This page intentionally left blank.
Streamlining Data Management in Drug Product Commercialization and Manufacturing

by

Spencer C. Anderson

Submitted to the MIT Sloan School of Management and the MIT Department of Mechanical Engineering on May 9, 2014 in Partial Fulfillment of the Requirements for the Degrees of Master of Business Administration and Master of Science in Mechanical Engineering

Abstract

Effective execution and alignment of data management across development and manufacturing teams is essential for Amgen’s Drug Product Technology group to realize its main goals of shortening the development timeline and ensuring robust commercial manufacturing. The right data management strategy can help address these goals by accelerating development work and regulatory filing as well as improving commercial manufacturing efficiency. In the face of challenges associated with rapid growth and an expanding product pipeline, Amgen’s commitment to standardizing development work and digitizing both clinical and commercial manufacturing has introduced many opportunities for new data management initiatives, improvements, and a revamped overall data management strategy.

We identify a framework for the development of a data management strategy for the Drug Product Technology group to enable greater efficiency and alignment across development and manufacturing teams. The primary steps in data management and objectives at each step were determined. While a full data management strategy has been recommended to the Drug Product Technology group as a set of current and future projects, this thesis focuses on three specific case study projects within the overall strategy: (1) data generation and collection in drug product manufacturing, (2) real-time multivariate statistical process monitoring of lyophilization in clinical manufacturing, and (3) integration of development study data through electronic lab notebook software. Based on insights from these case studies, we make specific recommendations for further improvements in data management.

Thesis Supervisor: Roy E. Welsch
Title: Professor of Management Science, MIT Sloan School of Management

Thesis Supervisor: Daniel D. Frey
Title: Professor of Mechanical Engineering, Mechanical Engineering
This page intentionally left blank.
Acknowledgments

I would like to give my sincerest thanks to the Leaders for Global Operations program, my classmates, and everyone at Amgen who supported me in the completion of this thesis. First and foremost, I owe much to my advisors, Roy Welsch and Dan Frey, and my Amgen mentors, Wenchang Ji, Narinder Singh, and Ed Walls, whose knowledge, passion, and guidance were both inspiring and invaluable to the success of the internship. Secondly, thank you to the amazing LGO alumni at Amgen for their constant support and advice. Chris Garvin, Matt Kasenga, Amy Lee, Marie Wolbert, and Seth White, you were always there when I needed encouragement and your friendship made for an incredible internship experience.

Lastly, I thank my family. They made me who I am today and I owe everything to them and their support. There is nothing more important in my life.
This page intentionally left blank.
Table of Contents

Abstract ................................................................................................................................. 3
Acknowledgments .................................................................................................................. 5
Table of Contents .................................................................................................................. 7
List of Figures .......................................................................................................................... 8
List of Tables ............................................................................................................................ 8

1 Introduction ......................................................................................................................... 10
   1.1 Project Motivation ......................................................................................................... 10
   1.2 Problem Statement ......................................................................................................... 11

2 Background and Literature Review .................................................................................... 12
   2.1 Biotechnology Industry ................................................................................................. 12
   2.2 Amgen, Inc .................................................................................................................... 13
      2.2.1 Drug Product Commercialization and Manufacturing Network ......................... 14
      2.2.2 The Drug Product Technology Group ................................................................. 16
      2.2.3 Overview of Data Management ........................................................................... 16

3 Project Approach ................................................................................................................. 22
   3.1 The Data Management Cycle Framework ..................................................................... 23
   3.2 Strategy Development ................................................................................................... 23
      3.2.1 Data Generation ..................................................................................................... 24
      3.2.2 Data Analysis ......................................................................................................... 25
      3.2.3 Integration ................................................................................................................ 26
      3.2.4 Information Access / Connectivity ........................................................................ 26
   3.3 Strategy Project Landscape and Case Studies ............................................................... 27

4 Analysis and Results ............................................................................................................. 29
   4.1 Generation - Drug Product Manufacturing Data Collection - Case Study #1 .......... 29
      4.1.1 Data Collection Technical Assessment ................................................................... 29
      4.1.2 Recommendations ................................................................................................. 33
   4.2 Analysis - RT-MSPM - Case Study #2 ........................................................................ 34
      4.2.1 Lyophilization Overview ........................................................................................ 34
      4.2.2 Lyophilization Equipment Preparation and the Sterilization Cycle ..................... 36
      4.2.3 Real-Time Multivariate Statistical Process Monitoring ........................................ 38
      4.2.4 Application ............................................................................................................. 41
      4.2.5 Results .................................................................................................................... 44
      4.2.6 Recommendations ................................................................................................. 50
   4.3 Integration – ELN BioBook – Case Study #3 ............................................................... 52
      4.3.1 Photosensitivity Study ............................................................................................. 54
      4.3.2 Study Standardization ............................................................................................. 56
      4.3.3 Photosensitivity ELN BioBook Template ............................................................... 57
      4.3.4 Assay Results and Automated Graph .................................................................... 59
      4.3.5 Recommendations ................................................................................................. 61
   4.4 Information Access / Connectivity - Remote Monitoring ............................................. 63

5 Summary of Recommendations and Conclusion ............................................................... 65
List of Figures
Figure 1: Amgen's Drug Product Commercialization and Manufacturing Network.. 15
Figure 2: Matrix of IT Decisions and Structures (Weill and Ross, 2004)............... 21
Figure 3: Organizational Objectives of the Data Management Cycle....................... 24
Figure 4: Drug Product Commercialization and Manufacturing Responsibilities...... 27
Figure 5: Flow Chart of Drug Product Manufacturing Unit Operations..................... 31
Figure 6: General Temperature Settings and Product Behavior during the Lyophilization Cycle .................................................................................................................. 35
Figure 7: Water Phase Change Diagram Overlaid with Lyophilization Cycle Steps... 36
Figure 8: Temperatures and Pressures throughout the Sterilization Cycle ............ Error! Bookmark not defined.
Figure 9: A Control Ellipse for Two Dependent Variables (Montgomery, 2009)...... 40
Figure 10: Hotelling T² Control Chart for Purge and Heat Up Phase (Screen Grab from Simca-P+)........................................................................................................ 43
Figure 11: Cooling Phase Excursion in Test Case #1 (Simca-P+ Screen Grab)......... 45
Figure 12: Variable Contribution Plot for Test Case #1 Excursion ......................... 45
Figure 13: Variable Analysis of "Equipment A" Power Output for Test Case #1 ...... 46
Figure 14: Purge & Heat Up Phase Excursion in Test Case #2 (Simca-P+ Screen Grab) ....................................................................................................................... 47
Figure 15: Variable Contribution Plot for Test Case #2 Excursion ......................... 48
Figure 16: Variable Analysis of "Equipment D" Temperature for Test Case #2....... 49
Figure 17: Final Evacuation Phase in Test Case #3 (Simca-P+ Screen Grab)......... 50
Figure 18: Flow Diagram for Drug Product Evaluation in the Photosensitivity Study ....................................................................................................................... 55
Figure 19: Screenshot of Photosensitivity Template in IDBS ELN............................... 58
Figure 20: Screenshot of Photosensitivity “Assay Results BioBook Spreadsheet” in IDBS ELN................................................................. 59
Figure 21: Screenshot of Assay Results Graph in IDBS ELN...................................... 60

List of Tables
Table 1: Sample of Recommended Process Monitoring Parameters during Buffer Preparation........................................................................................................... 32
Table 2: Prioritization of RT-MPSM Model Implementation.................................... 51
Table 3: Prioritization of Development Studies for ELN BioBook Roll-Out............. 62
This page intentionally left blank.
1 Introduction

1.1 Project Motivation

While Amgen continues to grow and increase revenues year over year (Amgen, 2014), a few major challenges lie in the years ahead. First, several of Amgen's current blockbuster drugs are nearing the end of their patent life. Secondly, as more and more companies improve their drug discovery capabilities, competition to be the first to bring a new drug to market has increased. Both of these challenges have pushed Amgen to actively expand its product pipeline through acquisitions and broader in-house R&D. As of early 2014, Amgen has 19 products in Phase 1, 13 products in Phase 2, and 14 products in Phase 3 (Amgen, 2013). The current focus is on being the first to develop and commercialize these drug candidates, while continuing to emphasize safety and Amgen's mission of "serving every patient every time".

Accordingly, Amgen's Drug Product Technology group has adopted two main goals: shortening development timeline and ensuring robust commercial manufacturing. While the pipeline expansion increases the chances of commercializing the next blockbuster drug, it also adds complexity to an already complicated development and commercialization process. As outlined by Garvin (2012), better alignment across drug product development and manufacturing groups is crucial to realizing these goals. An effective data management strategy across drug product development, commercialization, and commercial
manufacturing groups is a key piece of this overall alignment, and is the focus of this thesis.

Improving the efficiency and alignment of data management across development and manufacturing teams can help address these goals by accelerating development work and regulatory filing as well as improving commercial manufacturing efficiency. In the past, development studies, regulatory filings, and clinical runs have been performed on a case by case basis and powered by manual data collection and analysis. On the commercial side, although some parameters have been captured electronically for some time, many important process parameters are still being collected manually and stored in paper batch records that can only be accessed for retrospective analysis. Now, Amgen is committed to standardizing development work and digitizing both clinical and commercial manufacturing. These actions have enabled many new data management initiatives that can streamline the development process, regulatory filing generation, and commercial manufacturing process monitoring.

