Spring 5-14-2022

The Intersection of Robotic Process Automation and Lean Six Sigma Applied to Unstructured Data

Emily McIntosh
Southern Methodist University, emcintosh@smu.edu

THE INTERSECTION OF ROBOTIC PROCESS AUTOMATION

AND LEAN SIX SIGMA

APPLIED TO UNSTRUCTURED DATA

Approved by:

Dr. Aurelie Thiele
Associate Professor

Dr. Richard Barr
Associate Professor

Dr. Harsha Gangammanavar
Assistant Professor

Dr. Eli Olinick
Associate Professor

Mike Wudke
Industry Representative

THE INTERSECTION OF ROBOTIC PROCESS AUTOMATION

AND LEAN SIX SIGMA

APPLIED TO UNSTRUCTURED DATA

A Praxis Presented to the Graduate Faculty of the

Bobby B. Lyle School of Engineering

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Engineering

with a

Major in Engineering Management

by

Emily McIntosh

B.S., Management Science, Southern Methodist University
M.S., Operations Research, Southern Methodist University
M.S., Systems Engineering, Southern Methodist University

May 14, 2022

ACKNOWLEDGMENTS

This work could not have been accomplished without the guidance of my advisor, Dr. Aurelie Thiele and the rest of my supervisory committee: Dr. Richard Barr, Dr. Harsha Gangammanavar, Dr. Eli Olinick, and Mike Wudke. I would also like to express my gratitude for the late Dr. Jeffery Kennington who first introduced me to the world of Management Science.

McIntosh, Emily B.S., Management Science, Southern Methodist University
M.S., Operations Research, Southern Methodist University
M.S., Systems Engineering, Southern Methodist University

The Intersection of Robotic Process Automation
And Lean Six Sigma
Applied to Unstructured Data

Advisor: Dr. Aurelie Thiele

Doctor of Engineering conferred May 14, 2022

Praxis completed April 1, 2022

While new Artificial Intelligence (AI) technologies gain traction in the workplace, there seems to be more buzz around these newer advances, including Robotic Process Automation (RPA), than more established process improvement techniques such as Lean Six Sigma. This praxis research uses Lean Six Sigma as a framework for effectively deploying these emerging technologies, a challenge for 86% of companies (Ernst & Young, 2021). This research is applied to one of the legal industry's most resource intensive processes – eDiscovery in the environment of a Big 4 accounting firm that provides services to corporations and legal professionals alike.

Electronic discovery (also known as e-discovery, ediscovery, eDiscovery, or e-Discovery) is the process of identifying, collecting, producing, and presenting electronically stored information (ESI) in response to a request for production in a lawsuit or investigation. ESI can include any type of electronically stored file and commonly includes emails, documents, databases, media files, social media, and web sites. The lifecycle of eDiscovery has been defined by the Electronic Discovery Reference Model (EDRM) as having the following phases: Information Governance, Identification, Preservation. Collection, Processing, Review, Analysis, Production, and Presentation. To move through the phases of the EDRM historically requires a significant investment in time, technology, and human resources.

This project had its origins as an automation effort driven by the technical advances in RPA solutions. However, RPA became a tool for to enable the program – not the solution itself. The DMAIC framework (Define, Measure, Analyze, Improve, Control) of Lean Six Sigma laid the foundation for a more wholistic analysis of the EDRM including the identification of processes that required revision prior to their automation. The Define phase identified the resource intensive strain moving through the EDRM causes corporations, vendors, and litigators. Through the measure phase, an opportunity to provide better results faster, and therefore cheaper was quickly identified. Through the analysis, several unnecessary handoffs, extraneous processes, and general bottlenecks in the process were refined. Through the Improve phase, automation played a significant part in realizing the efficiencies identified in the analyze phase. Finally, the controls phase not only put these improved processes into place but also quantified the value of ensuring these procedures were thoroughly deployed.

This research is organized using the DMAIC framework to articulate the process for completing the research, the gains and efficiencies made throughout the analysis, and to measure the impact and success of the overall program enhancements.

The impact of this project is measurable not only in the reduction of defects as defined by Lean Six Sigma, but also a significant improvement in time required to complete these processes. Even more satisfying, these efficiencies have a measurable, financial impact that has currently been realized north of $5 million USD in one year alone. This impact led to the solution becoming a finalist for an industry award where it was presented to over 3,000 industry professionals. Furthermore, the reduction and automation of manual, tedious tasks have also led to more enriching work for resources

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

This work is dedicated to my family.

CHAPTER 1

INTRODUCTION

Robotic Process Automation (RPA) involves using virtual “bots” to perform routine tasks and has become a buzzword in the professional services industry. Management research, including Forrester, Gartner, and EY among others, has clearly established that RPA technology will continue to be more widely adopted. However, automating inefficient processes has the potential to create significant resource constraints in addition to multiplying errors in processes that are not consistently accurate. This research applies the principles of Lean Six Sigma as the foundation of building an RPA program – specifically in the context of the Electronic Discovery Reference Model (EDRM). This chapter is organized by first discussing the background and context, followed by the research problem, the research aims, objectives and questions, the significance and finally, the limitations.

1.1 Background

In order for this research project to be possible, it is first important to understand the background of both RPA and Lean Six Sigma. Additionally, it will be important to understand the eDiscovery lifecycle in order to appreciate the application of these two efficiency methodologies.

1.1.1 Robotic Process Automation

RPA is an umbrella term for applications that operate on the user interface of other systems in the way a human would do instead of via back-end integrations. RPA aims to replace human input on manual, repetitive tasks by automation done in an "outside-in'' manner (van der Aalst, 2018). Gartner defines RPA as "a productivity tool that allows a user to configure one or more scripts (which some vendors refer to as "bots") to activate specific keystrokes in an automated fashion. The result is that the bots can be used to mimic or emulate selected tasks (transaction steps) within an overall business or IT process. These may include manipulating data, passing data to and from different applications, triggering responses, or executing transactions. RPA uses a combination of user interface interaction and descriptor technologies. The scripts can overlay on one or more software applications." (Gartner, 2021)

Precursors to RPA include screen scrapers, workflow automation and management tools, and Artificial Intelligence (AI). The term can first be traced back to the early 2000s but leading companies including UiPath, Automation Anywhere, and Blue Prism, began gaining traction around 2013. According to Gartner, by 2018 RPA platforms make up the largest growth in software development companies.

1.1.2 Lean Six Sigma

Lean Six Sigma marries the quality focus of Six Sigma with the cost saving principles of the lean management style and can be summarized as:

1. Reducing defects and errors through a data-driven approach that focuses on root-cause analysis
2. Focusing on the management strategy of employees distinguishing the business from the bottom line
3. Combining engineering philosophy with low-risk techniques

Six Sigma aims to reduce defects by focusing on the 5 Ss. The original 5 Ss are Japanese words, with their English translations below (BPI, n.d.):

- Seiri (Sort) - Removing unnecessary items from working area
- Seiton (Straighten, Set) – Storing items in an orderly way for easy maintenance
- Seiso (Shine, Sweep) – Maintaining a clean environment so that problems can be easily identified
- Seiketsu (Standardize) – Define standards across areas
- Shitsuke (Sustain) – Maintaining standards for the long-term

While this project is not related to work performed in a physical location, the 5Ss are still relevant when applied to the virtual environment and data storage.

Finally, Six Sigma relies on the DMAIC framework which will be used as the foundation for this research:

- Define **–** The main priority of the define phase is to clearly articulate and refine the problem including why it is a problem. While it may sound trivial and an overly burdensome administrative task to dedicate an entire phase of the lifecycle to simply defining the problem, this phase is crucial in establishing clear objectives and aligning stakeholders early on in the process of what the issue is that is trying to be solved.
- Measure - The measure phase builds off of the define phase by gathering data and establishing a baseline of the current state. The data that is gathered during the measure phase will be critical to properly analyze the problem in the next phase. As it is often said in regards to data quality, "Garbage in, garbage out."

