention, and the National Institute of Health Sciences (Japan).

15. ISPE, *Baseline®Guide: Volume 1 - Bulk Pharmaceutical Chemicals,* First Edition, June 1996. (to be reissued in 2004).

16. Quote from Edwin Rivera Martinez of the FDA's compliance office of the Center for Drug Evaluation and Research, from Synthetic Organic Chemical Manufacturers Association (SOCMA), Current Good Manufacturing Practices Workshop," Washington, D.C., September 23, 2002.

About the Authors

Stephen Hall is a chemical engineer (University of Pennsylvania, 1974) with a Masters in business Administration (Drexel University). He started in the process development department at Occidental Chemicals before joining Corporate Engineering at SmithKline Beckman (now GSK) in the 1980s. Hall has worked in the pharmaceutical engineering consulting field for nearly 20 years as project manager and process engineer. He's been a speaker at ISPE regional and national meetings and has contributed papers to *Chemical Engineering* and *Chemical Engineering Progress*. He is now Director of Process Technology at AMEC's pharmaceutical engineering center in Lawrenceville, NJ. He can be contacted by email: stephen.hall@amec.com or by telephone: (609) 219-9266.

AMEC, 3131 Princeton Pike, Bldg. 5, Suite 105, Lawrenceville, NJ 08648.

Andy Stoker graduated in chemical engineering from the University of Cambridge in 1976. His early roles included scale-up, design development and plant production support for Boots Pharmaceuticals and Fisons Pharmaceuticals. He holds a PhD from the University of Edinburgh for a study of pharmaceutical technology transfer with special reference to India and Indonesia. He has recently worked as a manufacturing strategy analyst for a world-class pharmaceutical company. His current role is Pharmaceutical Specialist in the UK offices of AMEC, a global engineering and construction company (www.amec.com). He can be contacted by email: andy.stoker@amec.com or by telephone: +44(0)1925 281800 ext.1625.

AMEC, 410 Birchwood Blvd., Birchwood, Warrington, Cheshire WA3 7WD, United Kingdom.

This article discusses two life cycle documents: the User Requirements Specification (URS) and the Functional Specification (FS).

GAMP Life Cycle Documents as Effective Communications Tools

by Matt Bothe, PE

Introduction

Imagine a resource from which engineers, designers, and end users can draw practical information that compiles the good practices, procedures, and technical competence of top operators and researchers, and has validation significance. This compilation of best practice may be found in the GAMP Guidance, a comprehensive reference of good practice, tailored specifically to automated systems. An article covering all aspects of GAMP could not be justified if limited to a few pages or less; therefore, this article covers two key life cycle documents:

- the User Requirements Specification (URS)
- the Functional Specification (FS)

The User Requirements Specification (URS) is intended to define end user expectations with regard to features, functions, and overall appearance of proposed process equipment and associated operations. The Functional Specification (FS) is intended to provide a working interpretation of the URS and an implementation strategy by external consultants or contractors. Unlike the Code of Federal Regulations, neither the URS nor FS are directly enforced by FDA mandates. Nevertheless, the development and adherence to both makes practical sense with regard to speedy development, competitive engineering, and the FDA-enforced validation practices.

URS and FS Defined

Before the URS and FS can be defined in accordance to GAMP guidance, it is important to understand the association they have with the stages of development or "life cycle" of a particular product line. The series of life cycle documents prepared up to and throughout validation, start-up, and commissioning, begin-ning with the URS and FS, represents only one stage of overall product development, and does not stop at the successful completion of Performance Qualification (PQ) testing. However, since much of the work is required to essentially "break the ground" of a new production line (or product line enhancement), many entities with a broad range of differentiated tasks are often involved, thereby requiring some mechanism to organize the thought processes and communication among players; and as a tool to coordinate the construction effort. Following PQ and throughout the useful life of the product line, the end user is generally responsible for the continuation, refinement, and maintenance of these life cycle documents although some external consultation or support may be retained. Therefore, the life cycle concept as defined by the GAMP guidance is "an approach to computer system development that begins with identification of the User Requirements, continues through design, integration, qualification, user validation, control and maintenance, and ends only when the commercial use of the system is discontinued." The significance of the FS is emphasized as a key component to life cycle documentation due to its inseparable link to the URS.

According to the GAMP Guidance, the URS "describes what the equipment or system is *supposed* to do, and as such is normally written by the user. The URS may be sent to suppliers as part of the vendor selection process. This version should include all essential requirements (*musts*), and if possible, a prioritized set of desirable requirements (*wants*)." Therefore, the URS is essentially a document that is generated (directly or indirectly) by the end user, often with assistance from an external consultant working on behalf of and sensitive to the performance issues expressed by the end user. It defines the key aspects of system per-

formance following start-up/commissioning as verified by PQ, often initiated during Site Acceptance. Following end user sign-off and distribution, the end user is generally bound by the "agreed-upon" terms and conditions, as detailed in the URS throughout the detailed design and validation periods.

The FS, on the other hand, "describes the detailed functions of the equipment or system, i.e., what the system *will* do. An initial version of the FS may be produced as part of the supplier response. Further revisions of the FS are prepared in conjunction with the user. The FS links to OQ which tests all the functions specified," communicated by those typically responsible for interpreting the URS in an effort to satisfy the objectives of the end user. Therefore, the FS represents the effort generally put forth by those contracted by the end user (or contractor representing the end user) to engineer, design, and build equipment and/or systems to meet end user expectations as outlined in the URS. The assessment performed by the end user of such effort generally occurs during Factory Acceptance Testing (i.e., at the site of development) through the execution of Operational Qualification (OQ) test protocols assembled by validation consultants, auditors, or groups within the end user hierarchy.

Both the URS and FS are vital documents essential to a successful validation program -- the details of which are typically outlined in the Validation Master Plan (VMP). The relative positions of these key life cycle documents (underlined) are illustrated in Figure 1.

Since the author of the FS is typically a contractor hired by the end user to satisfy the objectives of the URS, a traceability matrix is usually advised to objectively identify and link all characteristics and propositions of the FS to the requirements detailed in the URS. Although not generally required, the preparation of a matrix is strongly encouraged to help ensure the FS is an accurate and complete interpretation of the URS. An effective matrix example (Table A) is one that is structured in tabular form, listing all requirements of the URS, followed by a brief description of the methods, procedures, and tangibles (i.e., hardware and software) proposed by the FS to address each item. A 'checked by' column should be included for verification.

URS and FS Enforcement

Although the URS and consequently the FS are not enforceable through FDA validation policy (i.e., 21CFR Parts 11, 210 and 211), they are widely accepted and frequently referenced by FDA compliance auditors in determining the "validatable" state of a process unit, control system, or entire operation. By using the GAMP life cycle documents as communication tools to collect and refine important information pertinent to end user performance expectations and operational requirements, elements of the FDA enforcement can be derived through validation-specific detail as referenced in the URS and FS. These "enforcement factors" add "bite" over standard "Scope Documents" and "Operational Guidelines" most often applied as communication tools.

Since the End User may often lack the time and focus on any particular project (often due to other priorities and/or

resource allocation issues), one may perhaps agree that it can be quite a challenge to secure the assistance of those within an end user hierarchy most closely associated with the operation and maintenance of associated equipment. As a consequence, the dissemination of reliable and complete information on end user-anticipated performance regarding new or revised operational equipment more often than not falls short of an accurate and comprehensive reflection of the actual needs of the ultimate users.

For the sake of accuracy and coverage of information to be included in a user requirements document, management support is essential to ensure all ultimate users (i.e., operations and maintenance personnel) are involved. This end user-sponsored mandatory involvement should persist throughout the evolution of the URS - a level of involvement mandated by the end user management team whose enforcement policy should not interrupt normal activities of the employees, but should be strongly encouraged such that appropriate measures and precautions are applied (in whatever form that is best for the end user). If human resource allocation shortfalls exist and persist, the end user management team should strongly consider external support that can be readily relied upon to fill these voids. The key objective here is to ensure that the URS is an accurate and complete reflection of the precise end user expectations after project completion. Some time also should be allocated by the end user hierarchy to review the content of the FS to ensure external contractors are not only properly interpreting all provisions of the URS, but possess the capacity and competencies to do so.

URS/FS Evolution

Among the multitude of steps required for project planning and execution, all associated disciplines should find resolve through the evolution of a vehicle tied to regulatory mandates and enforcement. From an automation point of view, the URS is typically the first document to be issued following the Basis Of Design (BOD) and VMP documents that govern the design and validation efforts respectively. During the early phases of project execution, the URS may function more like a "scope of activities" document than a document with validation significance. The evolution of the URS (throughout the project life cycle) progressively adds credibility to the document as a "validation mandate" - to be verified by way of PQ testing during the Site Acceptance Testing (SAT) - accepted by not only the end user, but third-party regulatory auditors as well. The FS, generated after the URS has been drafted, may follow a similar evolutionary trail with emphasis placed on equipment operations (verified via OQ protocol execution) as opposed to performance. In many cases, the end user justifiably places more emphasis on the FS (particularly if the engineering and design of a particular process is placed in the hands of outside contractors) to ensure compliance to the requirements of the end user - even if a URS has not formally been prepared. Note, however, given the circumstance just mentioned, an FS without a formal URS places more risk than necessary on the contractor. Therefore, the evolution of a

URS Part	Description of URS Part	FS Part	Description of FS Part	Checked By
2.1	Control System shall utilize brand X PLC...	2.2	Model "A" PLC from X, Inc. shall be applied to perform sequential logic...	QRS
3.4	Analog I/O shall be 4-20mADC with 24 VDC primary power for...	3.4	The PLC analog I/O cards shall be provided with 250 Ohm precision...	TUV
5.2	All discrete tie-ins from field devices shall be fieldbus...	6.3	The discrete field devices shall be linked together using Devicenet...	WXY
7.1	Control System Historian shall be 21 CFR Part 11 compliant...	8.2	The Historian shall consist of OS/drivers, applications S/W, H/W...	QRS

Table A. Sample matrix.

formal URS, with "FDA bite," is strongly encouraged to improve two-way communications and assist in the solidified notion of shared accountability.

Benefits

The advantages derived through the use of GAMP-supported documents as vehicles to establish initial contact and maintain continuous dialog between the engineering/design and build contractors, and between the contractors and end users alike are numerous; ten of these are listed:

1. URS as Viable Communications Tool

The importance of assembling input from the end user cannot be underestimated. After all, the focus of most any design project, pharmaceutical or otherwise, is to provide the end user the equipment and services they require to sustain a competitive margin in their respective market. Communicating the needs of the end user to those entities with the resources to fulfill the expectations of the end user is often a complex task. This is a particularly viable statement if the end user may not be familiar with the variety of options available or lack the resources to best achieve their ultimate goals. Therefore, the preparation and subsequent utilization of GAMP life cycle documents (at least in principle) are well positioned to "unravel" the complexities of associated information exchanges - primarily through the URS and FS as effective communications media.

2. URS as Primary Reference Document

Considering that the BOD reflects the conceptual expectations of the end user with regard to capacity and approach, the URS (as a working and evolutionary document) should be prepared to provide direction and distinction far beyond the intent of its earlier ancestors (including the VMP). The URS must contain sufficient detail to accurately and effectively direct the authors of the FS to assemble a design consistent with the needs of not only those marketing the products of the end user, but also should include the needs of the operators, maintenance technicians, parties involved with safety/environmental regulatory enforcement, procurement specialists, and onsite engineers. The URS and subsequent FS are essential components in the assurance that detailed design efforts, as described in the Detail Design Specifications and other related life cycle documents, follow the appropriate path toward compliance (to both the end user and regulatory agencies).