1.2 Problem Statement

Historically, the approach to data management in Amgen’s drug product manufacturing and development network has been largely inconsistent from product to product, site to site, and team to team. The data required for regulatory approval and process understanding has been determined partly on a case by case basis, analysis for common studies has been time consuming and varies widely depending on the scientist performing the analyses, data is stored in many different
systems and often in paper form, and gaining access to relevant data can take several days to weeks. A comprehensive data management strategy is needed to create alignment across teams in the drug product development and commercialization organization and promote a more streamlined approach to data collection, data analysis, data integration, and information access. This project outlines a potential data management strategy in both development and commercial manufacturing to achieve these benefits.

2 Background and Literature Review

2.1 Biotechnology Industry

While biotechnology in general encompasses any use of living organisms, biological process, or biological systems to create a product, the healthcare biotechnology industry specifically focuses on the production of antibodies, recombinant therapeutic proteins, and vaccines. These products are developed from living organisms (as opposed to chemical compounds that make up traditional pharmaceutical drugs) and are designed to target specific cells associated with various medical conditions. “In contrast to most drugs that are chemically synthesized and their structure is known, most biologics are complex mixtures that are not easily identified or characterized” (FDA, 2009).

The first non-vaccine biologic, a biosynthetic insulin developed by Genentech, was introduced in 1982. Since then the industry has experienced
incredible growth, posting $262 billion in revenue in 2013 and 11.0% growth from 2008 to 2013 (IBISWorld, 2013).

Due to the difficulties of successfully carrying a biologic drug candidate through clinical trials and to the market, it is estimated that 90% of the money spent researching new treatments is spent on failures, meaning that biotechnology companies can spend as much as $2 billion for each successful drug (Biotechnology, 2014). In this environment, competition can be fierce, especially when public research findings highlight an opportunity to develop and commercialize a new blockbuster drug with a greater chance of success. Plunkett (2013) presents one current instance of such competition:

A new blockbuster is on the horizon that will dramatically lower cholesterol levels. Based on recent findings of a rare mutation found in two women with abnormally low LDL levels, Amgen, Pfizer and Sanofi were in a race to test and seek approval for a drug that mimics the mutation as of mid-2013. Each firm's drug is a biologic (a monoclonal antibody made in living cells), which means enormous development and production expenses. However, the potential profits from such a drug are so great that the manufacturers are sparing no expense.

Winning the race to the market comes with a potential first-mover advantage, but more importantly, means there is more time to earn revenue from the drug before patent expiry. In short, every month cut out of development is an extra month of revenue before the drug's patent expires.

2.2 Amgen, Inc.

Amgen, headquartered in Thousand Oaks, CA, discovers, develops, manufactures, and delivers human therapeutics with one simple mission: "To serve
patients. Since its founding in 1980, Amgen has grown to become the world’s largest independent biotechnology company, boasting a staff of approximately 20,000, total revenue of $17.3 billion, product sales of $16.6 billion, and R&D expenses of about $3.4 billion for the year 2012 (Amgen, 2014).

"From process development and clinical manufacturing to full-scale therapeutic protein production, Amgen has built one of the industry’s largest operations." (Amgen, 2013). In 2013, Amgen broke ground on an innovative manufacturing facility in Singapore, complementing a worldwide footprint of manufacturing and distribution facilities in California, Colorado, Kentucky, Puerto Rico, Rhode Island, Ireland and the Netherlands.

Overall, Amgen is a well-positioned and well-run company, but with the complexity added by a continuing expansion of its pipeline and the pressure of increasing competition, efficiency in drug development and manufacturing is becoming even more critical.

2.2.1 Drug Product Commercialization and Manufacturing Network

The project scope is centered on the development, commercialization, and manufacturing of Drug Product (DP). Drug Product typically encompasses the formulation, fill, and finish (FFF) of the ultimate product. Before the FFF activities begin, bulk Drug Substance (DS) is produced, frozen, stored, and transported to the DP groups where the DS is thawed for use. After DS thaw, the DP FFF process is initiated with formulation and mixing, Bioburden filtration, and then an in-process hold. Next, DP goes through sterile filtration, product filling, and lyophilization (for
some products). Finally, the DP is inspected manually or automatically, labeled, packaged, and shipped. Overall, the Drug Product teams are responsible for producing the right dosage and presentation of finished drug product from the original bulk drug substance.

Before the FFF process can be used in clinical or commercial manufacturing, the appropriate drug product formulation needs to be determined and the appropriate process parameters for each unit operation (manufacturing step) need to be specified and tested at the lab scale. Once the formulation and process is defined, a process called technology transfer introduces the new formulation and process to the clinical environment and irons out any inconsistencies that arise due to scale up or equipment differences. After clinical trials are cleared, another technology transfer brings the formulation and process to the commercial manufacturing site(s). These four major operations (pictured in Figure 1) are owned by four separate functional groups that make up Amgen’s Drug Product Commercialization and Manufacturing Network. These groups are Drug Product Development (DPD), Drug Product Technology (DPT), Clinical Manufacturing, and Commercial Manufacturing.

Figure 1: Amgen's Drug Product Commercialization and Manufacturing Network

One of the major goals of the groups that make up the DP network is to achieve better alignment across groups and to approach projects and solutions from
a network level view. This commitment is ever more important given that the development is a decentralized network, making communication between groups difficult.

2.2.2 The Drug Product Technology Group

This project focuses on the Drug Product Technology group and its crucial connections across the Drug Product Commercialization and Manufacturing Network. DPT is made up of several smaller groups that mainly act in a process development and manufacturing support capacity. Once the Drug Product Development group recommends a formulation, DPT performs process development, process characterization, and validation studies to assess drug product manufacturability and recommend specific equipment settings and ranges for critical process parameters. In addition to manufacturing recommendations, these studies provide a significant portion of the data used to support regulatory filings. Manufacturing recommendations are initially made to Clinical Manufacturing, and then after some evaluation of the clinical operations, recommendations are also made to Commercial Manufacturing. Beyond the initial process recommendations, DPT plays a critical role in supporting Clinical Manufacturing and Commercial Manufacturing through technology transfer and non-conformance investigations.

2.2.3 Overview of Data Management

A Quick History
The use of data to make key decisions is now commonplace thanks to several advancements in data management technology. Prior to the 1950’s the infant state of data management, requiring the tabulation of countless punch cards, made the process of gleaning knowledge out of data very cumbersome and limited. To effectively use data improvements in both the storage of data and in processing speed were required. This was achieved over the next several decades with the advent of the disk drive by IBM in 1956, the introduction of the first database by General Electric in 1961, and the construction of the first relational database by Honeywell in 1976. Data management was positioned to enable the effective use of data in the decision-making process, symbolized by Proctor and Gamble’s use of the first “business intelligence” system in 1985 to leverage sales data for retail decisions.Shortly thereafter, “data warehouses” and IT departments became a critical component of most firms and the data management focus shifted from storing and processing data to finding the best ways to use data (Hayes, 2002).

**Big Data and Analytics**

With continued increases in the capacity for and the availability of data, new solutions and concepts, such as “Big Data” and “Analytics”, have captured the attention of many contemporary firms. Gartner (2013) differentiates big data from traditional data using the “3 V’s model”, which attributes volume (amount of data), variety (type of data), and velocity (speed of processing) as the defining characteristics of big data. Analytics refers to the targeted extraction of information or knowledge from data through analysis and pattern recognition. While big data and analytics can provide great insights and effectively supplement decision
making, their indiscriminate application in every situation and industry should be cautioned.

In recent years, many companies have successfully applied big data and analytics to their decision making processes. Many of the success stories have come from the technology and retail industries and rely heavily on customer data that had previously been unused. For example, a large factor in Amazon’s advantage over traditional book stores came from their ability to tailor reading recommendations to customers based on data collected from their browsing and purchase history. Similar benefits of big data can include improved demand forecasts, more finely targeted marketing, and more optimized operations. Overall, “companies that have incorporated data and analytics into their operations show productivity rates 5 to 6 percent higher than those of their peers” (Gordon 2012).

While the benefits of big data and analytics are clear, there are risks involved. First of all, big data essentially uses the past to predict the future and clearly fails when relied upon for strategic forward thinking. In the biotechnology space, it is advisable to take a more selective approach, targeting the use of big data toward specific deliverables to avoid the potential risks and superfluous costs associated with a blind adoption of big data and analytics. The process and strategy must be fully evaluated before making the investment.

Secondly, the technologies inherently decrease human involvement in the decision making process, replacing human intuition and analysis with strict IT processes. On one hand this aspect of the technologies could reduce data
verification needs, but minimizing human involvement does not always maximize value. This was highlighted in a 2005 chess tournament organized by world champion Gary Kasparov, eight years after he was defeated by the computer program Deep Blue. The tournament pitted combined teams of computers and chess players against each other. At this point, the community knew that combined teams were superior to humans or computers alone because the combined teams can use their complementary analytic strengths. “Computers don’t make mistakes; they are highly precise, while humans can use intuition and lateral thinking” (National Research Council, 2012). Even still, in a tournament with several chess grandmaster and supercomputer teams competing, a shocking victor emerged. ZackS, a team of two amateur chess players, using laptops to run open-source chess software dominated the field. The National Research Council goes on to describe Kasparov’s reaction to the outcome:

“Weak human + machine + better process was superior to a strong computer alone and, more remarkably, superior to a strong human + machine + inferior process” This revelation points to the essential evolution of the conclusion from Deep Blue in 1997 – that humans working together with machines can solve big data challenges better than computers alone. Tackling big data means more than just algorithms, high-performance computing, and massive storage – it means leveraging the abilities of the human mind.