- Analyze – While tempting to jump straight to the improve phase, the anlayze phase forces initiatives to get to the true root cause of issues and not make costly assumptions that can lead to subpar results.
- Improve – The improve phase begins by determining potential solutions, then prioritizing and deciding which ones to implement, and finally implementing the solutions. It is helpful during the improve phase to consider the final control phase and ensure that processes are engineered in a way that can be easily controlled.
- Control – Finally, the control phase aims to ensure the improvements made during the initiative are sustainable. A big component of this phase includes change management which builds off of the very first phase define by aligning stakeholder priorities to the newly refined process.

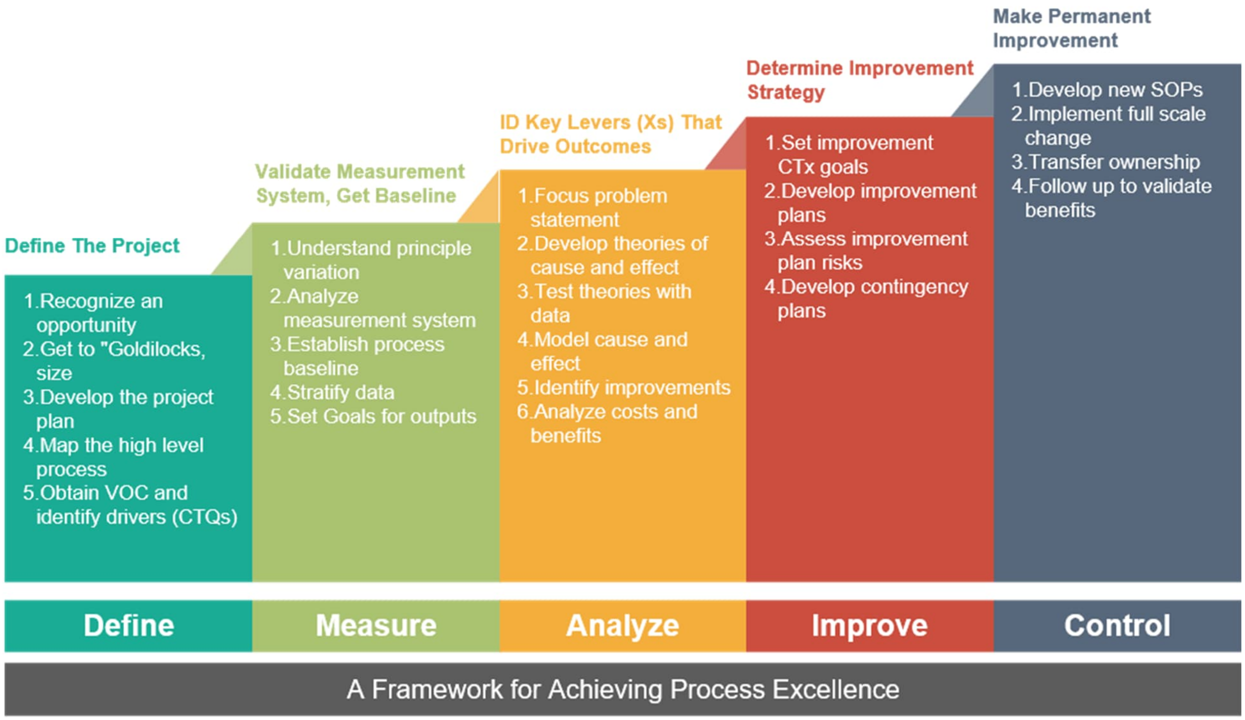


Figure 1 DMAIC Framework for Achieving Process Excellence through Lean Six Sigma

1.1.3 eDiscovery

Complete Discovery Source defines eDiscovery as “the electronic aspect of identifying, collecting and producing electronically stored information (ESI) in response to a request for production in a law suit or investigation. ESI includes, but is not limited to, emails, documents, presentations, databases, voicemail, audio and video files, social media, and web sites.” (The Basics: What is e-Discovery?, n.d.) Said differently, eDiscovery is the process of identifying and collecting relevant data, typically unstructured, and preparing it in a way that it can be easily searched and reviewed so that relevant information can be produced to the opposing party, typically a regulating body such as the Department of Justice (DOJ), Securities Exchange Commission (SEC), or Public Company Accounting Oversight Board (PCAOB) here in the United States.

The Electronic Discovery Reference Model (EDRM) provides a framework for the process of data as it flows through the lifecycle (Electronic Discovery Reference Model, 2020). As you move across the model from left to right, the volume of data increases while the relevance of the remaining data increases. Said differently, as data moves across the EDRM it is continuously analyzed and culled so that the most important pieces of data for each matter are analyzed in additional detail.

It is important to note that the EDRM model while presented linearly more often than not is performed concurrently. Additionally, one may not need all of the steps outlined in the diagram. For example, data is often identified and preserved for a matter but never processed, reviewed, or produced, if a settlement is reached. Alternatively, it is common during the review phase of an engagement to identify an additional relevant custodian and go back and collect and process his or her data for additional review.

Additionally, the EDRM is an iterative process that can have many phases being performed simultaneously. For example, one may be reviewing data that has already been identified, preserved, collected, and processed when they find a key piece of data involving an employee whose data was not originally collected. In eDiscovery, this new employee would be referred to as a custodian, defined as a person “having administrative control of a document or electronic file" (Microsoft, 2020). It is quite common for the scope of custodians to expand as review is completed. When this happens, the review of data can continue while the new custodians’ data is collected.

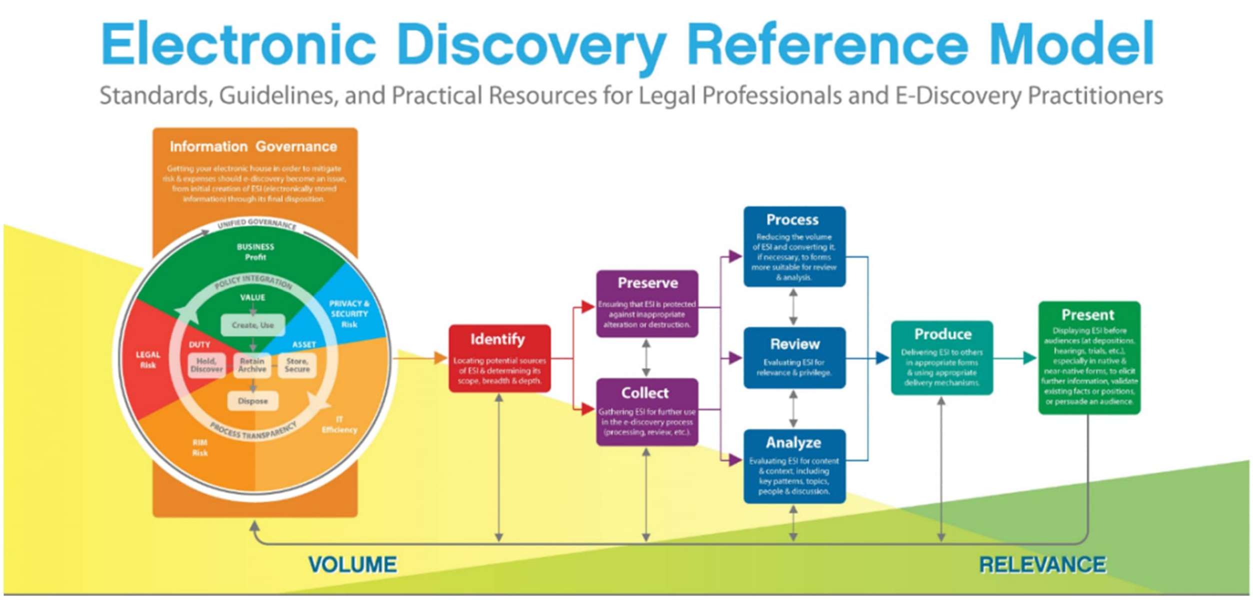


Figure 2 Electronic Discovery Reference Model (Electronic Discovery Reference Model, 2021)

1.1.3 Information Governance

The EDRM begins with an Information Governance (IG) model. Information governance can be defined as “the specification of decision rights and an accountability framework to ensure appropriate behavior in the valuation, creation, storage, use, archiving and deletion of information. It includes the processes, roles and policies, standards and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals (Gartner Glossary, 2021)." IG is the foundation for organizations to be prepared should eDiscovery arise and includes the entire data lifecycle – from the creation of data through its disposition. It includes not only data but also an organization's technology and related processes, policies, and strategies.