3. FS as Key Contractor Document

The FS is the key life cycle document by which external contractors base their designs. Therefore, an accurate, complete, and end user-approved version of the FS is essential. Many external integrators and programmers, for example, impose substantial "cost adders" to their proposed project budgets without formal end user acceptance of the FS and all its constituents. These "cost adders" are justified and often necessary to cover the increased risk of frequent and unsupported "after-the-fact" design changes. Therefore, the allocation of some review time toward FS acceptance often saves the end user considerable costs due to "risk reduction," as well as a reduction in unsolicited change requests. As a side note, the end user should be cautious of any contractor who claims the FS is an unnecessary project component for the sake of "budget reduction" and competitive bidding.

4. Required by Validation Auditors

Although not an FDA mandate per se, most (if not all) entities involved with the validation of automated systems reference the URS and related documents to enhance the efficiency and effectiveness of the validation effort, while reducing the total amount of preparation work, risk, and associated costs. Therefore, the use of GAMP guidance encourages the end user to directly involve those responsible for the operation and upkeep of the equipment and systems engineered and specified.

5. URS/FS and the Divisions of Responsibilities

The associations between the URS and FS as described in this article provide clear distinction as to the roles and responsibilities of the external contractors/suppliers and end user. The importance of various project controls aimed at regulating the effects end user "wish lists" and contractor "extras" have on project budgets and schedules is paramount.

6. Evolution of URS/FS

The evolution of the URS/FS helps to ensure continuity from one execution phase to the next, and may, depending on the nature of the documents prepared, assist in interdisciplinary coordination - often a complex "web to untangle."

7. Management Of Change (MOC)

The URS/FS can be viable tools during their evolution to help manage change requests by both the end user and contractors associated with a particular project. The value of the Trace-

ability Matrix can be felt by ensuring the FS is updated, reviewed, and approved accordingly. The URS/FS should not replace existing MOC procedures on either the end user or contractor side, but be structured to work with these (often existing) sets of vital policies and procedures. Note that it is not uncommon for an entity's MOC processes to have strict CFR implications due to their potential impacts on the environment and safety (i.e., 29 CFR, Part 1910.119).

8. Work Scope Boundaries
Some end users desire turnkey installations with little internal employee involvement, others desire absolute control. In either case, the URS, through an element of enforcement, and the FS, as a formal interpretation of user requirements, may better define the lines between designer creativity and user expectations, helping to minimize the occurrence of stray tasks.

9. Establishing Positive Relationships
Using the URS as a "scope of design" and the FS as an "acknowledgement of design," their transformations to comprehensive and formal validation documents for use by external validation consultants is greatly facilitated (and may even be greatly appreciated by both auditors and end user alike, thereby leading to additional opportunities for relationship building).

10. Standardization through GAMP
The URS, FS, along with other key life-cycle documents, are established validation deliverables described in the GAMP guidance to provide an element of standardization throughout the industry. This standardization effort provides a common language, terminology, and procedural task flow essential to accurate, comprehensive, and competitive project implementation and maintenance - it simply makes sense.

Significance of 21 CFR Part 11
With the increasing availability of highly sophisticated computer systems capable of processing and centrally registering enormous quantities of information, the advantages of full-scale automation had become evident. However, with the increased power and capacity of computer systems, coupled with the desire to electronically register, process, and file production data in the form of batch records, security likewise had become an area of considerable concern. For this reason, the FDA (in association with other regulatory agencies, standards organizations, manufacturing facilities, and contractors) have compiled a set of rules that the end user, and those in association with the end user, must comply with in order to maintain, manage, and preserve the integrity of historical electronic records. This set of rules are collectively presented and organized in the 21 Code of Federal Regulations (CFR) Part 11 (or Part 11).

The Part 11 structure consists of two key components: 1) Electronic Records and 2) Electronic Signatures. Per code, the concept of Electronic Records (or Erecs) "applies to those in electronic form that are created, modified, maintained, archived, retrieved, or transmitted, under any records requirements set forth in Agency regulations." Therefore, for example, it is not in the best interests of the end user to assume all data in electronic form are subject to Part 11. Therefore, the URS should be the key component precisely differentiating what information collected (by the end user) is vital for maintaining and ensuring the safety and efficacy of their human-consumable product offerings from data to be used for academic and internal purposes.

To ensure continued compliance to Part 11, the Electronic Signatures (or Esigs) component links the electronic record (and any data contained within the record) to an individual or group that can verify the information is derived from reliable validated sources.
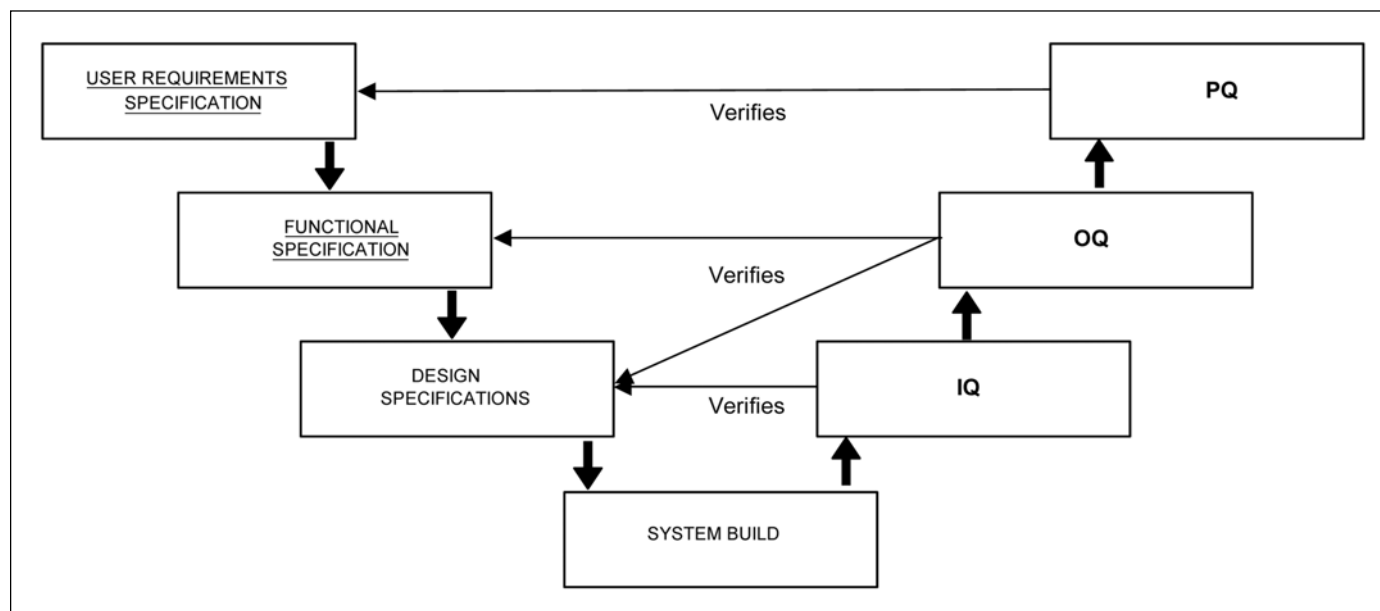

Figure 1. A basic framework for specification and qualification.

In addition to the identification of data subject to Part 11 policy, the URS (and subsequently the FS) should contain references and/or statements regarding specific procedures detailing the handling, security, and authentication of such information (via username/password prompts, audit trails, and other metrics). The overall objective of the URS/FS combination is to ensure end user compliance to Part 11 and the total elimination of "483" violation letters issued by FDA auditors.

This article is not intended to cover all the details of Part 11 policy or its implementation. However, the importance of the URS/FS as vital communications tools among those parties involved in Part 11 interpretation, implementation, and compliance is stressed. Note that Part 11 is a subcomponent of validation and is not intended to replace any components of the validation process.

Conclusion

Until the next revolutionary communications processes emerge, and more viable tools are identified, GAMP philosophy can be readily applied in validated projects such as those with an automation scope. Starting with the URS/FS combination of life cycle documents, the GAMP methodology may be employed to help manage dialog between the contractors and associated parties (including the end user), establish enforcement protocols, and cover bases of inconsistencies and deviations. As the URS and subsequently the FS evolve throughout the detailed design phases, for example, the compromise between flexibility and rigidity can lead to stronger and more profitable end user-contractor relationships well into the future…and of course, a positive reputation breeds a persistent flow of opportunities (not only for the end user through increased profitability, but for the contractor's continued contributions to the bottom line of the customer).

References

1. *GAMP Guide for Validation of Automated Systems*, GAMP 4 (ISPE (GAMP Forum), 2001).

2. Code of Federal Regulations 21, Parts 11 (August, 1997), 210, and 211.

3. CRB Reference Library on GAMP-related Policies and Procedures. CRB Consulting Engineers, Plymouth Meeting, PA. Various applications and releases.

About the Author

Matt Bothe is a licensed Professional Engineer in four states (NY, CT, PA, and TX) and serves as Senior Automation Engineer for CRB Consulting Engineers, Inc. at their Plymouth Meeting, PA office. He is responsible for the specification and selection of instrumentation and process control methodologies (for both continuous and batch operations), and provides comprehensive services in the areas of FAT/SAT/IQ/OQ/PQ, 21 CFR Part 11, and general validation support. He possesses BS degrees in both chemical and electrical engineering from NC State University, Raleigh, NC, a Certificate of Graduate Completion in hybrid circuit design from National Technological University, Fort Collins, CO, and an MBA from Texas A&M-Corpus Christi. He is currently an active member of ISPE and ISA, and former Key Person Chairman for the IMC. His experiences are primarily centered on Instrumentation and Process Controls (including Advanced Control) in a variety of industries that include automotive, power, commodity and bulk chemicals, and biopharmaceuticals. Some of his major accomplishments include the design and installation of a comprehensive 17-node Power Quality Monitoring System, design and installation of redundant model-based PLC-resident Rectifier Control systems for a Chorine-Caustic production facility, advanced custom controller based on Iterative Convergence Control Technology for a major chemicals manufacturer, and participated in and supervised numerous large and small DCS/PCS installations in large scale chemical facilities. In support of his areas of responsibilities, he has created numerous database and software applications using VBA, SQL, and Visual C++. He can be contacted at: matt.bothe@crbusa.com.

CRB Consulting Engineers, Inc., 220 W. Germantown Pike, Suite 170, Plymouth Meeting, PA 19462.

This article discusses an innovative approach to applying single-use disposable fluid path technology to final fill and finish operations.

Single-Use Disposable Filling for Sterile Pharmaceuticals

by Brett Belongia, PhD, Robert Blanck, and Steve Tingley

Introduction

Growth in new protein-based biotechnology products, vaccines, gene therapy drugs, and synthetic chemical entities in injectable form has fueled the need for accurate, safe, and easy-to-use process systems that meet the demands of today's fast paced drug development cycles and increasingly stringent regulatory requirements. An industry need for improved economics, enhanced speed to market, and prevention of cross contamination has driven the development of disposable components. New enabling technologies for disposable manufacturing are optimizing biopharmaceutical drug production by eliminating the need for the conventional cleaning, autoclaving, and steam sterilization of equipment. The recent advent of gamma presterilized flow paths, reservoirs, storage containers, sterile filters, and connecting devices is helping to achieve the highest level of safety by providing for the operator-free intervention of sterile parts and elimination of aseptic assembly of equipment.