This caveat to big data again echoes the importance of a thoughtful process and should be even more relevant in the biotechnology industry where the human element is generally highly educated, insightful, and creative. The invaluable instincts and knowledge of individual biotech scientists and engineers should not be disregarded in the application of big data and analytics.
Lastly, an overarching risk comes from the almost fanatical support for the use of big data and analytics in recent years. In this environment, and given the aforementioned risks, it is critical to analyze the appropriateness of big data and analytics for each application and determine effective uses of the concepts before making any large investments.

**IT Organization and Data Management Decisions**

One of the major areas of focus for a firm aiming to effectively use data is IT governance or organization. The main data management questions faced in this area include what decisions need to be made and who makes these decisions. Weill and Ross (2004) highlight these questions and provide potential answers in the form of a matrix of five major IT decisions and six IT decision making structures (Figure 2). The five major IT decisions, moving from general to specific, are:

1. IT Principles – General IT strategy decisions
2. IT Architecture – Definition of firm-wide technical requirements
3. IT Infrastructure – Determination of firm-wide IT services to provide capabilities before specific needs are known
4. Business Application Needs – Identification of the business requirements for purchased or internally developed IT applications
5. IT Investment and Prioritization – Decisions on how much and where to invest in IT

The six archetypes for IT decision making structures outlined by Weill and Ross, ranging from more centralized to less centralized, are:

1. Business Monarchy – A senior business executive or a group of senior business executives make decisions
2. IT Monarchy – IT executives make decisions
3. Federal – Coalition of C-level executives and business representatives from each operating group make decisions
4. IT Duopoly – IT executives and one other group of business leaders make decisions

5. Feudal – Business unit or process leaders make separate decisions based on their own needs

6. Anarchy – Each individual or group makes their own decisions

Figure 2: Matrix of IT Decisions and Structures (Weill and Ross, 2004)

Within a single firm, each of the IT decisions can and should be tackled by a different decision making structure, depending on which archetype fits the industry, firm maturity, and firm strategy, but Weill and Ross were able to make some overarching observations of which decision making archetypes are most common. They observed that the federal archetype is the most commonly used across all IT decisions, although IT principles are determined most by duopoly while IT architecture and IT infrastructure decisions are determined most by IT monarchy.
The process development and manufacturing focus of this thesis aligns best with the last two IT decisions: business application needs and IT investment and prioritization. These two IT decision categories were observed to be addressed mostly by the federal, duopoly, or business monarchy archetype. Specifically for decisions regarding business application needs, Weill and Ross recommend not to use the feudal model even though there may be pressure for each business unit to focus on their own constituents. Although each group may have their own individual needs, having business application needs addressed by each group can proliferate the use of different systems across the organization. In an organization seeking to integrate systems and make decisions from a more global view, it is important to make IT decisions using a more centralized model, such as a Federal model or an IT Duopoly model. An IT Duopoly model, where IT decisions are made by IT executives and one other group, such as subject matter experts, would fit well in the biotechnology environment where the input of process leaders is critical.

3 Project Approach

Before a data management strategy could be developed for the drug product technology group, we first had to better understand the steps involved in data management and decide which attributes were important to the organization at each step. From there, specific objectives were formed and existing and potential new projects were researched to help determine a project landscape that would best meet these objectives. To organize how we thought about the steps of data management and the important attributes of each step we built a framework around
what we call the data management cycle. Then, with the framework and exemplary attributes in mind, we explored how current and future projects would fit into an effective data management project landscape through three specific case studies of existing projects and interviews with several project leads and department heads.

3.1 The Data Management Cycle Framework

A more organized view of data management was shaped by first breaking down the flow of data into four distinct steps: data generation, data analysis, integration, and information access. By and large, to successfully pull value out of data, it needs to go through these four steps. First, the data needs to be collected and stored. Next, analysis needs to be performed to create comprehensible information from the data. Then the data and analysis needs to be integrated with data and analysis from other sources to provide a full picture of information. Lastly, the information needs to be accessed by those who can use it to make decisions and create value. This four step process is what we call the data management cycle. Completing the data management cycle will produce value from data, but the effectiveness of our data management strategy depends on how we execute each of the four steps.

3.2 Strategy Development

Our overall strategy consists of organizational objectives for each step of the data management cycle, as well as a proposed project landscape with specific projects to address each objective. To form the organizational objectives we defined attributes
of each step in the cycle that would help fulfill the needs of the overall development and manufacturing process. Figure 3 shows these objectives for each step of the data management cycle that will be discussed further in the following subsections. Ultimately several projects that could effectively meet each objective were identified. A subset of the projects, including three detailed case studies, will be discussed in the results section of this thesis to give examples of how each objective across the data management cycle can be met.

![Data Management Cycle Diagram]

**Figure 3: Organizational Objectives of the Data Management Cycle**

### 3.2.1 Data Generation

In determining the critical attributes of data generation for Amgen's Drug Product Technology group, there were two critical questions that we focused on; how should the data be collected and what data should be collected? Our answer to
the first question was placing an emphasis on electronic data collection. Electronic collection allows us to track more parameters, perform instant analysis, and avoid transcription errors and data verification steps. For the second question, we decided on a targeted approach to data collection. This means collecting only data that fulfills specific needs for organization-wide commercialization and manufacturing deliverables such as regulatory filings and process monitoring. This approach to data collection is extremely important to efficiency in data management because it conveys that we are only interested in the data that allows us to commercialize our product and deliver it to patients safely. We may be tempted to gather other nonessential data that may seem important to our individual group, but if this is outside the scope of what truly creates value for patients and the business, it only serves to lengthen the development timeline and increase cost. The answers to these questions led us to our overall objective for data generation. Collect the right data at the right time and in the right place from a network level view.

3.2.2 Data Analysis

Data analysis in the commercialization and manufacture of biologics can be complicated, time-consuming, and repetitive. However, the repetition that comes from moving several products through the same pipeline across multiple manufacturing sites and scales provides some promising opportunities. First of all, repetitive analysis can be automated. In addition, advanced analysis techniques, such as predictive models, can be employed to make decisions based on known product attributes and data from previous products or other manufacturing sites.
By investing some time upfront, both of these approaches have the potential to significantly reduce the workload required for each new drug, manufacturing scale up, or manufacturing site change. Overall, our objective is to apply automation and advanced analysis techniques to leverage cross-product and cross-site data.

3.2.3 Integration

There are several groups across the drug product commercialization and manufacturing network and each has specific needs and requirements. Accordingly, the groups often use different data storage systems, analysis packages, and even different lab notebook systems. In order to maintain the functionality and benefits that each group gets from their unique systems, we propose instead to focus on developing a network wide system that can aggregate information from each of these lower level databases and integrate it in one place where unique analysis and reporting can be performed. To streamline the issue of data and information storage, our ultimate objective is to smoothly integrate data, analysis, reporting, and validation into one user-friendly system.

3.2.4 Information Access / Connectivity

Currently, getting information from individual systems used by various groups across the drug product network requires informal searching and often requires extensive use of email to find out who knows the location of the desired information. Even when the appropriate location is found, receiving the requested information can take days to weeks. Fixing this issue is crucial because rapid information flow is critical to effective data management and can have a significant
impact on the development timeline. Electronic data capture and integration will inherently improve information access because the information will be available as soon as it is produced. Making the information readily accessible then depends on the development of systems that allow searching of the most up to date information on drug product development and manufacturing parameters. Overall, connections need to be created across the drug product network to enable real-time access to data and the development and manufacturing processes themselves.

3.3 Strategy Project Landscape and Case Studies

Finally, after the objectives for each step in the data management cycle were determined, a project landscape of existing and potential future projects that could help realize the objectives was built. Projects were identified through interviews with team leads and department heads, focusing on the needs of each department and how they related to the overall responsibilities of the drug product commercialization and manufacturing network, as shown in Figure 4.

Figure 4: Drug Product Commercialization and Manufacturing Responsibilities
The core data management responsibilities of the commercialization groups are to gather, analyze, and report data for the purposes of pre-commercial regulatory filings and technology transfer from the development labs to the clinical, or commercial, manufacturing site(s). Information for regulatory filings is specifically delivered in the form of Marketing Applications and Investigational New Drug Applications, while information for technology transfer is specifically delivered in the form of the Manufacturing Batch Records that provide instructions for the clinical and commercial manufacturing processes.

The core data management responsibilities of the commercial manufacturing groups are to gather, analyze, and report data for the purposes of commercial regulatory filings and to ensure overall product quality and efficiency. Information for regulatory filings is specifically delivered in the form of Annual Product Reviews, while information for the tracking of quality and process efficiency is specifically delivered through on-site process monitoring and remote monitoring capabilities.