1.1.4 Identification

Identification is the process to identify potentially relevant pieces of information that can include ESI for individual custodians or enterprise systems. Additionally, paper sources, such as notebooks, printed binders, etc., can also be deemed relevant and are typically scanned so that they are then stored electronically. The aim of the identification phase is for the legal team to develop and execute a plan to identify and validate potentially relevant ESI sources including people and systems (EDRM, 2010).

Figure 3: Identification Phase Workflow (EDRM, 2010)

1.1.5 Preservation

In short, duty of preservation is triggered when “litigation is filed, threatened, or is reasonably anticipated.” (American Bar Association, n.d.) This is typically achieved by placing a legal hold in place which is a written directive to avoid the destruction of any potentially relevant information. Typically, custodians are alerted of a legal hold, and Information Technology (IT) departments can create, or modify existing, back-up policies to preserve relevant information. Data should be preserved in a way that is designed to be “legally defensible; reasonable; proportionate; efficient; auditable; broad but tailored; mitigate risks.” (EDRM, 2020)

Figure 4: Preservation Phase Workflow (EDRM, 2020)

1.1.6 Collection

Unsurprisingly, collection is the process of acquiring data that is identified during the identification phase and meets the same requirements of the data preservation phase: legally defensible, proportionate, efficient, auditable, and targeted. The Sedona Conference is a nonprofit organization that includes legal and technology professionals to study the advancement of eDiscovery. Sedona Principle 6 states that “responding parties are best situated to evaluate the procedures, methodologies, and technologies appropriate for preserving and producing their own electronically stored information." (19 Sedona Conference, 2018) This means that the producing party is responsible for determining the appropriate method for data collection, in addition to other processing, review, and production activities. The EDRM Collection Guide defines the process as follows:

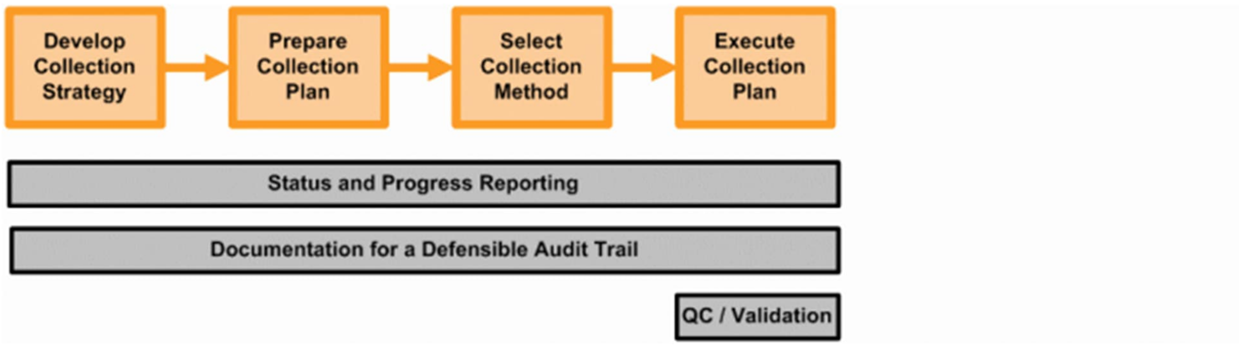


Figure 5: Collection Phase Workflow (Electronic Discovery Reference Model, n.d.)

There are several tools in the industry to collect data into what is known as a forensic image. The tool used is heavily dependent on the type of data being collected but common tools include EnCase, AccessData-FTK, and Macquisition.

1.1.7 Processing

The next phase of the EDRM is processing, which will be the primary focus of this research project. In short, processing includes the transformation and preparation of collected data into a format that is ready for searching and ultimately review. This typically includes:

- Extraction of key metadata fields such as To, From, CC, BCC for emails, author for documents, and relevant date fields into a structured database for searching
- Extraction of file content text and subsequent indexing for searching

- Flattening of files so that files are available outside of their containers such as word documents in a zip folder or attachments to an email
- Exclusion of non-relevant system files
- Deduplication which is the process of identifying the unique document population
- Migration to a review platform

The EDRM defines processing's objective to "perform actions on ESI to allow for metadata preservation, itemization, normalization of format, and data reduction via selection for review." (EDRM, 2020)

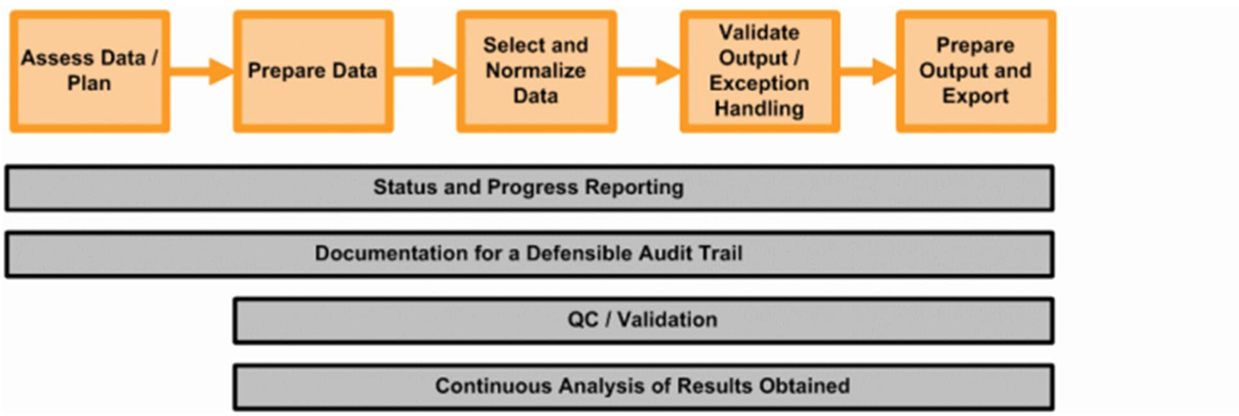


Figure 6: Processing Phase Workflow (EDRM, 2020)

1.1.8 Review and Analysis

Review is typically handled by attorneys, paralegals, or other legal professionals and aims to "gain an understanding of document content while organizing them into logical sub-sets in an efficient and cost effective manner. Develop facts, reduce risk, reduce cost, leverage technology, facilitate collaboration and communication." (EDRM, 2010)

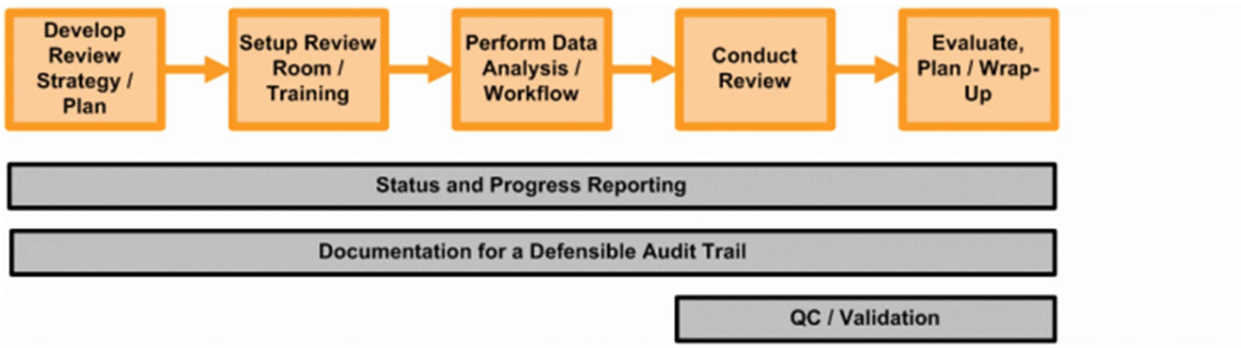


Figure 7: Review Phase Workflow (EDRM, 2010)

In short, review is the process of identifying relevant documents to produce and privileged documents to withhold.