Recent innovation has brought the benefits of single-use disposable manufacturing to final fill and finish operations for sterile injectable drugs, ophthalmics, and large volume parenterals. The primary benefits of such single-use technology are the absence of risk associated with product cross contamination along with fast, easy set-up. These benefits can be translated directly to bring drugs to market faster and more cost effectively, primarily by facilitating multi-product filling facilities and reducing the validation burden. The complete disposal of all product contact surfaces eliminates the need for conventional equipment cleaning cycles and minimizes downtime between filling campaigns. Disposable filling lines offer added operator and product safety with the complete containment of the drug, making these filling lines especially suited for cytotoxic or biohazard fills.

Conventional Filling Techniques

In conventional filling operations, product contact surfaces such as tubing, filling needles, reservoirs, and pump components are typically autoclaved or steam Sterilized-In-Place (SIP) before use. Autoclaved components are aseptically assembled on the filling line which is located under laminar flow conditions in a cleanroom or barrier isolator. The contamination of equipment and product contact surfaces can occur at every intervention point during the installation process. The assembly process is usually time consuming and is complicated by the restricted movement of an operator wearing protective gear in the cleanroom environment.

Figure 1. A disposable filling system showing the disposable reservoir, tubing and needle assembly along with hardware for mounting on a filling line.
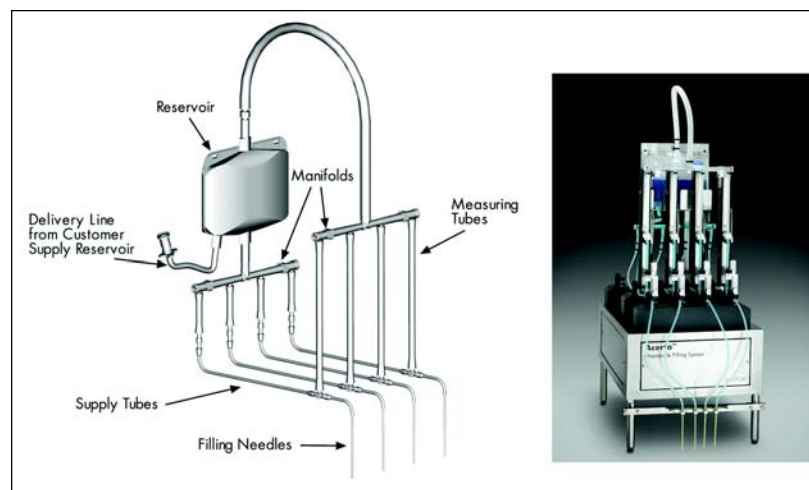
After the filling operation is complete, the product contact parts must be disassembled, cleaned and sterilized before reuse. Cleaning regimes usually involve the use of caustic and/or cleaning solutions along with Water-For-Injection (WFI). Depending on the type of filler and associated reservoirs, the cleaning process may take several hours and require special cleaning equipment. It is important that the cleaning process be validated before implementation to ensure the absence of product carryover from previous batches, and that proper cleaning cycles are performed by following predetermined Standard Operating Procedures (SOPs). Adding to the complexity of the cleaning, accurate cleaning data must be recorded and kept for possible regulatory inspections. These required procedures often delay the rapid introduction of a product into the marketplace or add a level of inefficiency to the daily manufacturing environment.

Conventional filling machines use either a piston pump, diaphragm pump, or time-pressure system to deliver the product to the vials. In a piston pump system, a reciprocating piston accurately dispenses a predetermined volume after each positive stroke of the pump. The accuracy of the pump is determined by the precise fitting of the highly machined parts and is most accurate when new. Filling accuracy can be compromised by pump wear during the filling cycle with the distinct possibility of introducing particles into the product. Care must be taken to prevent pump damage during the regular assembly and disassembly of components after filling. Damage to the machined parts compromises accuracy and precision, and in most cases, renders that part useless. It is common for companies to prepare spare pumps and components for use in the event of mishandling or problems during critical filling operations. In some instances, piston systems can be steamed in place, but close tolerances around the piston and the cylinder complicate the sterilization process.

Time-pressure filling involves an enclosed pressurized reservoir, flexible tubing connecting the reservoir to the filling needle, a pinch valve, and a timer. The tank is usually pressurized by a nitrogen supply tank and the desired fill volume is achieved by a timed opening and closing of the pinch valve. The technique has minimal moving parts and has gained some acceptance because it is relatively simple to use. The disadvantage of time pressure filling is the variability of the process, especially for lower volume fills. The components also require regular cleaning and sterilization with a steam-in-place set-up.

Disposable Filling Technology

Disposable filling is based on a liquid dosing principle that uses pre-sterilized disposable components for all product contact surfaces. Gamma pre-sterilized and pre-assembled components provide greater filling process security by eliminating the need for the aseptic assembly of pumps, reservoirs, tubing, and needles. The technique provides an improved level of flexibility and ease of use by allowing the easy changeover of product contact surfaces in a cleanroom or barrier isolator environment.

One type of disposable filling technique is offered by peristaltic pump systems. These are single or multiple head pumps that operate by pushing liquid through a flexible tube with a hardened roller. In a filling line, the pump is operated intermittently, dispensing a measured amount of liquid through the filling needle to the vial on the filling conveyor. The pump tubing may be replaced for each filling batch and may be pre-sterilized with an autoclave before insertion into the non-sterile rotating pump head. The system is relatively simple, and there are no piston pumps to assemble or valves that might contribute particulates.

The accuracy and reproducibility of the peristaltic pump system relies on tubing consistency in the pump head. Natural variation in the tubing diameter, wall thickness, and flexibility can affect filling accuracy, while operational variability can result due to stretch, compression, fatigue, and tubing wear. As with conventional filling systems, peristaltic filling systems require that product contact parts such as filling needles be prepared, cleaned, sterilized, and reassembled each filling day. This requirement necessitates a validated cleaning protocol for the safe operation of the system.

A second, novel, disposable filling technology has recently been developed by Millipore Corporation in conjunction with Filvek. The technology provides for an accurate volumetric fill driven by gravity which does not require the use of a pump or a pressured system. The disposable filling system, as illustrated in Figure 1, comprises a hardware component and a disposable pre-sterilized assembly of tubing, reservoir, and filling needles. The disposable component is discarded after each filling operation, obviating the cleaning and sterilization of the product contact surfaces. The single-use device also prevents the possibility of cross contamination between batches. This allows for the complete segregation of drugs using the same filling lines and is especially useful for multi-use facilities or contract filling operations. The integrated design of the disposable filling device, including needles, tubing, and reservoir provides an extra level of operator safety during equipment disassembly, minimizing the risk of exposure to the drug product. This margin of safety becomes especially advantageous for use with cytotoxic or biohazard materials - *Figure 1*.

The operation of this dispensing system is based on a gravimetric volumetric fill principle and works with a fill and discharge cycle as seen in Figure 2. The reservoir, filled to a set level with the sterile drug, serves as a capacitance buffer and is fed from an external sterile holding tank. The volume of liquid in this reservoir is precisely controlled by a sensor located external to and behind the reservoir which detects and maintains the liquid level to approximately 0.5 liters. This control is achieved by a feedback loop from the sensor to a peristaltic pump or a pinch valve fed from the sterile holding tank.

When the upper pinch valve (Figure 2) opens, liquid, driven by a small head of pressure, flows from the reservoir into the measuring tube. The upper sensor on the measuring
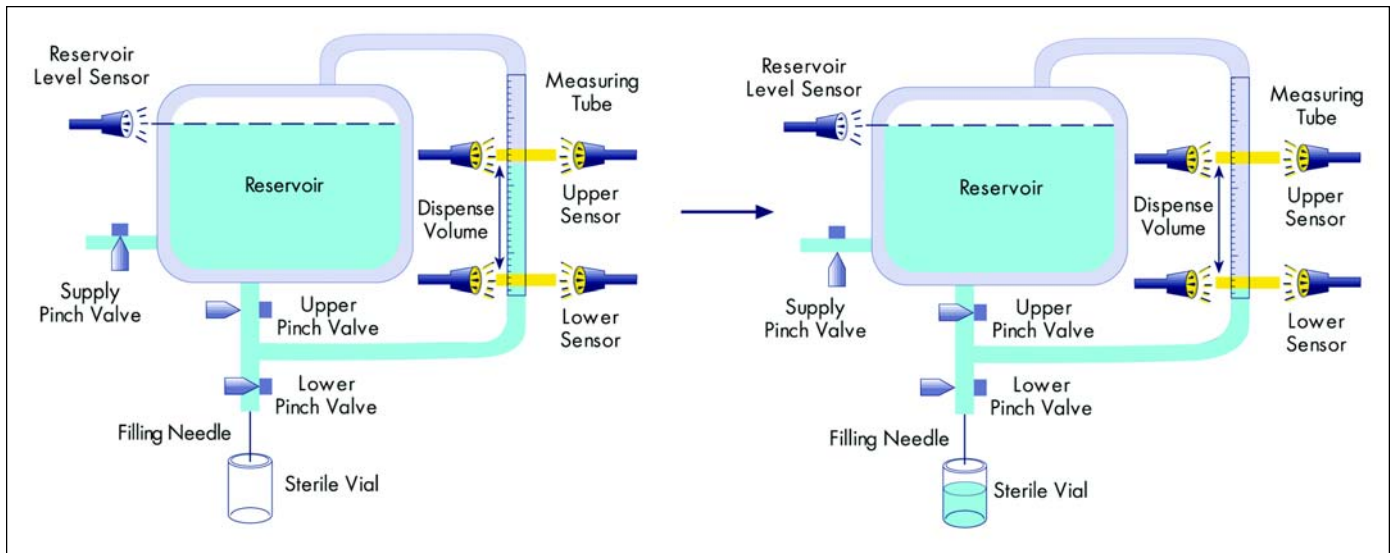
Figure 2. A disposable filling system operation. Charging the measuring tube and subsequent dispensing of the sterile product through the filling needles.

tube recognizes the liquid meniscus and closes the upper pinch valve. The measuring tube is now filled with a reproducible volume of liquid. The charged measuring tube is now ready for dispensing to the vial. Depending on the number of filling heads required for the line, the reservoir may feed several measuring tubes. To allow for the equilibration of pressure in the measuring tube with the reservoir head space, the top of the tube is vented back to the reservoir. This provision serves to maintain the sterility of the liquid to be dispensed by forming a closed circuit - *Figure 2*.

In the vial filling cycle, the drug is transferred from the measuring tube to the filling needle located on the filling line. Once the vials are indexed, the lower pinch valve is opened. Again, driven by a head of pressure, the liquid flows past the lower pinch valve into the vial. When the meniscus triggers the lower sensor, the lower pinch valve shuts. A fillcycle has been completed which is capable of delivering a repeatable volumetric fill. The cycle is then repeated as the next vial is indexed to the filling position.

The volume of the fill is determined by adjusting the positions of the sensors which are located on an automatic servo mechanism. During the initial fill, weight checks of the product in the vial are performed and adjustments are made by changing the sensor position.

The disposable components have been designed as an integral assembly to meet specific customer filling needs for one-, two-, four-, six-, eight,- or twelve-needle systems. The reservoir, tubing, measuring tube, and filling needles are manufactured with polymers commonly used in the pharmaceutical and medical device environment such as LDPE, EVA, and platinum-cured silicone. Filling needles are manufactured with a rigid gamma sterilizable polymer such as polyimide or polycarbonate. Materials are selected for their low extractables, low particulate introduction, and are USP Class VI tested.

Other Integrated Sterile Components

The scope of the disposable components used for this filling system may be expanded to include other disposable devices such as sterile filters, sterile valves, connecting tubing, and supply bags. Figure 3 illustrates two possible sets of components for a filling line. The pre-assembled, gamma sterilized components can be multiple bagged before shipment from the supplier, allowing for staged opening of the package in sequential cleanrooms. The disposable components are designed to snap in place in a process that is estimated to take approximately 15 minutes - *Figure 3*.