The proposed data management strategy project landscape was built around projects that directly contributed to each of these responsibilities and deliverables. Each individual project also targeted at least one specific objective from the aforementioned data management cycle framework. While the overall project landscape includes several projects that cannot be detailed due to proprietary information, the following sections will explain one project that fits with each step of the data management cycle in either drug product commercialization or drug
product manufacturing, including in-depth analysis into three case study projects that were the focus of my internship.

4 Analysis and Results

4.1 Generation - Drug Product Manufacturing Data Collection - Case Study #1

The first case study project developed while working with the Drug Product Technology group explored what data to collect during the drug product manufacturing process, both at the clinical and commercial scales. The main deliverable for this case study was a technical assessment (TA) providing a holistic list of recommendations for which operating (inputs) and performance (outputs) parameters to track during each unit operation of the manufacturing process.

4.1.1 Data Collection Technical Assessment

For any given product, there are already requirements on critical parameters that must be monitored during manufacturing to meet certain regulatory requirements. However, tracking supplementary parameters can offer additional benefits. Overall, the technical assessment is meant to serve as a universal reference for work in clinical monitoring, process characterization, technology transfer, and commercial monitoring. First, the recommendations can be valuable in the clinical monitoring phase, where new monitoring plans can be adapted from the parameter list and data can be potentially leveraged in subsequent phases. The list can also be used to steer the implementation of electronic data collection in clinical manufacturing. Secondly, the guide can be helpful in process
characterization and technology transfer work by promoting better alignment of process parameters between sites. Lastly, in commercial monitoring, the technical assessment can be a foundation for real-time multivariate statistical process monitoring (discussed further in section 4.2) and help increase continuity in the development of specific master monitoring plans.

The recommended parameters were first determined through a fundamental analysis of each unit operation, using existing In-Process Control documents, Process Characterization Summary Reports, and Master Monitoring Plans for reference. The operating and performance parameters of the various unit operations were then cross-checked with documentation and current manufacturing programs in clinical, commercial, and contract manufacturing. Specifically, research focused on clinical manufacturing in Thousand Oaks, commercial manufacturing in Puerto Rico, as well as contract manufacturing sites. The list was then refined by consulting staff members from development teams, manufacturing site representatives, and unit operation subject matter experts who provided recommendations regarding which parameters were important to their specific unit operation(s).

The recommended process monitoring parameters covered each unit operation from drug substance freezing to drug product inspection. The full scope of covered unit operations is illustrated in Figure 5.
The document source listing contains information on each parameter’s in-process control classification, if applicable. Classifications for specific process characterization and monitoring plans should be deliberated and determined by the actual team creating the plan, because these classifications can vary depending on the drug, equipment, processes, and sites.

Along with the list of recommended parameters to be monitored, the full technical assessment also defines the specific step in each unit operation when the parameter should be tracked, the parameter units of measure, a justification for tracking the parameter, reference document sources, as well as the recommended data collection source(s). In general, transitioning to electronic collection of data where feasible is recommended as it provides a more efficient way to monitor trends and share information across functional groups. An example of the
recommended process monitoring parameters during the buffer formulation step is provided below in Table 1. Several pieces of information have been removed for proprietary reasons.

**Table 1: Sample of Recommended Process Monitoring Parameters during Buffer Preparation**

<table>
<thead>
<tr>
<th>STEP</th>
<th>PARAMETER</th>
<th>UNIT</th>
<th>TYPE</th>
<th>JUSTIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>Buffer Batch Size (Final Tank Weight)</td>
<td>kg</td>
<td>Operating</td>
<td>Tank weight to determine concentrations and yield at various time points</td>
</tr>
<tr>
<td>General</td>
<td>Room Temperature</td>
<td>°C</td>
<td>Operating</td>
<td>Process consistency</td>
</tr>
<tr>
<td>Solution Formulation</td>
<td>Amount of WFI added</td>
<td>kg</td>
<td>Operating</td>
<td>Affects concentrations.</td>
</tr>
<tr>
<td>Solution Formulation</td>
<td>Amount of Excipient(s) added</td>
<td>g</td>
<td>Operating</td>
<td>Affects concentrations.</td>
</tr>
<tr>
<td>Solution Formulation</td>
<td>Duration of Addition for Each Excipient</td>
<td>kg / min</td>
<td>Operating</td>
<td>Shown to affect viscosity and concentrations of DP</td>
</tr>
<tr>
<td>Solution Formulation</td>
<td>Mixing Time (duration)</td>
<td>Min</td>
<td>Operating</td>
<td>Indicates Process consistency. Affects product quality and uniformity of buffer</td>
</tr>
<tr>
<td>Solution Formulation</td>
<td>Mixing Speed</td>
<td>rpm</td>
<td>Operating</td>
<td>Shear effects and for formulation consistency</td>
</tr>
<tr>
<td>Solution Formulation</td>
<td>Mixing Tank Size</td>
<td>kg</td>
<td>Operating</td>
<td>Shear effects and for formulation consistency</td>
</tr>
<tr>
<td>Solution Formulation</td>
<td>Impeller Size</td>
<td>n/a</td>
<td>Operating</td>
<td>Shear effects and for formulation consistency</td>
</tr>
<tr>
<td>Solution Formulation</td>
<td>Buffer Hold Time Before Filtration</td>
<td>rpm</td>
<td>Operating</td>
<td>Process Consistency</td>
</tr>
<tr>
<td>Surfactant Prep</td>
<td>Amount of Surfactant Added</td>
<td>g</td>
<td>Operating</td>
<td>Affects concentrations.</td>
</tr>
<tr>
<td>Surfactant Prep</td>
<td>Amount of WFI Added</td>
<td>g</td>
<td>Operating</td>
<td>Affects concentrations.</td>
</tr>
<tr>
<td>Surfactant Prep</td>
<td>Amount of Other Raw Materials (RM) Added</td>
<td>g</td>
<td>Operating</td>
<td>Affects concentrations.</td>
</tr>
<tr>
<td>Surfactant Prep</td>
<td>RM Addition Rate</td>
<td>g/s</td>
<td>Operating</td>
<td>Process Consistency</td>
</tr>
<tr>
<td>Surfactant Prep</td>
<td>Temperature of WFI</td>
<td>°C</td>
<td>Operating</td>
<td>Indicates Process consistency.</td>
</tr>
<tr>
<td>Surfactant Prep</td>
<td>Diluted PS Hold Time (Duration)</td>
<td>Min</td>
<td>Operating</td>
<td>Ensure Process consistency</td>
</tr>
<tr>
<td>Testing</td>
<td>pH</td>
<td>N/A</td>
<td>Performance</td>
<td>Controls in place, but affects product quality</td>
</tr>
<tr>
<td>Testing</td>
<td>Conductivity</td>
<td>μS/cm</td>
<td>Performance</td>
<td>Controls in place, but affects product quality</td>
</tr>
<tr>
<td>Testing</td>
<td>Osmolality</td>
<td>mOsm/kg</td>
<td>Performance</td>
<td>Controls in place, but affects product quality</td>
</tr>
<tr>
<td>Testing</td>
<td>LAL - Endotoxin Level</td>
<td>EU/mL</td>
<td>Performance</td>
<td>Controls in place, but affects product quality</td>
</tr>
<tr>
<td>Testing</td>
<td>Bioburden</td>
<td>CFU/10 mL</td>
<td>Performance</td>
<td>Controls in place, but affects product quality</td>
</tr>
</tbody>
</table>
4.1.2 Recommendations

While the current technical assessment contains all recommended parameters for each unit operation in one document, it is recommended that unique documents be created for each unit operation to provide ownership to each specific unit operation expert. Cataloging these parameter recommendations in unit operation guidelines, for example, would allow the unit operation expert the flexibility to provide more detail and make changes whenever they deem necessary. Appropriate monitoring parameters can differ based on several factors, including SKU details and manufacturing site capabilities. Individual ownership would allow for greater accuracy in capturing the consequences of these factors.

Moving forward, it is recommended that an effort similar to the manufacturing data collection technical assessment be undertaken on the development side as well. There should be clearly identified owners for each development study, just as there are for each manufacturing unit operation. For the manufacturing side, parameters were identified for process understanding and regulatory needs. For development, the appropriate parameters required for each study should be standardized and documented. While an effort is underway to map out which studies provide data for regulatory filings, more work needs to be done to standardize and focus what data is collected in each of those studies. The case study presented in section 4.3 gives a potential tool to help in this area, but the content of the standardized studies should be determined by specific study experts and regulatory filing experts.
4.2 Analysis - RT-MSPM - Case Study #2

Our second detailed case study explores an example of the use of advanced statistical methods in data analysis. The case study involves using multivariate statistics to improve process monitoring in clinical drug product manufacturing. Real-Time Multivariate Statistical Process Monitoring (RT-MSPM) techniques have proven successful in drug substance manufacturing, but they have yet to be applied to drug product manufacturing. Improved analysis of drug product manufacturing through multivariate modeling techniques could provide many benefits, including a reduction in non-conformances and investigation time as well as potential yield increases. In pursuit of these objectives, a team was formed to model the Lyophilization (freeze-drying) step in the manufacturing process. The specific focus of this case study was on the modeling of a lyophilization equipment preparation cycle, which occurs before any drug product is placed inside the unit, in an attempt to catch equipment issues before drug product is put at risk.