1.1.9 Production and Presentation

Finally, data is produced in "an efficient and usable format in order to reduce cost, risk and errors and be in compliance with agreed production specifications and timelines" (Production Guide, 2010)

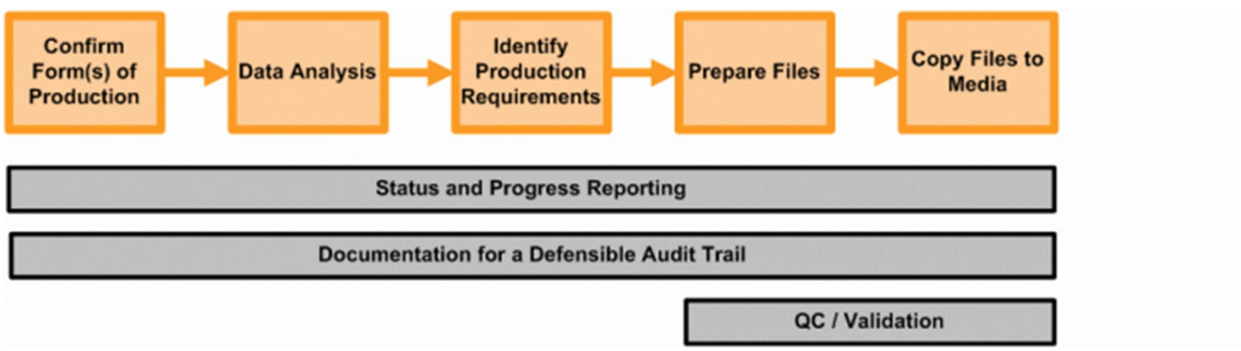


Figure 8: Production Phase Workflow (Production Guide, 2010)

Productions can have many formats but tend to stick to standard specifications defined by the Department of Justice (DOJ), Security and Exchange Commission (SEC), and the Company

Accounting Oversight Board (PCAOB) that include native files, itemized image files, extracted text files, and metadata load files.

1.1.10 EDRM Summary

In short, the EDRM can be summarized into the phases below:

1. **Identification –** Locating potentially relevant sources of ESI and any relevant paper sources including determining its scope, breadth and depth.
2. **Preservation –** Ensuring that ESI is properly maintained and protected against inappropriate modification or destruction.
3. **Collection –** Forensically capturing ESI for further use in the EDRM processes including processing, review, analysis, production and preparation.
4. **Processing –** Preparing collected ESI for review including data culling and easily consumable exported data.
5. **Review –** Evaluating processed data for case needs often including reviewing for responsiveness, privilege, and applying any necessary redactions.
6. **Analysis –** Evaluating ESI for content & context, including key patterns, topics, people and discussion for the case or investigation strategy
7. **Production –** Delivering ESI to the opposing party or regulatory body in appropriate forms, using appropriate delivery mechanisms.
8. **Presentation –** Displaying ESI before audiences (at depositions, hearings, trials, etc.), especially in native & near-native forms.

For the purposes of this research project, we will be focused on the process of preparing data for review which includes both the processing and review phases.

1.1.11 Trends in the eDiscovery Industry

Market Size

The report "eDiscovery Market by Component (Software (Processing, Review and Analysis, Identification, Preservation and Collection, and Production and Presentation) and Services), Deployment Type, Organization Size, Vertical, and Region - Global Forecast to 2025" claims the global eDiscovery market size is expected to grow from USD 9.3 billion in 2020 to USD 12.9 billion by 2025, at a Compound Annual Growth Rate (CAGR) of 6.6% during the forecast period with the majority of growth in North America. This is due to a variety of reasons including:

- Increased regulations under democratic administrations
- Filled vacancies in the Department of Justice (DOJ) and Securities Exchange Commission (SEC)
- Increase in whistleblowers due to rewards and potential compensation through the Dodd Frank Act
- Exponential growth in data volumes resulting in rise in costs in Discovery due to scale

While the industry may expect to grow ~6.6% in revenue, year over year, it's important to contextualize that growth within the pace at which data is growing. Christo Petrov identified 25 impressive data growth stats, of which I'll highlight my favorites below (Petrov, 2021):

- In 2020, every person generated 1.7 megabytes in just a second.
- Internet users generate about 2.5 quintillion bytes of data each day.
- 95% of businesses cite the need to manage unstructured data as a problem for their business.
- 80 – 90% of the data we generate today is unstructured

Technology Advances

Unsurprisingly, to keep pace with the increase in market and data size, the industry has made several advances in technology:

- Remote Collections – Historically data collections consisted of physical media devices such as desktop machines, laptops, and mobile phones. With the transition of data to the cloud and reduction in data storage costs, it is now more common for companies to have technologies in place to not only collect data but place legal holds and preservation measures in place as well.
- Targeted Collections – With advances in data collection platforms, such as Microsoft's O365 Compliance Suite, data can now be available to be searched prior to collection. For example, a case involving a pyramid scheme that took place within a two-year period can have emails exported that were sent only within that time period and only if they contained relevant search terms.
- Emerging Data Sources – With the emergence of new technologies and sources of data, such as the myriad of sources within the Internet of Things, the eDiscovery industry has to be prepared to not only identify and collect these data sources but also host them in a review-friendly way.
- Shadow IT – Similar to the emerging data sources, a seasoned eDiscovery professional knows they need to consider the applications employees may be using that aren't officially sanctioned by organization IT departments. Excluding these sources can leave major gaps in investigation data sets.
- Scalability – Finally, the industry must continuously evolve to meet the demands of the ever-increasing data footprints of today's organizations.

Analytics

Finally, while technology advances have assisted with the collection and preparation of the increased data volumes seen in recent years, analytics have been used to help consume this information in an efficient manner. The most common type of analytics used in Discovery matters is often referred to as Technology Assisted Review, or TAR which aims to predict whether a document is relevant or not. The ultimate goal is to build a model that gives reviewers confidence that additional documents reviewed will most likely not be responsive. There are typically 2 agreed upon types of models – TAR 1.0 which is supervised and TAR 2.0 which is unsupervised, or as it is more commonly known in the industry as “Continuously Active Learning.” This can also be referred to as “predictive coding.”

TAR is one of the simplest use cases for text mining because relevance, or responsiveness, is typically a broad concept based on if documents are related or not. The criteria are not often overly complex, so it provides a simple use case to provide a well-trained model. For example, in one of my recent matters we have been able to reduce the volume of documents needed for review by 99.8%. This leads to a significant cost savings since documents are typically reviewed by attorneys.

In addition to predictive coding, we also use the types of analytics below to streamline review:

- Communication analysis - Displays complete networks of communication based off email metadata and allows users to filter by person, domain, CC, BCC, etc.

- Clustering - Visually organizes the data set based on conceptual similarity and allows users to navigate the clusters like a map, quickly identifying neighborhoods of related documents vs. one document at a time
- Concept searching - Automatically expands terms, phrases, paragraphs, or even entire documents to reveal related concepts that may have been previously unknown to the user
- Term expansion/ optimization - Allows users to type in a term or list of terms and returns other terms closely related to those within the data set
- Textual near duplicate analysis - Identifies textually similar documents and groups them together; frequently used to assist in organizing the review process and increasing speed and consistency of review
- Email threading - Gathers an original email and all subsequent replies and forwards pertaining to that original email and determines which emails are inclusive (contain unique content not included in any other email); can be used to organize review or reduce review by nominating only the most inclusive email thread for review
- Language identification - Examines the extracted text of each document to determine the primary language and up to two secondary languages present

1.2 Research Aims, Objectives, and Questions

Given the lack of research regarding both RPA and Lean Six Sigma, this study will aim to evaluate the intersection of these two topics specifically applied throughout the EDRM from data ingestion through review.