System Hardware

The system hardware is composed of a stainless steel filling station and a user interface. The filling station has a stand for the mounting of the reservoir and individual supports for each of the measuring tubes. The measuring tube sensors are mounted on servo belts for the automatic individual volume adjustment of each dispensing line. The measuring tubes and flexible tubing are positioned with snap-in-place connections for easy assembly.

The filling station is designed to be integrated into a new or existing filling line and replaces the existing pumping station and surge tank. The fill valves and sensors are regulated by a programmable controller which is located in an accessible area away from the filling line.

Since disposable filling technology is new, project engineering and integration costs currently may be more costly than those for a conventional filling line. This is due to the custom engineering required to match the disposable system hardware components with the existing filling line design. Most modifications are expected to be the placement of the dispensing equipment, mounting needles on the filling bar, integrating the control software and ergonomic considerations for easy reservoir and needle removal. It is expected that these costs will drop as the technology becomes main-
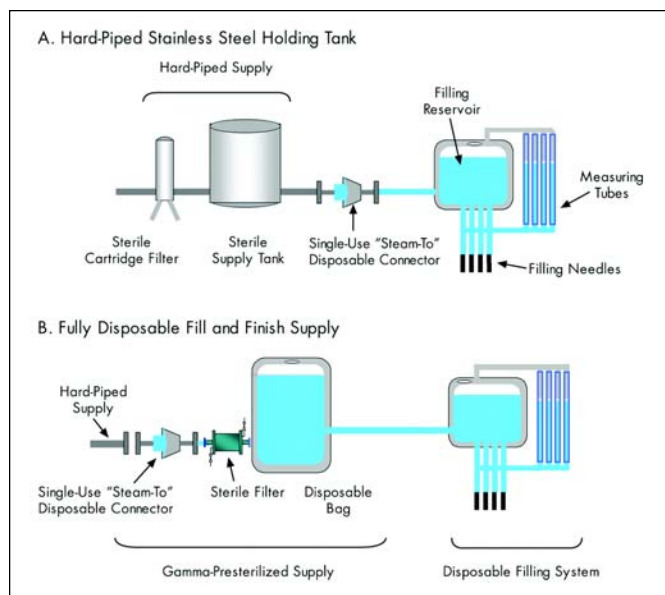
Disposable Filling Technology

Figure 3. Integrating disposable filling into a fill and finish process.

stream and filling machine companies design their equipment to accommodate disposable technologies.

Testing the Performance of a Disposable Filling System

The performance of a disposable filling system is best understood by assessing key process aspects of the technology. These key aspects include:

- range of fill volume and fill rates
- accuracy and repeatability
- system economics compared to conventional filling
- non-aqueous fills and other fill types
- product recovery

Range of Fill Volumes and Fill Rates

In the gravity-driven volumetric fill system, the relationship between fill volume and fill rate determines the choice of supply tubing, dispense tubing, measuring tube and needle diameters. The components must be chosen to meet the desired fill rates of the filling line, taking into account the effects of product viscosity and density as they relate to the solution's ability to flow through the tubing. Variations in other physical properties such as those found in emulsions and suspensions also impact the design of the disposable components. To accommodate a wide range of filling requirements, a selection of different reservoirs, measuring tubes, and needles must be available to meet the needs of the fill and finish operation.

To begin understanding the disposable filling system, the required sizing of the disposable reservoir assembly, and the application range of the current design, researchers developed a mathematical model describing the system dynamics to predict the range of operation for a given set of conditions. An estimate of the number of fills per minute per head can be determined by breaking the disposable filling system into three parts (supply, dispense, and component delay). For a

given fill volume of a known solution, the mathematical model can estimate a fill rate for each combination of reservoir, tubing, and needles.

To illustrate this model, researchers assessed two different theoretical configurations of disposable components for a disposable filling system. The disposable set-ups can be broken up into five distinct sections: reservoir, supply tubes, measurement tubes, dispense tubes, and dispense needles. Both of the example configurations shown in the Figure 4 diagram used the same size reservoir which held approximately 500 ml of solution. The red surface profile incorporated 1/8 inch internal diameter supply and dispense tubes along with a 1/4 inch internal diameter measurement tube. The dispense needle was a straight, five-inch tube of internal diameter 0.062 inches. The green surface used a supply tube, a measurement tube, and dispense tubes having internal diameters of 1/4, 3/8, and 1/4 inches respectively. These tubes were connected to a five-inch straight needle of internal diameter 0.125 inches. A range is shown for the operation of these two different disposable reservoir configurations. The surface profiles define the upper operating limit of each disposable configuration, and stable operation can be achieved at all operating conditions below the surfaces.

The process variables that determine the limits of operation are the solution viscosity and density, reservoir head height, disposable configuration (tubing diameter, needle size, etc.), and solution dispense volume. By setting the reservoir head height and picking a desired fill volume for a solution with a known kinematic viscosity (viscosity/density), the tube set needed for a required fill rate is easily determined. For example, to dispense 4 ml of water at 60 fills/min, tubing set B would be required since tubing set A would not be able to

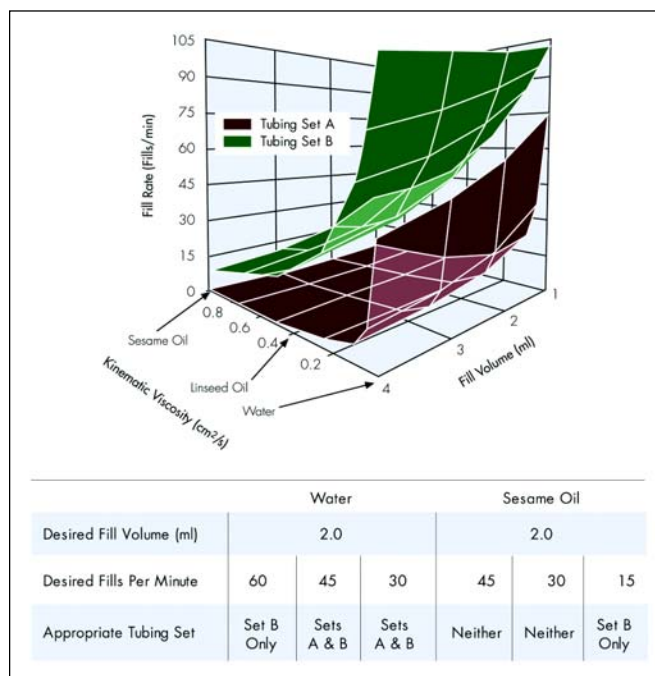|	Water			Sesame Oil		
Desired Fill Volume (ml)	2.0			2.0		
Desired Fills Per Minute	60	45	30	45	30	15
Appropriate Tubing Set	Set B Only	Sets A & B	Sets A & B	Neither	Neither	Set B Only

Figure 4. Example model predictions of fill rates for two different disposable configurations. The table gives examples of acceptable filling speeds and disposable component selection for a 2.0 ml fill derived from the model.

achieve this fill rate. Based on the three dimensional graph in Figure 4, an example of the selection of tubing sets for 2.0 ml oil and aqueous fills at different filling speed requirements is presented in the Figure 4 table.

The gravity-driven, volumetric dispensing mode of operation places some restrictions on the maximum dispense volume and the maximum fill rate. Current designs are limited to fill volumes below 20 ml although the technology may be applicable for larger volumes in the future. For 1.0 ml fills, maximum speeds are of the order of 14,400 fills per hour. However, it is evident from these test data that high fill rates are limited to small fill volumes and low kinematic viscosity solutions.

Other limitations of the technology may be found relating to the physical properties of the drug. Disposable filling has not been implemented to date for suspension fills due to the potential settling of particles in the reservoir. In this case, care must be taken to ensure adequate mixing of the suspension particles for accurate dosage dispensing. In other cases, high foaming characteristics may be problematic due to interference of foam on the light sensors. Also, drugs with high viscous properties will be limited inherently to lower fill rates. It is anticipated that next generation disposable filling system designs will overcome some of these issues.

Accuracy and Repeatability

To test the reproducibility of the disposable system, feasibility tests were performed on different disposable configurations. Samples were collected and weighed over approximately 30 minutes averaging around 1 sample per minute. Samples were taken over a five-hour test period with the first sample of each 30 sample set being taken at the beginning of each hour. After each test was completed, the average and repeatability of all the samples and each individual set of 30 samples was calculated. The highest and lowest values were then used to calculate the maximum positive and negative percent error from the mean. This is referred to as the positive and negative percent error, and the average of these values is the average error.

Figure 5 shows an example of a set of data collected over five hours. The target weight of the samples was 1.025 grams with a difference of 8 milligrams between the high and the low (max = 1.029 g, min = 1.021 g). The average, positive, and negative percent errors were 0.39%, 0.37%, and 0.41% respectively, and the system was running at 55 fills/min. The error lines at 0.5 and 1.0% deviation from the mean are shown on the figure for the reader's convenience.

Other feasibility testing using water has shown that a range of fill volumes between 0.5 and 15 ml is obtainable with the current disposable reservoir assemblies. Ongoing work with larger fill volumes will establish the performance characteristics and component sizing requirements of fills up to 100 ml. The current maximum fill rates have been found to be between 45 and 60 fills/min for small volumes, such as 0.5 and 1.0, and between 15 and 30 fills/min for volumes in the range of 2 to 15 ml. For all of these cases, the repeatability was found to be less than +/- 0.5%.
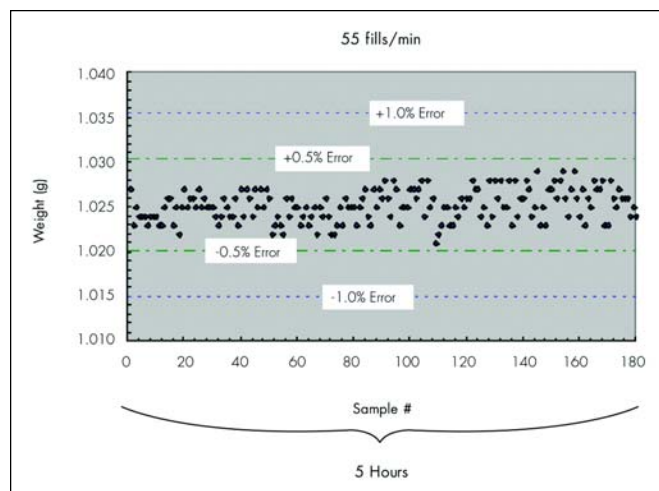

Figure 5. Reproducibility of the disposable filling system operating at 55 fills per minute with an average dispense weight of 1.025 g.

Non-Aqueous and Other Fill Types

To understand the performance on oil based fills, sesame oil was tested at three different dispense weights over 1.5 hours (60 samples). A different disposable assembly was tested than was previously used for water and the results are given in Table A. As can be seen on the chart, with a fixed tubing set, the fill rates varied with the size of dispense, but again, the average percent error remained less than or equal to 0.5%. The model predictions of oil fill rates are included on the table. As is evident from Table A, the model accurately predicts the fill rates for a given component assembly as long as the fluid characteristics are known.

It is expected that disposable systems will be applicable for use with suspension fills and nitrogen topped fills. For suspension fills, other procedures will be required to ensure mixing of the product during the filling operation. Oxygen sensitive products may be protected with a nitrogen blanket on the storage reservoir.