4.2.1 Lyophilization Overview

Lyophilization is a common unit operation in the production of biologics, used to extend the shelf life of the product. In cases when liquid biologics show stability issues before the desired shelf-life, freeze-drying of the product into a more stable solid form is often used. The lyophilization unit operation occurs near the end of the manufacturing process, after the drug product is filled into vials.

In the first step of lyophilization, the liquid product is frozen to create a crystalline matrix to allow the escape of sublimating water without changing the
composition of the rest of the formulation. In some cases, an annealing step can also be performed by raising the temperature for a short time during the freezing step to promote crystal growth and quality, thereby bolstering overall product quality.

Next, the product goes through "primary drying". In this step, vacuum is pulled to reduce pressure and the product is heated to induce the sublimation of the water in the frozen product until the product reaches a moisture content of roughly 3 - 5%. Most remaining moisture is removed during "secondary drying" through additional heating that removes ionically bound water (Freeze Drying, 2009).

The temperature settings and behavior associated with the general lyophilization steps, and their relation to the water phase change diagram, are shown in Figure 6 and Figure 7, respectively (Matthew, 2009).

Figure 6: General Temperature Settings and Product Behavior during the Lyophilization Cycle (Matthew, 2009)
4.2.2 Lyophilization Equipment Preparation and the Sterilization Cycle

Before the product lyophilization cycle can be run, however, the lyophilizer needs to go through equipment preparation steps to ensure the unit is in proper working condition and to meet certain regulatory requirements. The equipment preparation cycles include a leak-up test, to ensure no leaks are present when the lyophilizer is highly pressurized, and a sterilization cycle, to steam-sterilize the lyophilizer before product is placed in the chamber. The sterilization cycle was chosen as the subject of our multivariate equipment monitoring case study because it is the last preparation step before the actual lyophilization product cycle and it utilizes all of the equipment used in the product cycle, except for individual product temperature probes.

The sterilization cycle is broken down into 5 distinct steps: purge and heat up, sterilization, drying, cooling, and final evacuation. In purge and heat up, steam is introduced into the lyophilizer until the chamber pressure reaches 15.2 PSIA and
then vacuum is pulled to remove any condensate and non-condensable gases (Cappia, 2004). Next, the steam inlet valves are reopened until the chamber temperature exceeds 120°C to begin sterilization. Conditions are held for 30 minutes to complete the sterilization step. In the drying phase, steam is dumped out of the steam drain and vacuum is pulled, reducing the chamber temperature to below 2 PSIA. Conditions are held for another 30 minutes. Then a cooling cycle is initiated by activating the compressor and opening the cooling valves until the condenser reads -30°C. Finally, vacuum is pulled again to remove any remaining moisture or gas. The temperatures and pressures inside the lyophilizer chamber throughout the five stages of sterilization are shown in Figure 8.

![Figure 8: Temperatures and Pressures throughout the Sterilization Cycle](image)
4.2.3 Real-Time Multivariate Statistical Process Monitoring

Historically, data from manufacturing runs at Amgen was captured manually and analyzed after the fact. If a problem occurred during a run, staff members wouldn't have a chance to figure out why the process ran the way it did until days, weeks, or months later.

In recent years, however, a new approach to process monitoring has been implemented with great success in drug substance manufacturing and is now being expanded to include drug product manufacturing. Now, manufacturing data is captured electronically and analyzed and interpreted nearly instantaneously by a software program in the middle of the run. The software uses multivariate statistical analysis to monitor many variables, process steps, and their interactions simultaneously. Through this analysis, the software can notify staff members when weak signals have been identified that could potentially lead to real issues with the equipment or process. If practical, staff members can even make adjustments to equipment or the process in the middle of a run to resolve the issue.

This advanced approach to process monitoring relies on multivariate statistics. Multivariate process monitoring and control specifically addresses two weaknesses in univariate monitoring and control. First of all, RT-MSPM allows the tracking of more parameters that wouldn't be realistically captured by a univariate approach. When using univariate charts, the work load and attention required by manufacturing staff increases as the number of parameters increases, making it impractical to track more than a few parameters at once. Conversely, when using a
multivariate chart, all variables are described by one statistic in one chart regardless of how many parameters are included, so the required workload and attention does not increase with the number of parameters. Secondly, traditional univariate process control charts monitor the behavior of process parameters independently, neglecting any effects of interaction between parameters. Multivariate process monitoring captures these interactions, reducing the chances of missing a real out-of-control signal. If the interactions between parameters were the sole cause of an out-of-control process, the issue would be missed by univariate charts. This failure to reject the null hypothesis of an in-control process is called type II error. Consequently, taking parameter interactions into account with RT-MSPM reduces type II error.

Figure 9 shows a graphical depiction of how multivariate monitoring and control can catch an out of control process that would be missed by univariate monitoring and control. Consider two normally distributed parameters being monitored in a process. If both parameters are plotted on univariate control charts, each observation falls within the control bounds and the process looks to be in control. However, when the interactions between the parameters are considered the control region is better defined by an ellipse determined by the chi-square distribution, $\chi^2$, where

$$\chi^2 = \frac{n}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} \left[ \sigma_2^2 (\bar{x}_1 - u_1)^2 + \sigma_1^2 (\bar{x}_2 - u_2)^2 - 2\sigma_{12} (\bar{x}_1 - u_1)(\bar{x}_2 - u_2) \right]$$
and the parameters have means $\mu_1$ and $\mu_2$, standard deviations $\sigma_1$ and $\sigma_2$, covariance $\sigma_{12}$, and $\bar{x}_1$ and $\bar{x}_2$ are the sample average of the variables with a sample size of $n$.

Specifically, while observation 11 in Figure 9 is in control in both univariate control charts, when each point is translated to the joint control region the multivariate point for observation 11 plots outside of the control ellipse. Therefore, the process is actually out of control, even though both univariate control charts show otherwise.

![Figure 9: A Control Ellipse for Two Dependent Variables (Montgomery, 2009)](image)

The control ellipse illustrates the benefit of multivariate monitoring and control, but when more than two variables are involved the ellipse becomes difficult to visualize. With more variables it is useful to describe the process using a Hotelling $T^2$ control
chart, where the combined effect of all variables is described as one statistic called $T^2$.

$$T^2 = (x - \bar{x})'S^{-1}(x - \bar{x})$$

where $\bar{x}$ and $S$ are the sample mean vector and covariance matrix, respectively. A control chart can then be constructed using $T^2$.

4.2.4 Application

While RT-MSPM is being applied to the lyophilization product cycle as a first step in clinical drug product manufacturing, in this case study, models were developed for the lyophilization sterilization cycle for two major reasons.

First of all, for the product cycle, a new model needs to be developed for each new product that is manufactured, but this can prove to be a difficult and inefficient proposition. Several batches of historical information are needed to build an accurate model and most products are not in clinical manufacturing for more than ten batches, so the number of runs that can actually get the benefit of the RT-MSPM model is small. This issue can be addressed by developing a cross-product model that works for all new products, and this should be investigated further. However, in the absence of an effective cross-product model, an RT-MSPM model for the sterilization cycle can potentially capture any equipment issues in the lyophilizer, regardless of the product that is being run. Process specific issues will not be caught by the sterilization model, but in recent years equipment issues have contributed to the majority of lyophilization non-conformances.
Secondly, applying RT-MSPM to the sterilization cycle puts an emphasis on greater monitoring of equipment cycles when product is not yet at risk. Currently, only data critical to each equipment preparation step is monitored during these cycles. During the leak-up test, only pressure readings are monitored. During sterilization, the cycle is successful as long as a temperature set point is reached and sustained for 30 minutes. Using advanced monitoring during equipment cycles provides more opportunities to catch weak signals before they turn into real issues and before any valuable product is put at risk. It is also worth noting that the major expense in the RT-MSPM system is for the data servers and software license. Once a system is installed at a manufacturing site, the cost of building more models for more processes is negligible.

To create a control chart for the lyophilization sterilization cycle we used Umetrics’ Simca-P+ to generate Hotelling T² control chart bounds for each phase in the process, based on historical data. In this case, eight variables from 10 historical batches that showed expected performance were used to create the Hotelling T² control chart. Batches were completed within two years of model development and incorporated the following parameters (some parameters have been disguised):

- Chamber Pressure from Viatran Sensor
- Chamber Pressure Transmitter
- “Equipment A” (Power Output)
- “Equipment B” Temperature
- “Equipment B” Vacuum
- "Equipment C" Temperature
- "Equipment D" Temperature
- "Equipment E" Vacuum

Figure 10 shows the corresponding control chart for the purge and heat up phase of the lyophilization sterilization cycle.

![Hotelling T2 Range - Purge & Heatup Phase](image)

**Figure 10: Hotelling T² Control Chart for Purge and Heat Up Phase (Screen Grab from Simca-P+)**

After the model was developed its effectiveness at catching weak signals in equipment performance was tested by applying the model to three sterilization runs that preceded product cycle runs resulting in non-conformances. All three test cases resulted in the termination of drug product and all three tests were performed offline, after the fact. The three test cases are as follows:

**Test Case #1**
- Class 3 NC on April 17, 2013
- Blown fuse on "Equipment A" caused temperature ramp time deviation
- Batch Terminated
Test Case #2
- Class 2 NC on June 27, 2012
- Refrigerant leak prevented shelf temperature from reaching set point
- Terminated (Equipment not ready)

Test Case #3
- Class 2 NC on June 7, 2012
- Silicone oil was observed coming from a pinhole leak in the system
- Terminated (Oil may have entered the vials)

If the RT-MSPM model were to catch weak signals during the sterilization cycle, the non-conformance in the subsequent product cycle could have potentially been prevented and the drug product that was scrapped could have potentially been saved.