1.2.1 Research Objectives

- RO1 **–** To apply lean six sigma principles using the DMAIC framework to the EDRM lifecycle as a process improvement initiative with an emphasis on RPA

- RO2 – To evaluate the effectiveness of the process improvement from a financial, speed, and quality perspective
- RO3 – To provide a methodology to efficiently transform unstructured data into a structured state with minimal human intervention

1.2.2 Research questions

In Lean Six Sigma projects, the problem statement clarifies the situation by identifying the problem, its severity, location and financial impact. This problem can be represented as a general statistical problem with a fully inclusive null and alternative hypothesis where the burden of proof is showing that improvement was really achieved:

Ho: There was no improvement to the output of the process.

Ha: There was improvement in the output of the process.

1.3 Significance

This project has two distinct areas of significance – the rise of RPA deployments and adoptions as well as the rising pressure to reduce legal costs.

1.3.1 RPA Adoption

According to Gartner, RPA ecosystem was the fastest-growing segment of the global enterprise software market in 2018. It is no surprise given that according to Automation Anywhere, employees spend 10%-25% of their time on repetitive computer tasks and that a typical rules-based process can be 70%-80% automated (Automation Anywhere). Although RPA can be a significant enabler to drive innovation, organizations risk burdening IT resources if inefficient processes are automated. While an organization may have the resources to handle a process that takes a human ~30 minutes to complete, the same infrastructure may become overburdened if it is automated to be performed every 30 seconds. Additionally, if a process is error-prone, those defects will be exacerbated by a process that is automated to be repeated. It is always better to eliminate unnecessary steps than to automate them.

1.3.2 eDiscovery Costs and Legal Pressure

In January 2021, EY together with the Harvard Law School Center on the Legal Profession conducted 2,000 interviews within law departments to better understand the opportunities and challenges (Ernst & Young, 2021). One of the insights driven from the survey: Workloads are increasing faster than budgets meanwhile law departments are planning to even more ambitiously cut costs. The survey found that General Counsel expect workloads to increase roughly by 25% over the next three years and sadly 75% of these same respondents don't expect budgets to keep pace. Interestingly enough, this same survey also identified that C-suite executives are not supportive of critical investments in legal technology and process improvements, despite many of these cost reductions forcing operational change. General Counsel, on the other hand, report that increased use of technology offers the greatest opportunity for cost savings but they face challenges securing budget for technology and process improvement from the C-suite.

1.4 Limitations

This project has defined its limitations into four categories: scope, research methodology, resources, and applicability.

1. **Scope:** This research is focused on the ingestion, transformation, and export of unstructured data for eDiscovery purposes. It does not include research into other methods of accelerating the EDRM lifecycle including Early Case Assessment (ECA) which filters data prior to review promotion or Technology Assisted

Review (TAR) which uses text analytics to limit the review of non-responsive documents.

2. **Research methodology:** This project could be criticized for oversimplifying the EDRM by operating under the assumption that the data to be processed will be processed as expected using the available tools.

3. **Resources:** This research is limited to processing through approximately one petabyte of data and limited in terms of manual review of individual files.

4. **Applicability:** This research is limited in focus to a process specific to unstructured data that is only reviewed for relevance.

1.5 Structure

The praxis aims to describe the phases of the research and report the research findings in the chapters: Introduction, Literature review, Methodology, Results, and Conclusions. This introduction aims to provide a brief background about the problem that justify the study and its significance. The literature review found in chapter two includes an overview RPA and specifically discusses references in the literature to Lean Six Sigma, unstructured data, and references to the legal industry. Chapter three includes the research methodology including project selection and analysis methods, primarily the Lean Six Sigma DMAIC framework. Chapter four includes the results of this analysis including control charts and factor analysis to measure the financial, speed, and quality benefits of this initiative. Finally, the conclusions in chapter five include observations of impact and potential for further research.

CHAPTER 2

LITERATURE REVIEW

2.1 Outline

This literature review begins by providing background into Robotic Process Automation including its definition and background. It then explores RPA and unstructured data, RPA and Lean Six Sigma, and published case studies including a deeper dive into the legal industry. Finally, the review concludes with the assertion that based on available publications this praxis is timely and relevant to both academia and the industry.

2.2 RPA Definition and Background

RPA is an umbrella term for tools that operate on the interface of computer systems and applications in the way a human would do. RPA aims to replace people by automation done in an "outside-in'' manner on manual, repetitive tasks (van der Aalst, 2018). Gartner defines RPA as "a productivity tool that allows a user to configure one or more scripts (which some vendors refer to as "bots") to activate specific keystrokes in an automated fashion. The result is that the bots can be used to mimic or emulate selected tasks (transaction steps) within an overall business or IT process. These may include manipulating data, passing data to and from different applications, triggering responses, or executing transactions. RPA uses a combination of user interface interaction and descriptor technologies. The scripts can overlay on one or more software applications." (Gartner, 2021)

Precursors to RPA include screen scrapers, workflow automation and management tools, and Artificial Intelligence (AI). The term can first be traced back to the early 2000s but leading companies including UiPath, Automation Anywhere, and Blue Prism, began gaining traction around 2013. According to Gartner, by 2018 RPA platforms make up the largest growth in software development companies. Anagnoste takes the adoption one step further and claims we are in the midst of the "Robots Revolution." (Anagnoste, 2017)

2.3 RPA and Unstructured Data

In 2018, Gartner published an article titled "Beyond Tactical RPA: Planning a Strategic Automation Roadmap that Leads to Long-Term Value." (Gartner, 2018) The report explicitly states "processes with unstructured data are not suitable for most RPA tools." It then elaborates with its first guideline of forming an automation roadmap that the processes for consideration should have structured digitalized data. In 2021, Gartner has stuck to its assessment that RPA is best suited for structured data. (Gartner, 2021).

The academic publications align with Gartner's recommendations. In 2019, Syed et. Al asserts that a "major constraint for successful RPA deployment is the mandatory requirement for structured data." (Rehan Syeda, 2019) Within this literature review, 27 different academic articles were reviewed which directly asserted either the need for data to be structured or specifically claimed that unstructured data was not a candidate for automation. These references are included in Table 1: References to Structured Data Requirement within Literature. Other publications, including the Institute for Robotic Process Automation claim that the handling of unstructured data is "futuristic." (IRPAAI 2021)

Table 1: References to Structured Data Requirement within Literature

Year	Title
2016	Robotic process automation at telefónica O2
2016	Turning robotic process automation into commercial success - Case OpusCapita
2017	Automation of a business process using robotic process automation (RPA): A case study
2017	Resolving tussles in service automation deployments: Service automation at Blue Cross Blue Shield North Carolina (BCBSNC)
2017	Robotic Automation Process - The next major revolution in terms of back office operations improvement
2017	Robotic process automation: Strategic transformation lever for global business services?
2017	Software bots -The next frontier for shared services and functional excellence
2018	A new season in the risk landscape: Connecting the advancement in technology with changes in customer behaviour to enhance the way risk is measured and managed
2018	Apply RPA (Robotic Process Automation) in Semiconductor Smart Manufacturing
2018	Artificial intelligence in clinical development and regulatory affairs – Preparing for the future
2018	Artificial intelligence in smart tourism: A conceptual framework
2018	Automation in recruitment: a new frontier
2018	Delineated Analysis of Robotic Process Automation Tools
2018	How do machine learning, robotic process automation, and blockchains affect the human factor in business process management?
2018	How OpusCapita used internal RPA capabilities to offer services to clients
2018	How to choose between robotic process automation and back-end system automation?
2018	Identifying candidate tasks for robotic process automation in textual process descriptions
2018	Innovation in Pharmacovigilance: Use of Artificial Intelligence in Adverse Event Case Processing
2018	Minimal effort requirements engineering for robotic process automation with test driven development and screen recording
2018	Multi-Perspective process model discovery for robotic process automation

Year	Title
2018	Process mining and Robotic process automation: A perfect match
2018	Robotic Automation Process - The operating system for the digital enterprise
2018	Robotic process automation - Creating value by digitalizing work in the private healthcare?
2018	SNS Door Phone as Robotic Process Automation
2018	Survey of Drones for Agriculture Automation from Planting to Harvest
2018	The key factors affecting RPA-business alignment
2018	Towards a Process Analysis Approach to Adopt Robotic Process Automation

2.4 RPA and Lean Six Sigma

Lean Six Sigma principles are an established framework within the literature so for the purpose of this literature review, only works that mentioned both RPA and Lean Six Sigma were reviewed. Within the academic publications, only 3 academic publications were identified that mentioned both Lean Six Sigma and RPA. Anagnoste mentions the "need for future improvement in the area of multi-tenancy, context awareness, adherence to supporting methodology (e.g. Six Sigma)." (Anagnoste 2017) Forrester also notes the overlap between departments focused on Lean Six Sigma and automation (Forrester 2021). No case studies were identified that highlighted successful deployments of RPA together with the Lean Six Sigma framework.