Economic Review of Offsetting Costs

Disposable technology as a filling option is no more costly than traditional filling. There are many examples of drug filling operations that would benefit from a disposable filling option. Most notably, these include toxic, biohazard materials, and multi-product facilities. For example, the ability to dispose of a used filling system rather than exposing operators to potentially hazardous material is invaluable. The same applies to a multi-product facility where the ability to speedily turn around a filling line and not manage onerous line segregation practices would be very beneficial. In these situations, the value of a disposable system cannot be counted in terms of simple comparative economics. Yet, potential users are interested in understanding the cost of ownership. A preliminary comparison of the disposable filling system and a conventional filling system results in essentially comparable running costs. Full Cost-Of-Goods (COG) modeling is required to accurately estimate the complete economic benefits.

Disposable Filling Technology

	1.2 g Dispense		2.0 g Dispense		4.0 g Dispense	
	Head #2	Head #4	Head #2	Head #4	Head #2	Head #4
Average Weight (g)	1.195	1.199	2.055	2.036	4.028	4.040
Fills/Min	39		23		12	
Model Prediction Fills/Min	37		24		13	
Variation						
Weight (mg)	12	8	9	8	16	9
Avg. % Error	0.50	0.33	0.22	0.20	0.20	0.11
Pos. % Error	0.46	0.35	0.21	0.17	0.26	0.09
Neg. % Error	0.55	0.32	0.23	0.22	0.14	0.13

Table A. Fill rates and error measurements for sesame oil at three different dispense volumes.

The preliminary COG evaluation, shown in Table B, illustrates a simple cost trade off that centers around the expendable cost of the disposable system with utilities, labor, and maintenance costs for the CIP/SIP filling system. The total labor required for a conventional system, including autoclaving, system set-up, disassembly, cleaning and rinsing, is estimated to be as much as six times greater than that of a disposable system. The COG calculations for a single shift, 1.0 cc filling system operating at a rate of 14,400 fills/ hour for 225 days per year shows that the running cost of disposable filling is 8% less than the running cost for a traditional filler. Extending this analysis to include depreciation for the incremental capital cost of the disposable system hardware leads us to the conclusion that traditional and disposable filling systems have very similar cost of ownership models. In preparing for a new facility, the disposable option would potentially allow for considerable capital cost savings such as smaller autoclave, clean steam or WFI capacity, and

Operating Cost Summary		
	Conventional	Disposable System
Operating and Maintenance Costs	($/yr)	($/yr)
Set-up Labor	48,125	7,975
Utilities	6,250	438
Parts	15,400	1,000
Maintenance	3,000	550
Disposable Expendables	0	56,250
Total O&M	72,775	66,213
Capital	30,000	80,000
Total Cost of Operation (7 year amortization)	77,061	77,641
Additional Capital Requirements Autoclave and WFI System Capital	60,000	0
Total Cost of Operation (7 year amortization)	85,632	77,641

Table B. Operating cost of a disposable filling system compared to that of a piston filler.

related space requirements. When the capital costs of autoclaves and WFI equipment are included as part of the amortized operation costs, as would be found in a new plant installation, the disposable option is approximately 9% less than the conventional filling equipment. Savings associated with a reduction in validation effort, less plant downtime, and increased speed to market are not calculated here. These factors are expected to be significant and will favor the use of disposable filling equipment - *Table B*.

A downside to this technology can be found in the disposal of the component reservoirs, filling needles, and associated tubing. The used components require disposal in a safe mode in accordance with local and federal regulatory waste guidelines. For some biohazards or toxic materials, incineration or other special handling may be required. When evaluating disposable filling, the landfill or incineration issues need to be compared with the disposal of liquid wastes from rinsing and cleaning.

Product Recovery

The high recovery of product at the end of the process is essential for maximum yield in the filling operation. The disposable filling system is capable of virtually complete drainage of the buffer reservoir. Losses are expected to approximate those typically found in conventional dispensing systems such as positive displacement pumps or time pressure systems. For the enhanced product recovery of extremely valuable drugs, manual drainage of the tubing is possible, followed by the manual filling of the final vials.

Validation Considerations

Validation requirements for the disposable filling system are expected to be similar to those of conventional fillers.

Fill Reproducibility

Tests conducted to ensure accurate reproducible dosage volume is common to both traditional and disposable systems. As there are no fundamental differences to the use of both systems, this parameter will be validated in exactly the same way by conducting validation fill runs and checking weights.

Aseptic Process Sterility

Tests will be conducted to ensure that the filling process is capable of consistently producing sterile filled vials. This is achieved through multiple media fills of thousands of filled vials required to validate a new filling line and revalidate an existing filling line on a routine basis after significant process changes or sterility problems have been experienced. Media fills test the capability of the whole aseptic process to produce a sterile product. They are not specific to just the filling component and as such would be identical for both traditional and disposable filling technology.

Filling Sterility

For media fills to be successfully executed, the systems must be sterile prior to use. In the case of traditional filling, this requires SIP processes to be validated using thermocouples

and Biological Indicators (BIs). These will be placed throughout the filling system especially in areas where, due to trapped air, condensate or heat sink effects, the system may have difficulty reaching the required sterilization temperatures. In the case of filling systems that are disassembled and autoclave sterilized piecemeal, there is the additional challenge of aseptic assembly which requires detailed SOPs and routine operator training. The sterilization challenge with disposable filling is much simpler as the process is one of supplier audit with responsibility for disposable sterility being shifted to the vendor.

Gamma irradiation is used to sterilize a disposable filling assembly that is sealed at both ends. The irradiation process is validated by the use of radiation meters, and BIs . Validation is completed by the vendor according to AAMI standards and certified. The end user needs to audit the vendor for compliance and train operators on unpacking protocols to ensure no sterility breakdown during setup.

Fill System Integrity
In today's filling process, stainless steel piping is connected to filter housings, valves, tees, etc, using clamp connectors and gaskets. Routinely, the filters are tested for integrity, but not the pipe work and connections. By strict comparison to current practice, it would be easy to conclude that similarly the disposable system need not be tested. Currently, this new disposable filling technology is designed, packaged, and manufactured by the vendor to ensure integrity. Again, these processes should be audited by the end user. Although it has not yet been done, it is theoretically feasible to consider implementing a pressure hold test of the filling system post use.

Fill System Cleanliness
The major concerns are particles, cross contamination by multiple drug products, or CIP fluids. In traditional filling, drug cross contamination is managed by using dedicated filling pumps for each product and relying on appropriate line closing and opening protocols and operator training to avoid mix ups. CIP fluid contamination is managed through an onerous CIP validation and routine revalidation. Stainless steel piston pumps are well known for generating steel particles and require routine maintenance to minimize the problem. There are no cross contamination or CIP fluid contamination concerns with the disposable filler. The biggest concern is extractables and TOC. Traditional fillers also require this validation as they typically use several feet of platinum cured silicone tubing to deliver drug solution to the pumps and from the pumps to the vials. In the case of the disposable filler, it is proposed to handle extractable validation in precisely the same way the filter validation is conducted. A vendor will supply a service of product-based extractable testing.

Fill System API Binding
Several pharmaceutical products are known for their sensitivity to contact with stainless steel. This is so prevalent that

glass and ceramic piston pumps are often selected for filling operations. It is expected that the vendor would provide technical support for product specific binding validation.

The validation challenges are clear and more representative of the processes that need to prepare a filling system (e.g. CIP process for cleanliness and sterilization processes for system sterility) rather than concern over filling system performance.

Conclusions
Disposable filling systems can be an economical alternative to conventional filling machines where the benefits of single-use disposability are required. Preliminary data demonstrates that the disposable filling can be run at fill rates and accuracy levels that require the user to make almost no compromises with conventional filler performance. The disposable systems offer improvements in process flexibility and improved ease of use compared to conventional filling systems. The elimination of cleaning, sterilization, and aseptic assembly in the filling suite contributes to a reduction in validation effort and set-up time in the production environment. The systems lend themselves to both single-and multiple-use facilities attempting to meet today's rigid regulatory standards while enhancing the speed of products to the marketplace. For small R&D or clinical fill operations, disposability simplifies the development effort and reduces the amount of time required to validate the filling process. For production fill and finish, benefits are found in quick installation, rapid equipment change-over between filling campaigns, elimination of aseptic assembly, and prevention of cross contamination.

Disposable filling technologies are particularly useful for plants requiring a high degree of safety in their filling operations. Cytotoxic and biohazardous fills requiring the complete containment of product for operator safety and prevention of cross contamination between fills are often difficult in the filling suite. The ability to completely dispose of all product contact surfaces provides an extra level of safety not found with reusable dispensing components.

The costs of operating disposable filling systems are shown to be approximately equal to those of conventional filling technologies although the split between direct and indirect costs will change. Additional direct costs associated with single-use expendables will increase, and savings will be demonstrated in indirect costs associated with equipment clean-up, utilities, installation time, and the elimination of autoclave cycles. Significant additional savings can be anticipated when the reduced cost of validation effort, reduction in burdened labor requirements, and potentially faster access to market are included.

The operation of the disposable filling system shows the reproducibility of the process to be comparable to that of conventional dispensing apparatus with typically less than 0.5% variation for 1.0 ml or greater fill volumes. The equipment is demonstrated to have fill accuracies equal to or exceeding those of conventional time-pressure or piston filling lines with no accuracy drift due to pump wear or daily

warm-up. The selection of the disposable tubing and needles requires an understanding of the physical properties of the feed solution and a matching of these components to the solution. As the filling speed relies on gravity for its driving force, the proper selection of tubing diameter and needle sizes is needed to meet the accuracy and reproducibility requirements of the pharmaceutical manufacturer.

About the Authors

Brett Belongia is a Senior Application Engineer for disposable product technology in the Biopharmaceutical Division of Millipore Corporation. He evaluates new technologies and their development into product lines. Belongia's primary focus is to investigate and support new disposable technology applications. Previously, he investigated and developed novel purification processes as part of the new business development group of the Microelectronics Division of Millipore Corporation. Belongia holds a BS in chemical engineering from the University of Wisconsin and a PhD in chemical engineering from the University of Arizona. He presents posters and presentations at a number of industry conferences and has co-authored several referenced papers. Belongia has had two provisional patents filed for acceptance with others under review.

Robert Blanck is a Market Manager for the Biopharmaceutical Division of Millipore Corporation. With a primary focus on strategic market analysis, he plans worldwide activities for classical pharmaceuticals, vaccines, and plasma products. Blanck also covers disposable filling technologies and clarification activities related to biopharmaceuticals. Previously, Blanck held product management roles within Millipore and has been involved in tangential flow filtration processes for more than 20 years. Blanck holds a BA in biology from Franklin and Marshall College in Lancaster, PA and an MS from the State University of New York at Syracuse.

Steve Tingley is the Director of Biopharmaceutical Marketing for Millipore Corporation. He specializes in final fill sterile filtration validation and integrity testing applications. Having joined Millipore Europe in 1985, Tingley transferred to the United States operations in 1992. As part of Millipore's technology and strategic development teams, he has authored and presented many papers on sterile filtration applications. Prior to Millipore, Tingley spent seven years in drug development with Glaxo and SKB. Tingley holds a degree in pharmacology from the University of Leeds, England.

Millipore Corporation, 290 Concord Rd., Billerica, MA 01821.

This article describes the use of Computational Fluid Dynamics to investigate several options for creating homogeneous horizontally directed airflow in a buffer consisting of multiple shelves.

Numerical Simulations to Assess Airflow Behavior in Buffers

by Stefan Barp, Jos Corver, Dr. Alois Schaelin, and Dr. Paul Stewart

Introduction

Laminar airflow is widely used to protect pharmaceutical clean areas from ingress of particles. In small contained volumes, such as storage cabinets or buffers, the air is passed multiple times through HEPA filters to guarantee the required specification for particle content. Airflow systems designed for this purpose are called laminar airflow units and generate a unidirectional down-flow. In a buffer consisting of multiple shelves; however, it is not possible to achieve down-flow, and horizontally directed airflow has to be applied.