4.2.5 Results
Results for each case are discussed below and show that the developed lyophilization sterilization model has the potential to catch weak signals before they develop into real issues that could threaten drug product in the lyophilization product cycle.

Test Case #1
The first test case investigates the sterilization cycle preceding the occurrence of a product cycle non-conformance in which a blown fuse on "Equipment A" caused a temperature ramp time deviation, altering the product cycle recipe in a way that required the scrapping of the drug product in the lyophilizer. The Hotelling T^2 statistic for each time point throughout the corresponding sterilization cycle was plotted against the historical Hotelling T^2
control bounds and an excursion was observed 60 minutes into the cooling phase, as shown in Figure 11.

![Hotelling T2 Range - Cooling Phase](image)

**Figure 11: Cooling Phase Excursion in Test Case #1 (Simca-P+ Screen Grab)**

Upon discovery of the cooling phase excursion, Umetrics' Simca-P+ software provides a variable contribution chart (Figure 12) that describes the cause for the excursion. In this case the "Equipment A" power output parameter is shown to be behaving significantly different than it has in the past. This corresponds well with the cause of the non-conformance observed in the subsequent product cycle.

![Variable Contribution Plot for Anomalous Point](image)

**Figure 12: Variable Contribution Plot for Test Case #1 Excursion**
By examining the specific performance of the "Equipment A" Power Output (Figure 13) we can see that the equipment requires nearly 100% power output in this sterilization cycle in order to perform its task, while historically the required power output has been less than 40%.

![Variable Analysis - "Equipment A" Output](image)

**Figure 13: Variable Analysis of "Equipment A" Power Output for Test Case #1**

In a "live" scenario, this observation would point to an issue with "Equipment A", which could be investigated before any product was put at risk. It is probable, but not certain, that such an investigation would have resulted in the replacement of either the fuse in "Equipment A" or of "Equipment A" itself. If this were the case, the non-conformance and loss of product in the subsequent product cycle would have been avoided. This excursion was caught because RT-MSPM allows the tracking of several parameters simultaneously, while the traditional univariate methods only allow for tracking of a few parameters at once due to manufacturing staff workload constraints. The "Equipment A" Output parameter was not monitored by the previous univariate control charts, but was able to be included in the multivariate
model. As the model is refined, even more parameters can be added, increasing the chance of catching more issues like this one.

**Test Case #2**

The second test case investigates the sterilization cycle preceding the occurrence of a product cycle non-conformance in which a refrigerant leak prevented the lyophilizer's shelf temperature from reaching its appropriate set point. The batch of drug product from this run had to be scrapped due to the deviation from the suggested lyophilization recipe. The Hotelling $T^2$ statistic for each time point throughout the corresponding sterilization cycle was plotted against the historical Hotelling $T^2$ control bounds and an excursion was observed at several points along the purge and heat up phase, as shown in Figure 14.

![Hotelling T2 Range - Purge & Heatup Phase](image)

**Figure 14: Purge & Heat Up Phase Excursion in Test Case #2 (Simca-P+ Screen Grab)**

Upon selection of one of these excursion points, Umetrics' Simca-P+ software provides a variable contribution chart (Figure 15) that describes the cause for the excursion. In this case, four variables show behavior roughly one standard deviation away from historical performance. By themselves, each variable would
not cause an excursion. However, the multivariate analysis also considers the interactions and dependencies of the four deviating variables and reveals that the system as a whole is actually out of control. A basic analysis of these interactions in Appendix A shows a high correlation and covariance between each of the four variables, supporting the overall compounding effect that is observed when each individual variable is behaving slightly abnormally.

![Variable Contribution Plot for Anomalous Point](image)

**Figure 15: Variable Contribution Plot for Test Case #2 Excursion**

Figure 16 shows the performance of the "Equipment D" Temperature parameter, which gives an approximation of the temperature inside the lyophilization chamber. It is evident that the temperature reading exceeds the typical range seen in past batches. It is difficult to make any conclusive assertions, but it is possible that the system is taking longer to reach the desired temperature set points than in the past. Regardless, the multivariate analysis has picked up on some unusual performance that would justify investigation of the lyophilizer before proceeding to the product cycle. Whether the refrigerant leak had already occurred
or would have been discovered by the investigation is uncertain, but there is the possibility that the non-conformance and loss of product in the subsequent product cycle could have been avoided if the RT-MSPM system were in use.

![Variable Analysis – "Equipment D" Temperature](image)

**Figure 16: Variable Analysis of "Equipment D" Temperature for Test Case #2**

**Test Case #3**

The third and final test case investigates the sterilization cycle preceding the occurrence of a product cycle non-conformance in which silicone oil was observed coming from a pinhole leak in the system. Due to the risk of contamination, the batch of drug product from this run had to be scrapped. The Hotelling $T^2$ statistic for each time point throughout the corresponding sterilization cycle was plotted against the historical Hotelling $T^2$ control bounds but no excursion was observed, as shown in Figure 17. In this case, the RT-MSPM model would not have triggered an investigation and would not have prevented the non-conformance that occurred in the subsequent product cycle run. Overall, however, the sterilization cycle RT-MSPM model has shown the potential to catch two out of three of test cases.
4.2.6 Recommendations

First of all, we recommend moving forward with the implementation of RT-MSPM in the clinical environment and with the use of equipment cycle models where possible, as they show the potential for catching weak signals before they develop into real issues. However, there is still a question of where to invest time and resources first when it comes to building new RT-MSPM models.

To answer this question, a survey of the potential cost savings associated with four unit operation candidates was carried out. Since the cost of the RT-MSPM comes mostly from the installation of the server and software, the expenses associated with building each model were considered negligible and the required time and resources should be roughly the same for each model. Therefore, cost savings became the focus of the prioritization analysis. Four unit operations, lyophilization, vial filling, depyrogenation, and vial washing, were considered. Yearly total savings from the RT-MSPM models were broken down into three categories: Full-time employee (FTE) savings from non-conformance investigation.
duration reduction, yield savings, and saved lot savings. Results were based on 2013 clinical manufacturing data for each unit operation including number of NCs, number of clinical runs, number of vials per run, reject percentage, and terminated lots. Table 2 shows the results of the prioritization analysis.¹ ²

<table>
<thead>
<tr>
<th>Potential Savings</th>
<th>Lyophilizer</th>
<th>Filling (Vial)</th>
<th>Depyrogenation</th>
<th>Vial Washer</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC Investigation/FTE Savings</td>
<td>$1,400.00</td>
<td>$3,000.00</td>
<td>$8,000.00</td>
<td>$9,000.00</td>
</tr>
<tr>
<td>Yield Savings</td>
<td>$10,300.00</td>
<td>$8,500.00</td>
<td>$-</td>
<td>$-</td>
</tr>
<tr>
<td>Saved Lot Savings</td>
<td>$98,300.00</td>
<td>$78,900.00</td>
<td>$3,000.00</td>
<td>$2,400.00</td>
</tr>
<tr>
<td>Projected Yearly Savings</td>
<td>$110,000.00</td>
<td>$90,400.00</td>
<td>$11,000.00</td>
<td>$11,400.00</td>
</tr>
</tbody>
</table>

The prioritization analysis shows that the major benefit of the RT-MPSM models comes from the potential to save lost product lots. Building models for product cycles or equipment that will eventually contain product is more beneficial than building models for component preparation unit operations. Even if there are more non-conformances occurring in the depyrogenation and vial washing component preparation unit operations, the potential savings is small because no product is ever present in those unit operations. There is still a benefit to building those models, but models for product-centric unit cycles should be built first to capture the larger potential savings first. Recommendations for further work include the development of a cross-product lyophilization product cycle model and investigation into ways of leveraging the clinical scale lyophilization models to accelerate the development and implementation of a companion commercial scale model.

¹ Assumptions: 5 hours of investigation/NC, $110/hr FTE, 10% reduction in rejects, 25% lost lot prevention, ~$20 COGM/Vial
² Data and results disguised
4.3 Integration – ELN BioBook – Case Study #3

The third and final case study focuses data integration in drug product development through the use of IDBS’s Electronica Lab Notebook (ELN) software and its imbedded BioBook spreadsheet functionality. Traditionally, the Drug Product Technology groups have captured development study data in hand-written lab notebooks. This tried-and-true method captures the necessary information from the studies, but several benefits that could help reduce development time and improve process understanding only become realizable through the use of an electronic system. The BioBook system provides several of these benefits, including a platform to create templates of standardized studies, automation of analysis, and semi-automated report generation. However, the main “data management-centric” benefit of the system lies in its ability to integrate data from all development studies in one structured database to leverage cross-study and cross-product data.

On a cross-study level, integrating all development study data in one location allows drug product teams to incorporate data from previous studies for the same product without delay because all of the information is available in the same place. The drug product team can more easily take steps to build a more comprehensive picture of product attributes and performance.