2.5 RPA Case Studies

In addition to the workflow functionality, a review was also done of published case studies. Current research indicates a focus on back-office activities and specific industries including healthcare, banking, and insurance (Deloitte 2017). Published academic case studies include:

- OpusCapita's deployment in the Purchase-to-Pay and Order-to-Cash, also known as Quote-to-Cash, processes (OpusCapita 2016)

- Recruitment process and the Human Resource Management Services (Gupta 2018)
- HR and Global business service operations (Wilcocks 2017)
- Blue Cross Blue Shield's development of a dedicated automation team (Dunlap (2017)
- The processing of adverse event reports by a pharmaceutical company (Schmider 2019)
- Shared Services' software bots and a focus on centralization (Wilcocks 2021)
- The tourism industry's development of "smart destinations" (Tsaih 2018)
- Applications in the IT sector (Isaac 2018)
- A smart door phone that sends user alerts (Koboyashi 2018)
- The strategic deployment of drones for agriculture automation from planting to harvest (Kulbacki 2018)
- Remote control of equipment in semiconductor manufacturing (Lin 2018)

In the legal industry, there are references for the potential for RPA but little mention of successful deployments. Forbes asserts that while law firms are strong candidates to benefit from RPA most have "barely scratched the surface in exploiting the potential of RPA." (Bhutta, 2020) Exigent published eDiscovery as a potential for automation in May 2020 (Exigent, 2020). Unsurprisingly, the earliest reference to RPA in the eDiscovery industry identified belongs to the author and is an innovation award submission from September 2018 (Relativity, 2018).

Specific to eDiscovery, Nelson Mullins introduces Lean Six Sigma and how eDiscovery can benefit from it but without supporting implementation. (Encompass Contributor, 2017) Evergreen Editions publishes Lean Six Sigma but with the focus on quality of review, not gaining efficiencies in the flow of data. (Blair, 2011) In 2018, TCDI summarizes this author's feelings by declaring "Yes, everyone throws around the terms LSS, lean, Six Sigma and

DMAIC, but it seems most people can’t be bothered to really use it. This is a big mistake.” (TCDI, 2018)

2.6 Conclusion

Based on the identified literature, there appears to be an opportunity to fill a gap in the research by applying Lean Six Sigma principles to RPA initiatives. Furthermore, it is exciting to provide an alternative solution to the publications that cite the need for data to be structured to be a candidate for an RPA solution.

CHAPTER 3

METHODOLOGY

3.1 Outline

This Praxis is the culmination of the engineering principles learned within SMU's Operations Research and Engineering Management department (formerly known as Engineering Management and Information Systems) applied to a real-world problem identified within industry - specifically within Ernst & Young's Forensics practice. This chapter discusses the research methodology and aims to provide an overview of how the research was conducted and providing the steps and procedures to address the hypothesis to provide proof of process improvement. This chapter begins with project selection and analysis methods. From there, this report provides a deeper dive into the steps of the Lean Six Sigma DMAIC framework which provided the foundation for the process improvement initiative. Finally, this chapter describes the analytical methods determined to provide the results within the next chapter.

3.2 Project Selection

The American Society of Mechanical Engineers (ASME) defines eight domains of engineering management through its Book Committee's 2010 "Guide to the Engineering Management Body of Knowledge." (ASME Book Committee, 2010) These eight domains include:

- **Domain 1: Market Research, Assessment and Forecasting –** This domain describes the processes and activities involved in conducting market research which includes an overview of the market analysis process and provides guidance for making recommendations and communicating the results of the study to nonmarketing personnel. Environmental scanning techniques presented include benchmarking practices and processes, business forecasting methods, risk analysis strategies and trend analysis techniques.

- **Domain 2: Strategic Planning and Change Management –** This domain addresses the internal and external processes used to bring a new product or technology to market and includes methods for planning and implementing new technologies in new global destinations. Knowledge of sources available to conduct competitive intelligence will assist engineering managers in making appropriate recommendations and crafting plans to ensure the new technology being introduced will capture the desired share of the new market.
- **Domain 3: Product, Service and Process Development –** This domain begins by identifying the appropriate engineering disciplines involved in product development and discusses how to apply various disciplines to interpreting research findings and development results. A complete section concerning manufacturability of a product provides knowledge of pilot production and feedback processes, product feasibility planning, assembly and disassembly procedures, Kaizan and Lean production techniques. It also includes quality analysis techniques such as Total Quality Management (TQM) and Six Sigma.
- **Domain 4: Engineering Projects and Process Management –** This domain includes techniques involved with establishing financial resource requirements, making financial projections, developing a budget and measuring return on investment for the product

production project. This includes project management fundamentals, project management planning and control concepts, scheduling and budgeting practices and assessment of legal liabilities.

- **Domain 5: Financial Resource Management –** This domain includes procurement and contract procedures and techniques for proper documentation of contracts including Uniform Contract Format (UCF). It also includes discussion of project funding, proposals, and budgeting and cash flow management techniques.
- **Domain 6: Marketing, Sales and Communications Management –** This domain discusses marketing practices and sales and advertising plans, including methods to reliably measure and improve customer satisfaction ratings are presented.
- **Domain 7: Leadership and Organizational Management –** This domain examines the management styles and organizational structures most conducive to managing professionals, including engineers, whose primary labor is intellectual (as opposed to manual). Areas of focus include: management systems, organization structure, people orientation, and the internal and external business environments they operate in.
- **Domain 8: Professional Responsibility, Ethics and Legal Issues –** This domain aims to provide an understanding and application of regulatory requirements, codes and standards including methods for communicating such standards.

In its totality, the Guide offers a comprehensive view of the roles and responsibilities that an engineering manager assumes from the cradle to the grave in the engineering production process. Figure 9 provides a more comprehensive breakdown of the 8 Domains of Engineering Management.

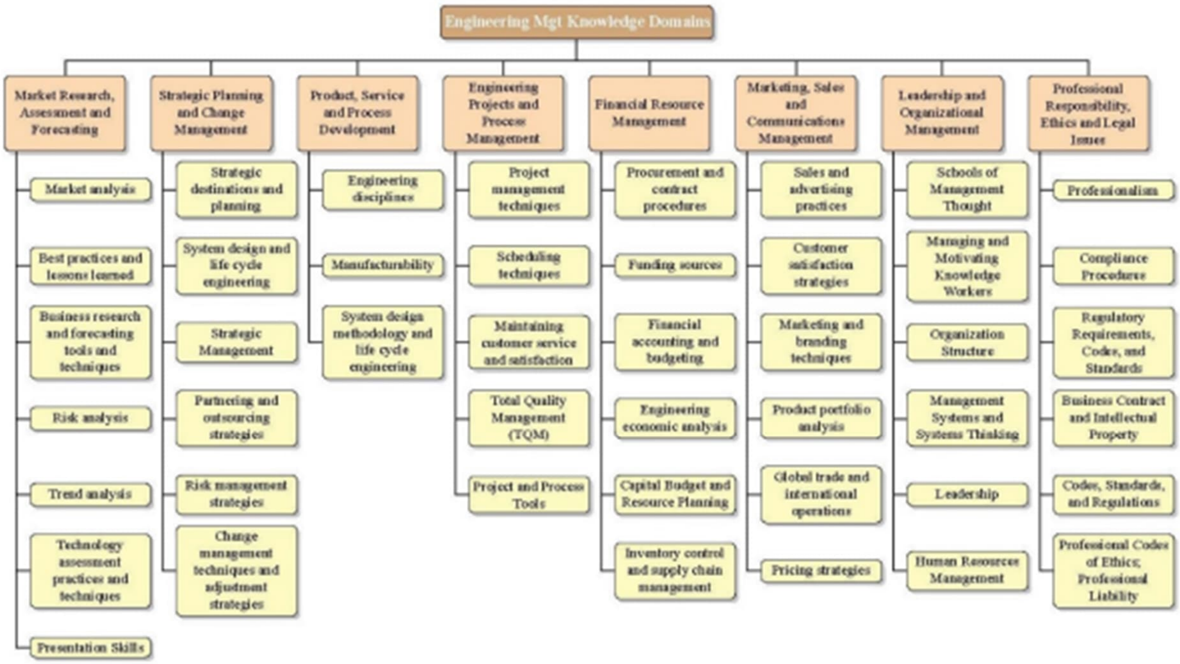


Figure 9: EMBOK's 8 Domains of Engineering Management

As a process improvement initiative, this research falls under the EMBOK's third domain: Product, Service and Process Development which specifically includes Six Sigma. It should be noted that as a Six Sigma project, particularly in the define and control phases, there are aspects of change management which would fall under the EMBOK's second domain – Strategic Planning and Change Management.