Since shelves create a number of obstacles, the flow is unlikely to be strictly laminar. However, it is essential that the velocity over stored vials is as homogeneous as possible, such that the air is equally filtered at each level and the same quality of air is guaranteed for the entire height of the buffer.

The design of airflow systems can be a tedious job of trial and error when tuning the airflow. The Computational Fluid Dynamics (CFD) method of numerical analysis is a valuable tool for the simulation of such non-laminar flows and may be applied to buffer design for the investigation of critical regions. CFD results in a 3D-vector field of airflow and post-processing software enables the visualization of the results as contour plots and particle trajectories. Further, by showing details of the particle trajectories, critical regions can be inspected in particular detail. The conclusions drawn from the results may be used in design recommendations.

This article describes the use of CFD to investigate several options for creating homogeneous horizontally directed airflow in a buffer consisting of multiple shelves. The results are intended to assist in the determination of optimal flow conditions in laminar airflow units.

Method
Description of the Situation
Buffers with shelves of widths ranging from 1.07m to 1.8m (3.5 ft to 5.9 ft) were modeled in the study - *Figure 1*. An airflow of 1.715 kg/s (60.5 oz/s) was used,

Figure 1. Outline of the geometry. The inlet channel is on the left side. The red planes indicate the perforated plates that are used to adjust the air distribution. On the front, the so-called 'wheel box' is outlined.
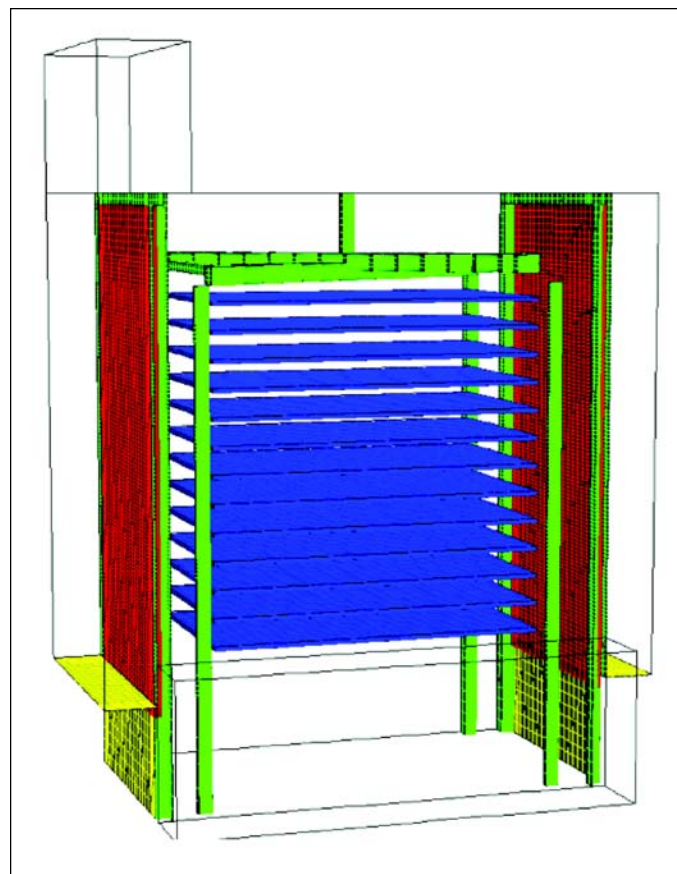
Case	Number of Loaded Shelves	Bottom stack	Porosities [%]	Pressure Loss Coefficient	Remarks
1	7	Yes	6, 4, 2.8, 1.9	550, 950, 1500, 2200	Porosity of original buffer
2	7	Yes		550, 591 608, 617	Optimized pressure loss coefficient distribution
3	7	Yes	6	550	Tapered walls of inlet and outlet channel
4	7	Yes	6	550	Vertical walls, constant coefficient
5	6 ½	Yes	6	550	Homogeneous perforation
6	13	No	6	550	Completely filled buffer

Table A. Overview of the simulation program.

which led to an average velocity of 0.46 m/s (90.6 fpm) after the perforated plate. The air-ducts extended only to the region where the shelves were located.

An opening at the unloading side was not considered and the loading door was closed. The structure also comprised a pushing system for automatic unloading of the shelves. This system consisted of a pusher that was actuated by pusher belts wound around wheels. The region where the wheels were located is referred to as the 'wheel-box.' It was not intended to actively have air pushed through this region and the consequences are considered in this article.

The buffer contained 13 shelves, each of which could be loaded with vials. The shelf-stack hung on a structure that could be moved upward and downward, such that each shelf could

be brought to the same level for loading/unloading. The shelves hung on rods suspended from adjacent shelves. Vertical guides kept the shelf-stack positioned in the horizontal plane. The buffer considered was designed to accommodate two vial packs per shelf and one case described in this article is devoted to the situation where only half of a shelf is loaded.

Three different geometries/loading stages were investigated:

- Geometry A: 7 shelves loaded, 6 shelves in the compressed bottom stack: the reference case

- Geometry B: 7 shelves loaded, 6 shelves in the compressed bottom stack. Walls of inlet and outlet channel tapered. The purpose of this case is to optimize the pressure situation in the ducts.

- Geometry C: 6 shelves loaded, 7th shelf loaded only on the right half, 6 shelves in the compressed bottom stack. This case simulates a potential worst-case situation during the loading of a particular shelf.

Although additional cases had been simulated, it became apparent those described provided the most relevant information since they account for buffers with both empty and filled regions. Therefore, these cases were chosen for consideration in this article.

Smoke Tests
In practice, smoke tests were used to visualize airflow. This 'smoke' actually consists of small particles that are transported by surrounding air. If those particles are small and have very little mass, they do not influence the airflow. Care must be taken to assure that the
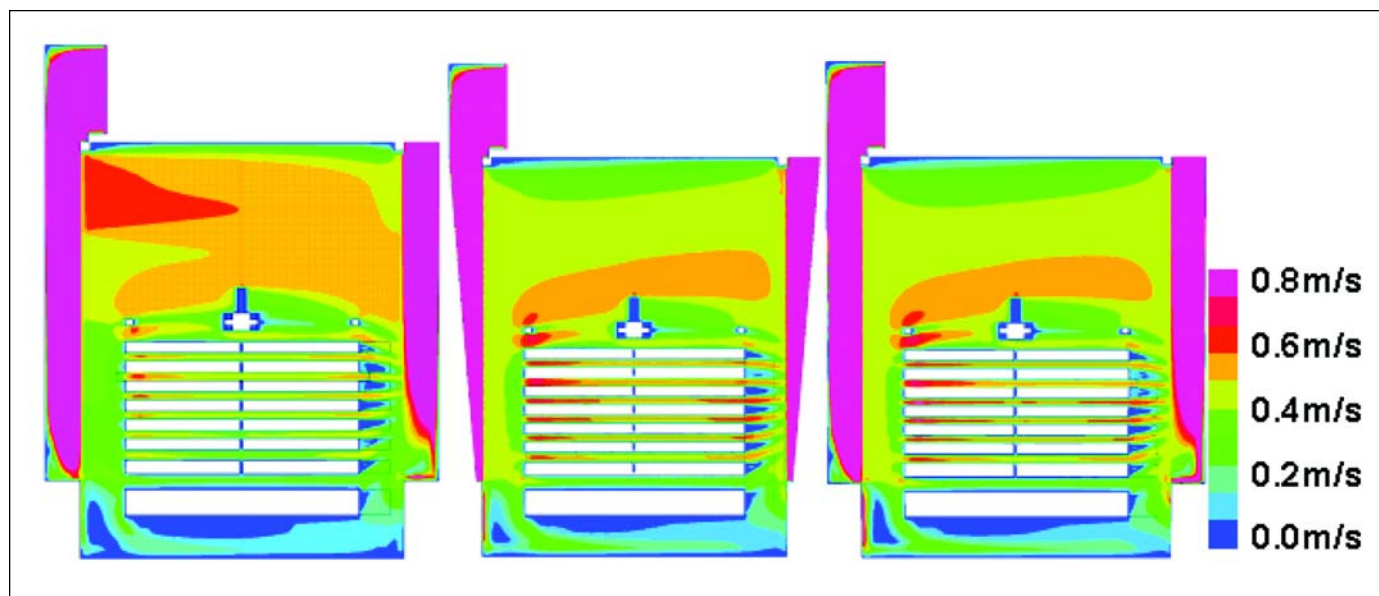

Figure 2. Homogeneous airflow distribution over the shelves: three alternatives. From left to right: optimization of porosity of inlet and outlet plates, utilization of tapered channels and homogeneous porosity of the inlet plates such that airflow impedance is the limiting factor.

injection of the 'smoke' does not influence the airflow.

Post processing capabilities of CFD software facilitate the simulation of smoke tests. The release of the particles from a line is simulated and it is assumed that the distribution of the particles is uniform. The paths of the particles are followed for 60 seconds. Visualizations show the path that the particles followed during the 60 seconds. The simulation is performed twice, releasing the particles along perpendicular lines.

The simulation of smoke tests may be used to support the validation of laminar flow units.

Variations

The inlet and outlet channels were separated from the buffer by perforated plates. The basic porosities (fraction open-to-total area) used in the calculations were from the original buffer. A pressure loss coefficient was determined from the porosities and was used in the simulation.

The porosity-pressure-loss coefficient correlation was taken from Wagner.[1]

The starting case related to a practical situation where the perforated plates varied from top to bottom, as a result of trial-and-error optimization. A few iterations were needed to represent the practical situation. Several methods were used to optimize the airflow situation, including optimization of the porosity distribution and optimization of the ducts surrounding the perforated plates.

Table A provides an overview of several calculated cases, from which the three situations discussed in this article were taken. The porosities are from bottom-to-top on the inlet side and from top-to-bottom on the outlet side.

Fundamentals of Computational Fluid Dynamics

In Computational Fluid Dynamics (CFD), the whole domain under consideration is divided into a large number of cells (typically 500,000 or more). For each cell, the transport equations for mass, momentum, energy (in cases with heat sources), turbulence quantities, and additional potential contaminant species, are set up and solved in a sophisticated iterative procedure. In the case of complex airflow patterns, care must be taken to ensure grid-independent solutions.

For the solution of the transport equations, commercial software was used with a body-fitted curvilinear structured multi-block grid. For accurate results, schemes with second-order discretion were used and multi-grid acceleration techniques were implemented to achieve fast convergence.

Turbulence Modeling

The application of CFD to ventilation began about 20 years ago and continues to prove successful for an escalating number of applications. The final principal problem relates to the turbulent character of airflow. A complete simulation of the full turbulent structures would require a model with sub-millimeter-scale flow structures, and as a general rule, this is not practical for current and projected computer technologies. Therefore, a large number of models have been developed to assess the turbulent character of flow.

Most turbulent flow models belong to a class known as the Reynolds Averaged Navier-Stokes (RANS) equations. The fundamental equations for the instantaneous fluctuating velocity components (u, v, w) are called Navier-Stokes equations. In a RANS model, these components are divided into averaged values, for which similar transport equations are solved, and fluctuating components (u', v', w' with an average value of zero), for which an additional model is set up.

Of the available models, the k-ε-model is a first-order RANS model, which requires a reasonable computational effort, and also is the most frequently used for flow problems in industrial applications. In the k-ε-model (after Launder and Spalding, 1974) two additional transport equations are solved for the two turbulence quantities k and ε (k is the isotropic turbulent kinetic energy (k = ½ (u'^2 + v'^2 + w'^2)) and ε is the viscous dissipation rate of k into heat, via molecular friction).