The cross-product benefits of ELN BioBook come from the use of standard study designs and templates. If used faithfully, the standard templates will encourage drug product teams to enter the same information for each study from product to product. Over time, a database of results for all products is built for each
study. Eventually, enough information will be available to perform cross-product analyses and determine how different product characteristics impact overall study results. In some cases it is even possible that predictive models could be built from the available information to avoid the execution of certain development studies altogether.

As of today, the Drug Product Technology staff has adopted the use of the electronic lab notebook system, but has only just begun to explore the potential of the BioBook functionality. To illustrate and examine the extent of the aforementioned benefits we have applied the Electronic Lab Notebook software and its BioBook functionality to one of the drug product development studies.

For the purposes of this case study the ELN BioBook system was used to develop a workflow for the photosensitivity study. This study was chosen because a study standardization effort was already underway and because the results of the study depend on collaboration with the Drug Product Development (DPD) group, which is responsible for performing assays on samples collected during the study. This collaboration was desirable because DPD had previous experience with the intricacies of BioBook and it allowed for the integration of data from two separate departments into one location. Generally, DPD and DPT exchange sample information and results through email or by sending links to separate ELN study entries. With BioBook, all information from both departments is entered into the system one time and stored in the same location, reducing transcription errors and data verification needs.
4.3.1 Photosensitivity Study

Photoexposure, both visible and ultraviolet, occurs throughout the drug product manufacturing process and protein drug products in liquid are susceptible to photo-degradation. Hence, it is critical to evaluate the susceptibility of liquid protein products to light exposure during manufacturing. The photosensitivity study is designed to determine the acceptable amount of photoexposure allowed during the manufacturing process.

Generally, the duration of exposure to light throughout the manufacturing process and the light intensity at each manufacturing site is known. From this information total expected photoexposure throughout the manufacturing process can be determined. The photosensitivity study subjects drug substance and drug product samples to the equivalent photoexposure in the testing lab. At several points during the study samples are removed from the testing environment and sent to the DPD team which performs multiple assays to determine the effect of photoexposure on the drug substance and drug product. If the assays show degradation in samples collected before the manufacturing process photoexposure is reached, steps need to be taken to reduce the light intensity at the manufacturing site or the light exposure duration during the manufacturing process. Overall, the objectives of the photosensitivity study are:

1. Evaluate the photosensitivity of the drug substance and drug product
2. Recommend the maximum allowed photoexposure for both visible and UVA light intensities during the commercial manufacturing process.

Control samples are also included in the study to isolate the effects of photoexposure. Before beginning the study, control samples of drug substance and drug product are set aside and sent for testing. These samples were shielded from light and were intended to provide the baseline conditions for both drug substance and drug product prior to light exposure. Additional samples of drug substance and drug product are wrapped in aluminum foil and placed in the testing environment alongside test samples. The aluminum foil protects these controls from light exposure, but they are still exposed to the same temperature for the same duration as the exposed test samples. Comparison of the controls to the exposed samples helps discern if any detectable changes observed by the selected product assays are due to light exposure. A simplified flow diagram for the samples involved in the drug product photosensitivity study is shown in Figure 18.

Figure 18: Flow Diagram for Drug Product Evaluation in the Photosensitivity Study
Multiple samples are collected at each time point because several assays are performed to assess product quality. Assays include inspection of product appearance, color, and clarity, sub-visible particles quantity, and protein aggregation.

4.3.2 Study Standardization

Study standardization is a crucial prerequisite to the use of BioBook for development studies. In order to reap the full benefits of the system and integrate cross-product data into a single database, the study first needs to be standardized and then run according to the standard protocol whenever the study is executed. Standardization is the most effective way to ensure that the appropriate data is collected and entered into the database.

At the time of the initiation of the BioBook implementation project, an effort was underway to standardize the design of the photosensitivity study, however, as the project moved forward, the official standardization of the study lagged behind the BioBook implementation schedule due to other pressing priorities. In order to move forward with the BioBook implementation a provisional standard study was developed by the BioBook project team. Past photosensitivity study reports were referenced and development scientists were interviewed to determine an appropriate provisional standard design to be used as the basis for the ELN BioBook study template.
The photosensitivity study description in section 4.3.1 reflects the basics of the resulting provisional standard study design. In addition, standard time points for sample collection were chosen and the standard assays to be performed for the study were specified. Both specifications were made with the expectation of providing the user the flexibility to adjust the time points and add additional assays in ELN, if necessary.

Overall, the purpose of standardization is to provide a common starting point and study design for anyone doing the photosensitivity study. The intent is to capture the same core information that is needed for regulatory filings and technology transfer whenever the study is done. Additionally, if the study is done the same way every time, data can more easily be compared from one study to the next. Once a standard design is determined, the intent is to use ELN BioBook to create a template for a study report that can be filled with all of the standard information upon study execution. However, since each drug product is unique, non-standard assays or other distinctive study design components may be required. Accordingly, sufficient flexibility needs to be designed into the ELN BioBook template.

4.3.3 Photosensitivity ELN BioBook Template

Given the provisional standard photosensitivity design, an ELN BioBook template was created provide a framework for future lab notebook entries and study reports. The template consists of pre-written common text with blanks for product specific information, user input tables that are used to fill in the blanks, and
results tables that are filled by automatic calculation based on the user inputs. Summary, conclusion, and some analysis sections are left blank for unique assessments by the author.

When an author conducts a photosensitivity they create a new lab notebook entry from the photosensitivity template. Figure 19 shows part of the template created for the photosensitivity study. The author simply enters product specific information into the yellow cells and the blanks in the stock text are automatically filled with the appropriate information from the table. A large portion of the lab notebook entry can be completed just from these user inputs.

Figure 19: Screenshot of Photosensitivity Template in IDBS ELN
4.3.4 Assay Results and Automated Graph

Next, the author needs to generate the assay results graphs using a BioBook spreadsheet specifically pre-designed for collaboration between DPT and DPD. The author, representing DPT, enters information about the samples submitted to DPD and, in turn, a DPD representative enters results from the assays performed on each sample. Figure 20 shows a portion of the photosensitivity “assay results BioBook spreadsheet”.

![Sample Information (Entered by DPT) and Assay Results (Entered by DPD) screenshots]

As the appropriate information is entered into the spreadsheet, BioBook automatically generates the necessary graphs for each assay. Figure 21 shows an example of a graph comparing the performance control and exposed samples for an

Figure 20: Screenshot of Photosensitivity “Assay Results BioBook Spreadsheet” in IDBS ELN

As the appropriate information is entered into the spreadsheet, BioBook automatically generates the necessary graphs for each assay. Figure 21 shows an example of a graph comparing the performance control and exposed samples for an
assay measuring protein aggregation and fractionation. From here the author places a copy of the graphs into their electronic lab notebook entry.

![Graph of SEC Results - Main Peak %](image)

**Figure 21: Screenshot of Assay Results Graph in IDBS ELN**

Finally, once the entire ELN entry is completed for the photosensitivity study, the author can convert the ELN entry into a Technical Report in Microsoft Word.

It is critical to note that significant upfront work was required to develop the templates that are used by the author to generate the pieces described in the previous sections. A framework template was built for the ELN entry, a BioBook spreadsheet template with programmed generation of the desired charts was built with significant help from the DPD BioBook developer, and an ELN to Word conversion template was created for the generation of Technical Reports.
4.3.5 Recommendations

Assuming that full implementation of ELN BioBook is adopted, a prioritization of which studies to incorporate first has been established based on interviews with subject matter experts and past study data. Development studies were scored in 4 relevant areas: Potential Time Savings (on a 0 – 10 scale), Product to Product Variation (0 – 10), Current Standardization Status (0 – 5), Type of Results (0 – 5).

The Potential Time Savings category seeks to quantify the benefit of ELN BioBook with regard to reducing the time required for development study analysis and reporting. The score for each study was determined by its frequency of use, typical report length, and number of unique analyses required. Product to Product Variation was determined through interviews with experts and depended upon each expert’s perception of how much a study’s design differs from product to product. Studies with the lowest variation from product to product would gain more benefit from the ELN BioBook system. Current Standardization Status and Type of Results were given half the weight of the other two categories. A study received a score of 5 if it was already standardized and another 5 if its results were mainly assays as it could easily leverage the work already done for the photosensitivity assay results spreadsheet. In total, each study received a score out of 30 potential points.