3.3 Research Methodologies

3.3.1 Lean Six Sigma Methodologies

Lean Six Sigma has two primary frameworks DMAIC and DMADV, depending on if the initiative is focused on improving an existing process or designing a new process. Their steps are summarized in Table 2: Lean Six Sigma Frameworks (Purdue University, 2021).

Table 2: Lean Six Sigma Frameworks

DMAIC (Existing Processes)	DMADV (New Processes)
Define - Define the problem, output to be improved, customers, and process associated with the problem.	**Define** - Define the process and design goals.
Measure - Collect data from the process to establish a baseline for the improvements.	**Measure -** Measure and identify critical-to-quality characteristics of the product, service or process. This includes risk and production capabilities.
Analyze - Analyze the data to find the root causes of defects.	**Analyze -** Analyze the data to find the best design.
Improve - Develop, test, and implement solutions to improve the process.	**Design -** Design and test the product, service or process.
Control - Implement process controls to sustain the improvements.	**Verify -** Ensure that the design output meets the design input requirements (verification) and that the designed product performs satisfactorily under real or simulated conditions of intended use (validation).

In addition to DMAIC and DMADV, there is also a less common method known as DMADVO which adds the additional step for Optimization after the DMADV process that recognizes the need to optimize design after implementation. Because this research focuses on improving an existing process, it follows the DMAIC framework as shown in Figure 10.

Figure 10: DMAIC Framework

3.3.2 Statistical and Data Analysis Methods

While Lean Six Sigma's DMAIC framework provides the methodology for the actual process improvement, additional statistical analysis methods are used to measure the results found in Chapter 4 to test the hypothesis. In addition to defining the 8 Domains of Engineering Management, the EMBOK also provides a framework for seven research methods within Engineering Management, defined in Figure 11.

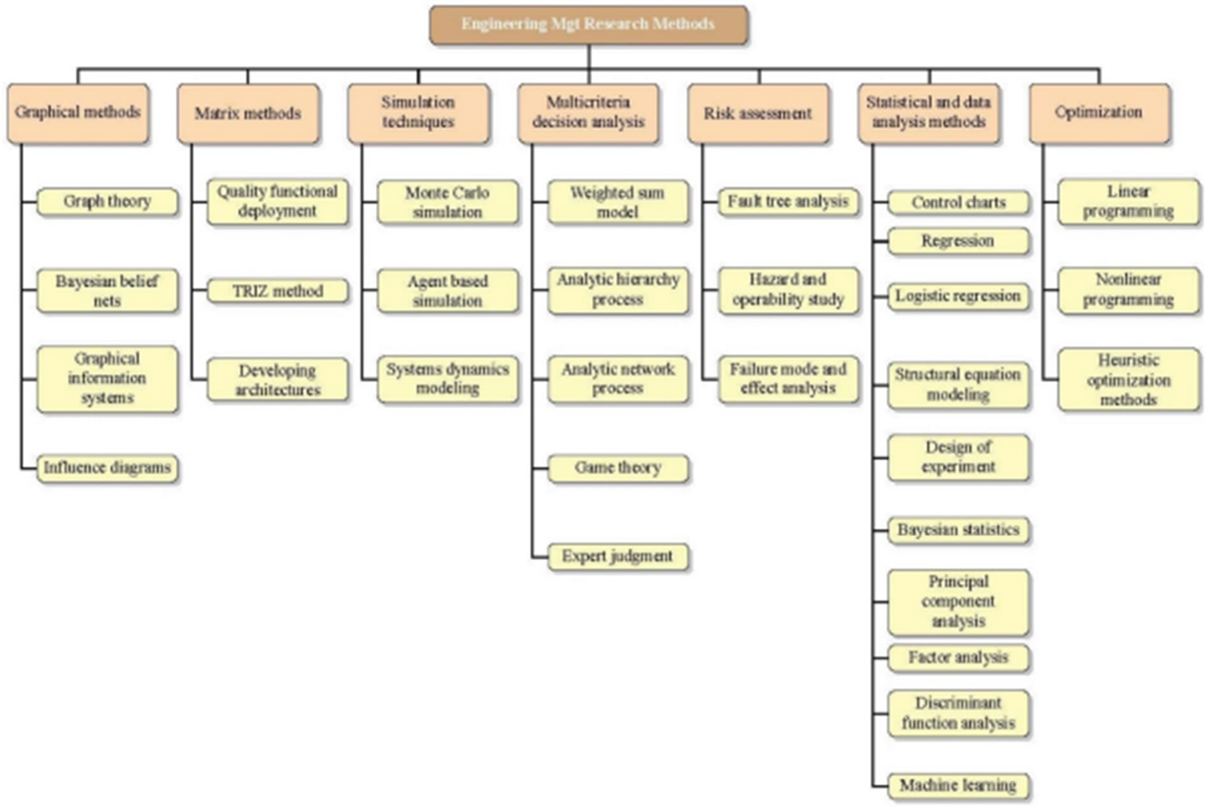


Figure 11: EMBOK's Research Methods

3.4 Define Phase

3.4.1 Define Phase Overview

The Define Phase is the first phase of the Lean Six Sigma DMAIC improvement process and aims to answer the question "What problem are you trying to solve?" As introduced in the first chapter, the main objective of the define phase is to clearly articulate and refine the problem statement of the initiative. By the end of this phase, teams should be able to answer the questions below:

- Who is your customer and what do they want?
- What is the problem you are trying to solve (or pain point)?

- What is the main measure of success?
- Does this solution already exist elsewhere within the organization?
- What are team roles and responsibilities?
- What is the high-level process?

3.4.2 Project Charter

The Project charter is typically the first step of a Lean Six Sigma project. It outlines the process improvement initiative for both the team and project leadership and should be continuously updated throughout the project. Project Charters can vary by organization, but typically include a Problem Statement, Business Case, Goal Statement, Timeline, Scope, and Team Members. Figure 12 shows the Project Charter for this initiative which uses EY's template. This charter also includes business measures, risks, and deliverables.