Other more advanced turbulence models include further terms or equations to correct for special flow situation effects that are not predicted effectively by simple models such as the k-ε-model.

An advance in accuracy would be the use of a second-order RANS model, where transport equations for second-order correlations of turbulence quantities, such as u'^2, v'^2, and $u'v'$, are set up and solved, or modeled separately.

In the derivation of all the RANS models, higher order terms appear in the equations for which additional modeling has to be developed. This is called the 'closure problem' of turbulence modeling.

At present, these advanced models continue to lack general applicability and the choice of the k-ε-model is still considered the most appropriate for general flow problems, such as those in a cleanroom.

Computational Effort Required

The computation time per case with 500,000 cells is in the region of 12–24 hours on a WindowsNT® workstation with a 1 GHz CPU. The memory requirement is somewhat more than 1 KB per cell, giving a total requirement of approximately 550 Mbytes for all 500,000 cells. For investigation of a specific situation, a parameter variation with 5 to 15 different cases usually has to be performed.

Geometrical Modeling and Boundary Conditions

Though still detailed enough to account for the influences on the airflow, the laminar flow units and flow obstacles (e.g., shelving) were modeled in a relatively simplified way. The air inlet and exhausts were modeled consistent with their design flow rates. In this environment, temperature effects are considered negligible, allowing the simulations to be performed under isothermal conditions.

Results and Discussion
Homogeneous Velocity Field

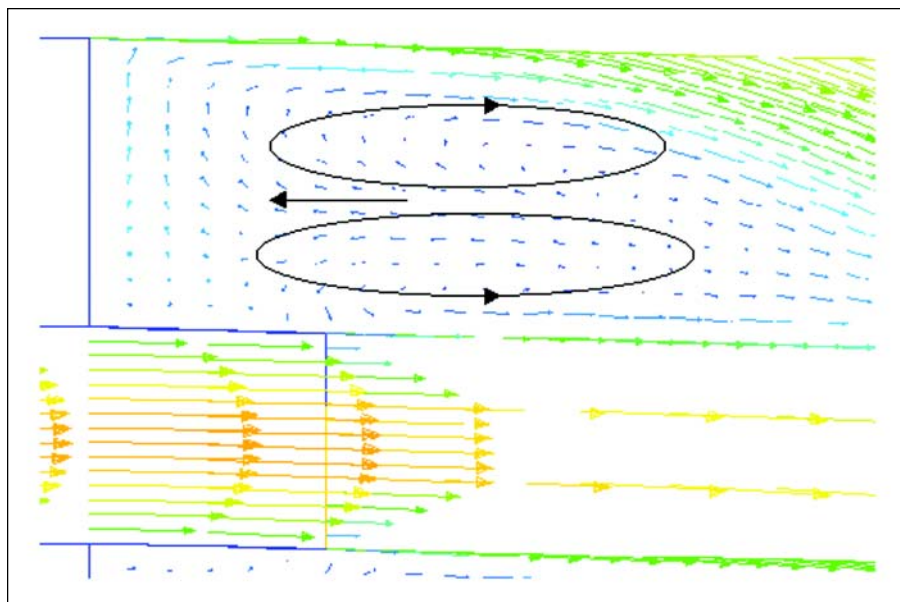Optimization porosity gradient: assum-

Figure 3. Airflow close after the shelves. In the region with very low velocities recirculating vortices appear.
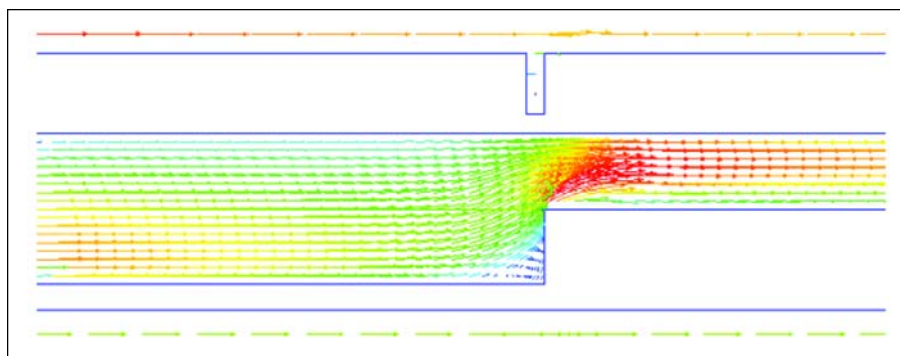

Figure 4. Airflow near vial front. The air accelerates and entraps air from the vial bottom region.

ing that it is essential to start with a homogeneous velocity field immediately after the air inlet plates, this was taken as the starting point. After the conclusion that the situation in the top part of the buffer was not optimal and the velocity between the shelves rather limited, other design aspects were taken into consideration.

Inclined walls: under the assumption that a homogeneous pressure gradient can be obtained from the shape of the inlet and outlet channels, the inclined wall case was calculated.

Uniform porosity: the inclined walls case also was based upon the application of a constant pressure loss on the perforated plates. The uniform porosity case was calculated to verify the contribution of the inclined walls of the ducts.

Details of Airflow and Particle Trajectories

Direction of velocity vectors and eddies: although it is hardly conceivable that the airflow can be made strictly laminar, it is generally understood that for optimal situations the flow should be unidirectional. Therefore, it is important to have a particular look in regions with discontinuities. The study focused on two regions: just behind the shelves and close to the region where the airflow strikes the vial pack.

Particle trajectories: When the velocity field is known, the path of theoretical particles can be traced. Therefore, two lines were determined where those theoretical particles would start and the course followed by the theoretical particles was followed for 60 seconds. The plots of the traces give insight of potential contamination.

Homogeneous Velocity Field

Figure 2 illustrates the general view of the flow velocity field. The purple colored regions are located inside the inlet and outlet channels. The starting situation is visualized in the left part of the figure.

- The porosity-distribution of the inlet plate was optimized to create a homogeneous airflow across the shelves.

- A depletion effect in the inlet and outlet channels is visible. This had no effect in the buffer due to the relative high pressure-loss over the perforated plates.

- The illustration also shows that in the region above the shelves the airflow velocity is quite high with velocity values of the order of 0.7 m/s.

- The vertical structures inside the buffer influenced the field and were the cause of the local velocity increase between the shelves, shortly after the entrance-region.

- In the region above the shelves, the airflow velocity was higher than elsewhere in the buffer and demonstrated some inefficiency.

- The middle part of Figure 2 illustrates the situation with inclined ducts. The porosity of the plates was uniform.

- The airflow in the region above the shelves was reduced, but between the shelves a slight increase was discernable and the velocity was above 0.4 m/s (78.7 fpm).

Implementing a homogeneous pressure loss over the inlet plates can create a very similar situation. The porosity that was chosen was 6%. The pressure loss coefficient appeared to be high enough (550) to ensure homogeneous inlet airflow. This is illustrated in the righthand part of Figure 2.
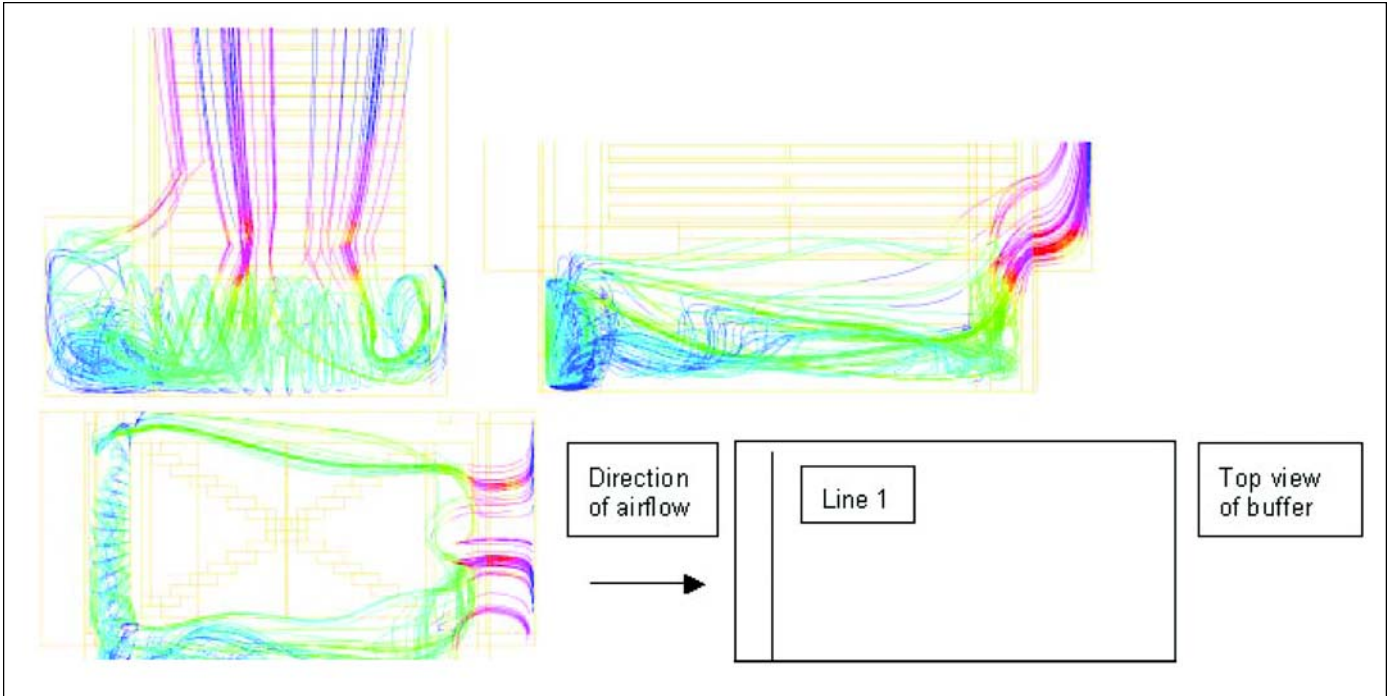
Figure 5. Streamlines, starting at a line on the floor across the buffer (line 1).

Discussion
on Plate Optimization

The distribution of the horizontal inlet velocity (v_{in}) through the perforated plates into the buffer depends on the chosen porosities of the plates and on the geometry of the inlet channel.

By choosing a high-pressure loss coefficient for the perforated plate, the inlet velocity becomes independent of the local Vertical velocity at a given height in the channel and also is independent of the channel geometry. Hence, a high constant pressure loss coefficient leads to a relatively constant velocity over the entire channel height. This is shown in the right-hand part of Figure 2.

The air enters the channel at the top with the inlet velocity, v_0, and the static pressure, p_0. Using Bernoulli's law, this leads to a static pressure at the vertical distance, z, from the entrance of:

$$p(z) = p_0 + v_0^2 \frac{\rho}{2} - v(z)^2 \frac{\rho}{2} \qquad (1)$$

From this static pressure, the inlet velocity, v_{in}, through the perforated plate can be calculated as:

$$p(z) = v_{in}(z) \frac{\rho}{2} \zeta(z) \qquad (2)$$

With a constant pressure loss coefficient, $\zeta(z)$, it is possible to achieve a constant inlet velocity, v_{in}, over the complete channel height, by choosing the channel geometry such that static pressure over the complete channel height, p(z/H), is kept constant. This requires from (1) that the vertical velocity at height z, v(z), is kept constant over the height.

To get a constant vertical velocity at height z, v(z), with a constant velocity, v_{in}, over the height, the continuity equation requires that the free area of the channel is reduced linearly. In this way, the total area where the air can leave any volume element in the channel is constant. This was achieved by inclining the back wall of the inlet channel as shown in the center of Figure 2.