The prioritization results shown in Table 3 have been disguised, but the major takeaways are the same. Studies with low variation from product to product that can leverage the existing “assay results BioBook spreadsheet” should be targeted first during the ELN BioBook roll-out.
Table 3: Prioritization of Development Studies for ELN BioBook Roll-Out

<table>
<thead>
<tr>
<th>Study</th>
<th>Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale-Down Process Evaluations</td>
<td>20</td>
</tr>
<tr>
<td>FIH Process Evaluations</td>
<td>19</td>
</tr>
<tr>
<td>Buffer Characterization</td>
<td>19</td>
</tr>
<tr>
<td>Particle Kinetics</td>
<td>17</td>
</tr>
<tr>
<td>Freeze-Thaw</td>
<td>17</td>
</tr>
<tr>
<td>In-Process Hold</td>
<td>17</td>
</tr>
<tr>
<td>Mixing</td>
<td>16</td>
</tr>
<tr>
<td>Cleanability</td>
<td>16</td>
</tr>
<tr>
<td>Photosensitivity</td>
<td>15</td>
</tr>
<tr>
<td>Binding (to filter) Kinetics</td>
<td>15</td>
</tr>
<tr>
<td>VPHP</td>
<td>15</td>
</tr>
<tr>
<td>Filling Evaluation</td>
<td>14</td>
</tr>
<tr>
<td>Shear Stress</td>
<td>13</td>
</tr>
<tr>
<td>Syringe Performance</td>
<td>12</td>
</tr>
<tr>
<td>Physical Properties</td>
<td>11</td>
</tr>
<tr>
<td>HUV</td>
<td>11</td>
</tr>
<tr>
<td>Inspection</td>
<td>9</td>
</tr>
<tr>
<td>Filtration</td>
<td>8</td>
</tr>
</tbody>
</table>

Overall, ELN BioBook shows great promise as a way to integrate development data into one place, but there is a significant effort required to build and maintain the required templates and spreadsheets for each study. Implementing BioBook for all development studies will likely take at least three years, moving at a pace of one study every two months. This is also assuming that a dedicated resource is assigned to develop the templates and spreadsheets and that development study experts are assigned to standardize each study before they are introduced into BioBook. It is critical that study standardization is completed before BioBook implementation because any changes in the study design will require changes to the BioBook templates and spreadsheets and will lessen the potential benefits of leveraging cross-product data over time. Another concern is that the long-term success of the
system will depend heavily on the software developer. Whether Amgen wants to have such a large undertaking tied to an external entity is questionable. Before moving forward with full implementation, it is advisable to explore a long-term relationship with the software developer to ensure on-going development support.

Focus should be placed first and foremost on study standardization and creating the ELN study templates as each of these areas comes with significant benefits without significant development requirements and the risks of relying heavily on external parties.

4.4 Information Access / Connectivity - Remote Monitoring

Although no case study was explored for the final step of the data management cycle, connectivity is essential in a decentralized environment to effective data management in a decentralized environment such as the Drug Product Commercialization and Manufacturing Network. Great work can be done in generating, analyzing, and integrating data, but if the resulting information is not easily accessible to inform critical decisions the value created in the first three steps can be lost or seriously delayed. Several examples from the perspective of the process development team within DPT highlight the need for streamlined information access and connectivity between groups.

DPT’s process development teams rely heavily on information produced by other groups and departments, including Drug Product Development, Clinical Manufacturing, and Commercial Manufacturing. In order to make accurate manufacturing recommendation and support on-going manufacturing efforts, the
process development teams need access to all manners of data and information, not limited to formulation data, clinical manufacturing batch record reports, and commercial manufacturing operating conditions. The faster the process development teams get this information, the faster they can prepare products for introduction into the clinical and commercial environments, and the faster they can help solve production issues. However, currently it is not unusual for process development teams to spend weeks tracking down data to address simple issues.

One such example comes from a recent technology transfer of a new product to a commercial manufacturing site. In preparation for the technology transfer, one of DPT’s process development teams ran a characterization study in the development lab and recommended an operating range for a filtration variable based on their equipment in the lab. When the process was run at the manufacturing site, however, the filtration variable was running outside of the recommended operating range. Upon notification of the issue, and after some site to site discussions about the occurrence, the process development had to request data from the manufacturing run. Manual batch records had to be compiled and emailed to the development team. Once the process development team had the data they were quickly able to determine the source of the issue, but the whole issue resolution process took close to two weeks.

One way to combat this issue would be to develop the capability to monitor manufacturing process parameters in real-time in a system with a user friendly interface and provide access to DPT’s process development scientists. This would
allow the development teams to see current manufacturing performance from anywhere in Amgen's network, analyze any unit operation, and solve a wide range of issues in real-time. Providing this kind of access could potentially reduce the time required for each investigation by weeks and significantly reduce the overall duration of technology transfer. This example shows just one potential area for improvement, but there are many more opportunities to realize the benefits of expanding access to data and creating similar connections between a number of other groups.

5 Summary of Recommendations and Conclusion

Through the aforementioned case studies and extensive interviews with subject matter experts and department heads, a data management strategy was proposed for the Drug Product Technology group within the Drug Product Commercialization and Manufacturing Network. The proposed strategy suggests broad objectives for improvement at each step in the simplified data management cycle, which guided the determination of a potential project landscape to meet each objective. Three specific project case studies within the project landscape were explored in the body of this thesis. For each case study, the data management benefits, potential pitfalls, and recommendations moving forward were discussed.

Specific recommendations include:
• Create unique documents for each unit operation based on the internally developed Data Collection Technical Assessment to provide ownership to each specific unit operation expert.

• Clearly identify owners for each process development and characterization study to encourage unilateral decision making and facilitate easier standardization.

• Determine, standardize, and document the appropriate parameters required for each development and characterization study based on what is absolutely necessary for regulatory filings and technology transfer to clinical and commercial manufacturing.

• Move forward with the implementation of RT-MSPM in clinical manufacturing and with the use of equipment cycle models to catch potential process issues before product is at risk.

• Prioritize the unit operations with the highest potential to save lost product lots when implementing RT-MSPM.

• Explore ways to develop a cross-product lyophilization product cycle model and ways to leverage the clinical scale lyophilization models to accelerate the development and implementation of companion commercial scale model.

• Target studies with low variation from product to product that can leverage the existing “assay results BioBook spreadsheet” during the ELN BioBook roll-out.

• Focus first and foremost on study standardization and creating the ELN study templates as each of these areas comes with significant benefits without significant development requirements and the risks of relying heavily on external parties.
Ultimately, making a significant improvement in data management efficiency will require a persistent effort and the continued support of the Drug Product Technology group's leadership. The proposed data management cycle framework has provided a simplified way of viewing and streamlining data management and the resulting project landscape has provided a near-term direction for making data management improvements. However, as many of the outlined recommendations suggest, success moving forward will rely heavily on the commitment of staff to standardization and a willingness to contribute significant efforts in continuous improvement projects. Leadership support and endorsement will be critical in fostering this commitment and realizing the goal of streamlining data management across the Drug Product Commercialization and Manufacturing Network.
6 References


### Appendix A: Variable Interactions from RT-MSPM Case Study #2

#### Multivariate Correlations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Chamber Pressure</th>
<th>Equipment B Temperature</th>
<th>Equipment D Temperature</th>
<th>Equipment E Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chamber Pressure</td>
<td>1.0000</td>
<td>0.6696</td>
<td>0.8278</td>
<td>0.7108</td>
</tr>
<tr>
<td>Equipment B Temperature</td>
<td>0.6696</td>
<td>1.0000</td>
<td>0.9599</td>
<td>0.8518</td>
</tr>
<tr>
<td>Equipment D Temperature</td>
<td>0.8278</td>
<td>0.9599</td>
<td>1.0000</td>
<td>0.9013</td>
</tr>
<tr>
<td>Equipment E Temperature</td>
<td>0.7108</td>
<td>0.8518</td>
<td>0.9013</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

#### Covariance Matrix

<table>
<thead>
<tr>
<th>Variable</th>
<th>Chamber Pressure</th>
<th>Equipment B Temperature</th>
<th>Equipment D Temperature</th>
<th>Equipment E Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chamber Pressure</td>
<td>95.52495</td>
<td>316.67736</td>
<td>242.95249</td>
<td>202.39478</td>
</tr>
<tr>
<td>Equipment B Temperature</td>
<td>316.67736</td>
<td>2341.2620</td>
<td>1394.8479</td>
<td>1200.7114</td>
</tr>
<tr>
<td>Equipment D Temperature</td>
<td>242.95249</td>
<td>1394.8479</td>
<td>901.81981</td>
<td>788.53667</td>
</tr>
<tr>
<td>Equipment E Temperature</td>
<td>202.39478</td>
<td>1200.7114</td>
<td>788.53667</td>
<td>848.71260</td>
</tr>
</tbody>
</table>

#### CI of Correlation

<table>
<thead>
<tr>
<th>Variable by Variable</th>
<th>Correlation</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equipment B Temperature by Chamber Pressure</td>
<td>0.6696</td>
<td>0.6363</td>
<td>0.7004</td>
</tr>
<tr>
<td>Equipment D Temperature by Chamber Pressure</td>
<td>0.8278</td>
<td>0.8086</td>
<td>0.8452</td>
</tr>
<tr>
<td>Equipment D Temperature by Equipment B Temperature</td>
<td>0.9599</td>
<td>0.9551</td>
<td>0.9643</td>
</tr>
<tr>
<td>Equipment E Temperature by Chamber Pressure</td>
<td>0.7108</td>
<td>0.6809</td>
<td>0.7384</td>
</tr>
<tr>
<td>Equipment E Temperature by Equipment B Temperature</td>
<td>0.8518</td>
<td>0.8350</td>
<td>0.8670</td>
</tr>
<tr>
<td>Equipment E Temperature by Equipment D Temperature</td>
<td>0.9013</td>
<td>0.8898</td>
<td>0.9117</td>
</tr>
</tbody>
</table>
Scatterplot Matrix

Chamber Pressure

Equipment B
Temperature

Equipment C
Temperature

Equipment E
Temperature

r = 0.66
r = 0.85
r = 0.71
r = 0.27
r = 0.95
r = 0.01
r = 0.01
r = 0.00