In a channel with constant free area, a general formula for velocity at height z, $v_{in}(z)$, can be obtained from (1) and (2):

$$v_{in}(z) = \sqrt{\frac{1}{\zeta(z)}\left(\frac{2}{\rho} p_0 + v_0^2 - v(z)^2\right)} =$$

$$\sqrt{v_{in0}^2 + \frac{v_0^2 - v(z)^2}{\zeta(z)}} \qquad (3)$$

By choosing a high pressure loss coefficient for the perforated plate, the inlet velocity becomes independent of the local velocity v(z) in the channel and also independent of the channel geometry. So, a high constant pressure loss coefficient ζ leads to a relatively constant velocity (v_{in}) over the channel height. This is shown in the right-hand part of Figure 2.

p_0 static pressure at the top of the inlet channel

v_0 vertical velocity at the top of the inlet channel

$v(z)$ vertical velocity at height z

z   vertical distance from the top in the inlet channel

v_{in} horizontal velocity through the perforated plates into the buffer

Detailed Topics
Direction of Velocity Vectors

From the contour pictures, it can be seen that there are various regions where the velocity is close to zero. The situation just behind the shelves is illustrated in Figure 3. The velocity is very close to zero, and due to some drag from the main flow region, a small vortex system is induced. In fact, two eddies can be identified. Since there is

no risk of this recirculation entering the region above the product, this situation is not critical. Nevertheless, this does illustrate the fact that areas with no direct forced airflow may develop eddies, and therefore, an uncontrolled mixture of air. This is further illustrated and clarified in the section on particle trajectories.

This buffer case is related to loading the shelf with two vial packs. In Figure 4, a worst-case scenario is illustrated where the first vial pack is loaded in a downstream position. The result displays an obviously undesirable situation where air gets pushed from platform height to vial height.

Both of the above aspects are deviations from the ideal unidirectional flow principle. The region behind the shelf is not a significant problem since there is no risk of particle ingress in the direction of the vials. The vial pack case illustrates that consideration should be given to first loading vials on the upstream side.

Particle Trajectories

Using post-processing facilities, it is possible to visualize the paths that particles carried by the flow are likely to follow under the assumption that the gravitational force is negligible in comparison to convective transport. Therefore, two lines were defined as nominal starting positions for a number of particles. In Figure 5, the particles were released (virtually) from the bottom of the buffer just following the inlet position. The particles were registered for a period of 60 seconds. The coloring of the paths relates to the air (particle) velocity at the respective positions.

Figure 6 illustrates the situation where the line of particles was located at the bottom of the buffer, just at the centerline parallel to the main air velocity direction.

The pictures show that the particles released in a region with little airflow could move along winding paths. The streamlines suggest that the region below the shelves can be a risk. The airflow velocity was very small, but a large eddy causes particles to be carried to higher regions and even between the shelves.

This picture also shows that the region defined as the 'wheel-box' experienced poor flow conditions. The situation without the 'wheel-box' also was simulated; however, the improvement was relatively minor.

The situation can be improved significantly by having forced airflow in the regions below the shelves eliminating regions with relatively low velocities which cause eddies to occur.

Verification

To verify the results, practice tests were performed on a buffer optimized for unidirectional airflow on all shelves. A three-step approach was taken:

1. CFD simulation to determine the outline
2. rough verification in a buffer with modeled shelves
3. verification using the real situation

The measured airflow velocities appeared to be within 20% of the predicted values, and unidirectional airflow was demonstrated between the shelves. Optimizing the air channels on the sides of the main inlet and outlet minimized the undesirable recirculation patterns below the shelves.

Conclusion
Homogeneity of Airflow

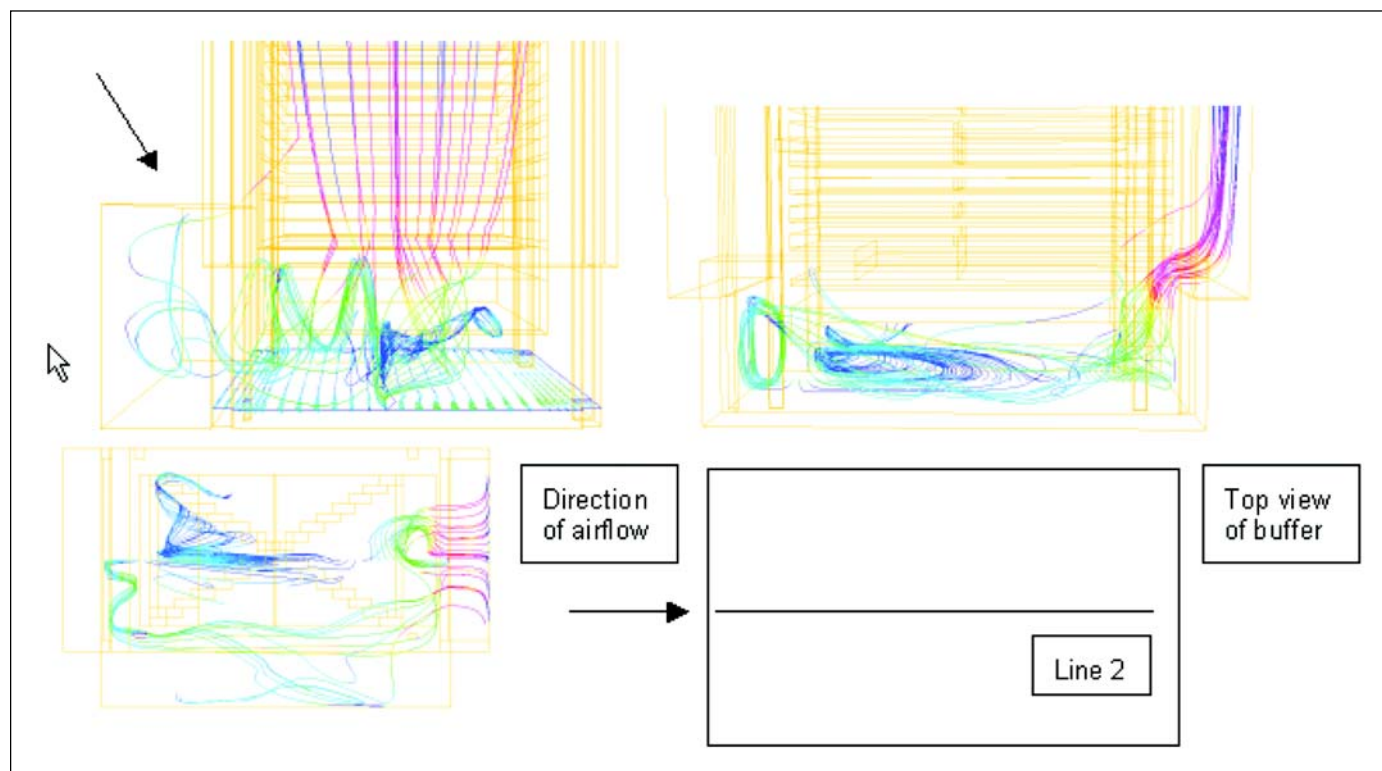When perforated plates are used for

Figure 6. Streamlines, starting at a line on the floor along the buffer (line 2).

final inlet of air in to the buffer, this provides the highest resistance in the ducts. Optimizing the airflow situation by changing the porosity or shape of the ducts will not improve the homogeneity. Conversely, it is important to make sure that all clean regions are supplied with air.

Recirculation Effects

The horizontal airflow does not stay unidirectional in the neighborhood of obstacles. When this occurs there is only a small risk of particle transport in the neighborhood of vial necks downstream. For this reason, it is recommended that the first vial pack is loaded upstream where the airflow is perpendicular to the loading/unloading direction. There is an advantage when the airflow is in the loading direction. However, the disadvantage in this situation is the absence of controlled airflow near the doors.

Particle-Contamination

Regions with stagnant air are likely to gather particles. Although these regions are not flushed by forced airflow, eddying flow is likely to occur. This flow can carry particles to other regions, and the present simulations show that even regions containing vials can be subject to particle contamination.

CFD Method

The CFD method provides a powerful tool for simulation of airflow in pharmaceutical equipment. Insights thus obtained can be used to shorten the design and optimization process. The 3D properties of the flow-field are in particular important to consider.

References

1. Wagner, Walter, Strömungstechnik und Druckverlust-berechnung, Vogel Fachbuch, Kamprath-Reihe, ISBN 3-8023-0449-7, 1992.
2. CFX-TASCflow Version 2.11, AEA Technology, UK, www.aeat.com/cfx
3. Wirth, A., Schälin, A., Computer Simulations of Airflows in Critical Cleanroom Areas. PDA Conference "Regulatory and Technological Challenges for One World," Basel, Switzerland, 2000.
4. Launder, B. E., Spalding, D.B., The Computation of Turbulent Flows. Computer Methods of Applied Mechanical Engineering, 3, 1974.

About the Authors

Stefan Barp studied mechanical engineering at the ETH Zurich (Swiss Federal Institute of Technology) and obtained his degree in 1993. He then joined the Combustion Laboratory at the ETH Zurich where he was responsible for the Computational Fluid Dynamics (CFD) in a development project of a new industrial low NOx burner. In 1995, he moved to the Room Airflow Laboratory at the ETH Zurich Room where he worked in fundamental research of turbulence modeling for buoyant flows. At this time, he also spent seven months at the University of Tokyo. He will finish the project in March 2002 with his PhD thesis. At the end of 1999, he joined Airflow Consulting.

AFC Air Flow Consulting, Langmauerstrasse 109, 8008 Zurich, Switzerland.

Ir. Jos Corver studied applied physics at the University of Delft (Netherlands) and obtained his degree in 1981 in dynamic calibration of a clinical electromagnetic flow meter system. He continued his research on blood flow behavior in the carotid artery to obtain early detection techniques of atherosclerosis. He joined Océ in 1984, a company that produces copier and printer systems. He worked on process development ranging from pure innovation onto introduction into production environments. He was responsible for the technical coordination of projects leading to new copiers and printers in both office and design engineering markets. He joined BOC Edwards Pharmaceutical Systems in 1999 to take responsibility for process, test, and validation including the development of new products.

BOC Edwards Pharmaceutical Systems, Steenstraat 7, 5107 NE Dongen, Netherlands.

Dr. Alois Schaelin has studied physics at the ETH Zurich (Swiss Federal Institute of Technology) and obtained his PhD in 1990 in semiconductor physics and optoelectronics. He joined then the Room Airflow Laboratory at the ETH Zurich and founded his own company Airflow Consulting in 1996. He is working primarily in professional applications of Computational Fluid Dynamics (CFD) for airflow in buildings, industrial ventilation, pharmaceutical applications, and related fields. He is also co-author of several handbooks, R&D, consultant, and visiting guest Professor at the Technical University in Aalborg, Denmark.

AFC Air Flow Consulting, Langmauerstrasse 109, 8008 Zurich, Switzerland.

Dr. Paul Stewart has spent the last 12 years working in the pharmaceutical industry for BOC Edwards Pharmaceutical Systems and has held senior roles in operations and business management in Europe and the US. He is currently responsible for technology and has been instrumental in developments in freeze drying and filling systems products, including several generations of lyophilizer automated loading systems. Stewart was born in England and studied at the Universities of Sunderland and Cambridge culminating in BSc and PhD degrees in mechanical engineering. During his early career, he specialized in computerization and automation in the manufacturing industry before moving into a position in manufacturing management within the BOC Group.

BOC Edwards Pharmaceutical Systems, 2175 Military Rd., Tonawanda, NY 14150.