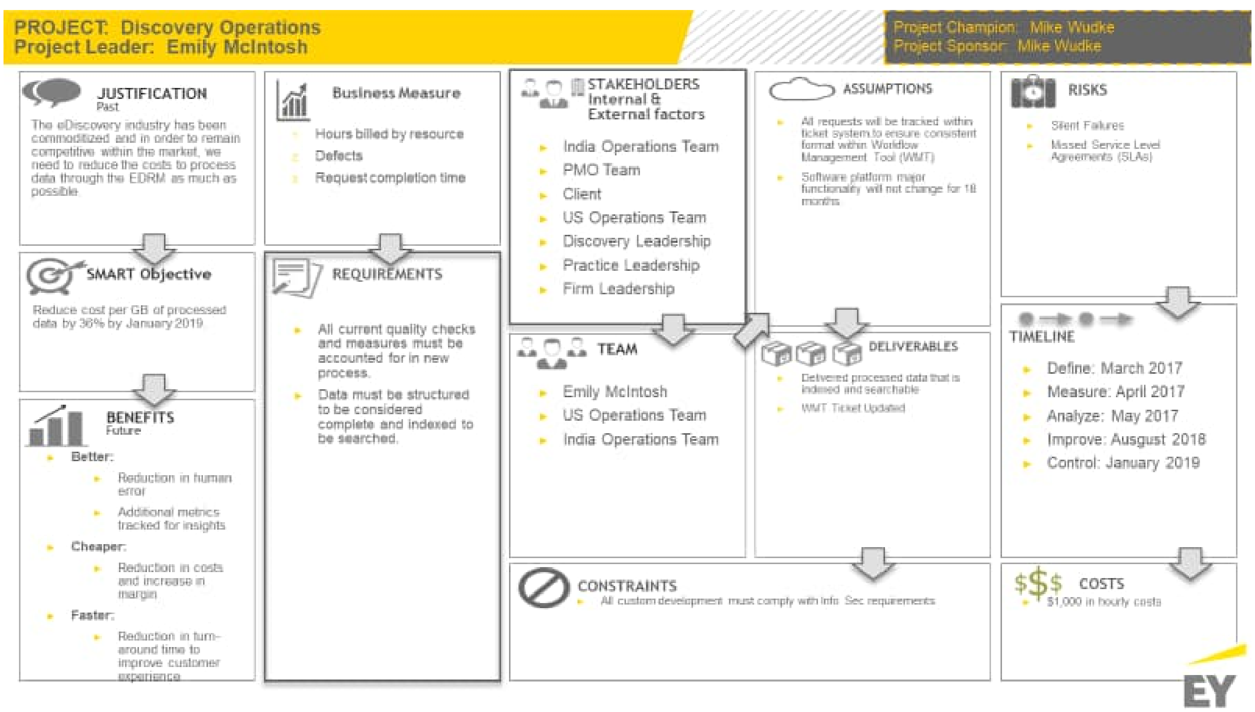


Figure 12: Project Charter

3.4.3 Customer Requirements

In order to address customer requirements, we must first identify who the customers are. Customers can be external or internal. For this initiative, three different clients with separate motivations were identified: our external clients who need their data processed and available for review, our internal Project Managers who are responsible for coordinating with our external clients, and finally our internal practice leadership who are focused on improving margin to keep our practice sustainable.

In order to meet customer needs, Lean Six Sigma projects rely on a variety of tools collectively referred to as VOC which stands for Voice of the Customer. Traditionally, VOC tools aim to help Lean Six Sigma teams identify Critical Customer Requirements also known as CTQs. For this project, the four questions below were answered for each of the three customers:

- **Customer:** Who is the customer?
- **VOC:** What does the customer want from us?
- **Key Customer Issues:** What does the customer want from us?
- **CTQ:** Summarize key issues and translate them into specific and measurable requirements.

As part of the VOC tools, some project teams may consider using a VOC translation matrix which helps translate customer comments into CTQs. Additionally, some teams may also consider using a VOC tree map which provides a methodology for breaking down general requirements into specific, measurable requirements. An example of a VOC Tree Diagram is given in Figure 13 which was developed by the organization GO Lean Six Sigma (GO Lean Six Sigma, n.d.).

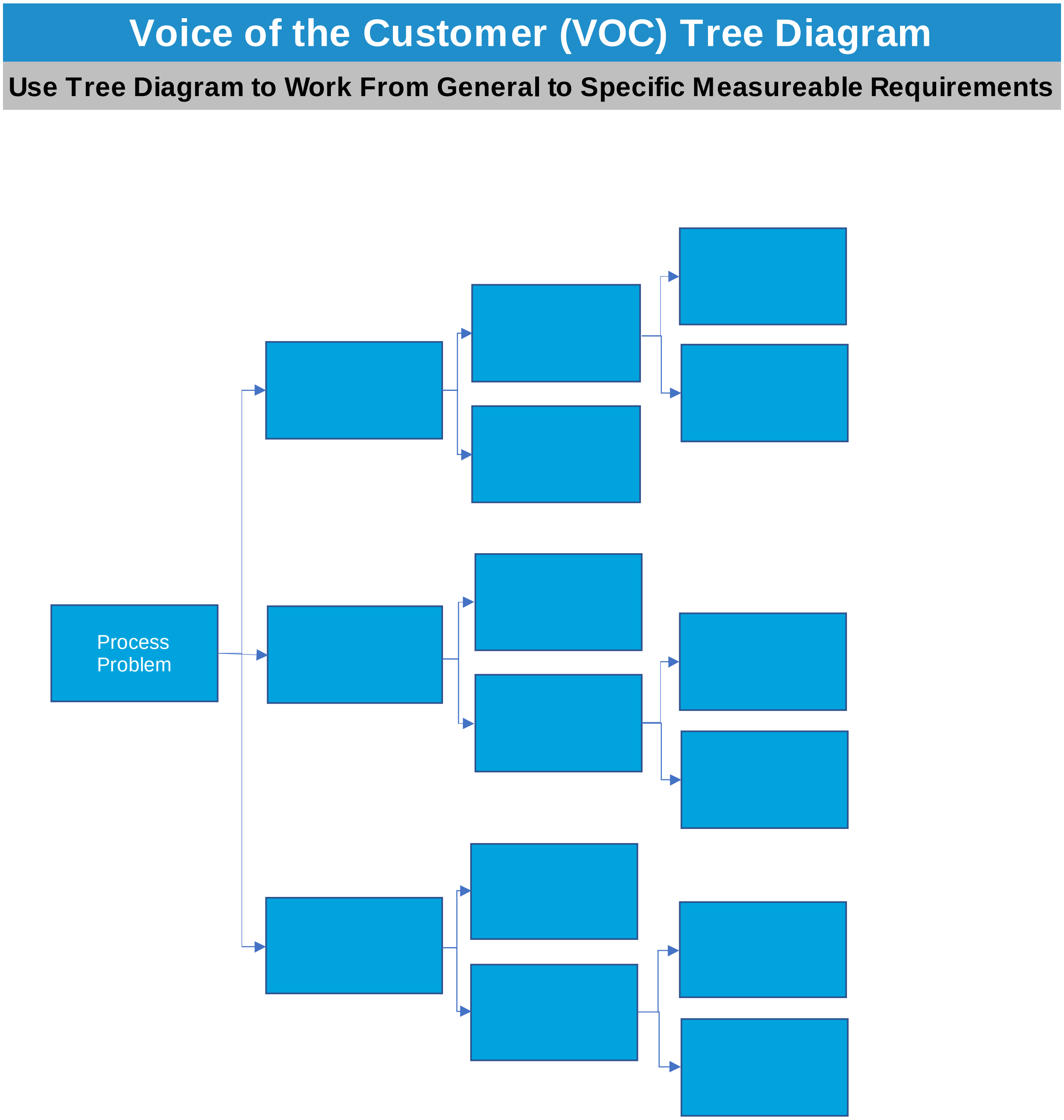


Figure 13: VOC Tree Diagram (GO Lean Six Sigma, n.d.)

3.4.4 Measure Phase Summary

Table 3 summarizes the EY Methodology used for the Define Phase of Lean Six Sigma projects.

Table 3: EY Define Phase Methodology Summary

Critical Questions	Standard Tasks	Required Tools
Who is your customer & what do they want?	Develop Questionnaire for VOC, gather known VOC data, Collect VOC Data / Define CTQ Driver Tree	Canvas and CTQC
What is the problem you are trying to solve (or pain point)?	Data collection	Canvas
What is the main measure of success?	Develop measure for determining success	Canvas SMART objective
Have you searched if the solution to this problem already exists within EY?	Review current projects	Slide describing search (EY Discover, TRACtion)
What are team roles & responsibilities?	Organize Team , First Team Meeting	RACI
What is the high level process?	Define the Process, Map High Level end-to-end Process	SIPOC

Lean Six Sigma also has other sets of tools to enable project teams throughout the Define Phase. The first example is an A3 report which is a one-page report on the background and current state of a project. Figure 14 is an example of a template that project teams can use to develop an A3 report.

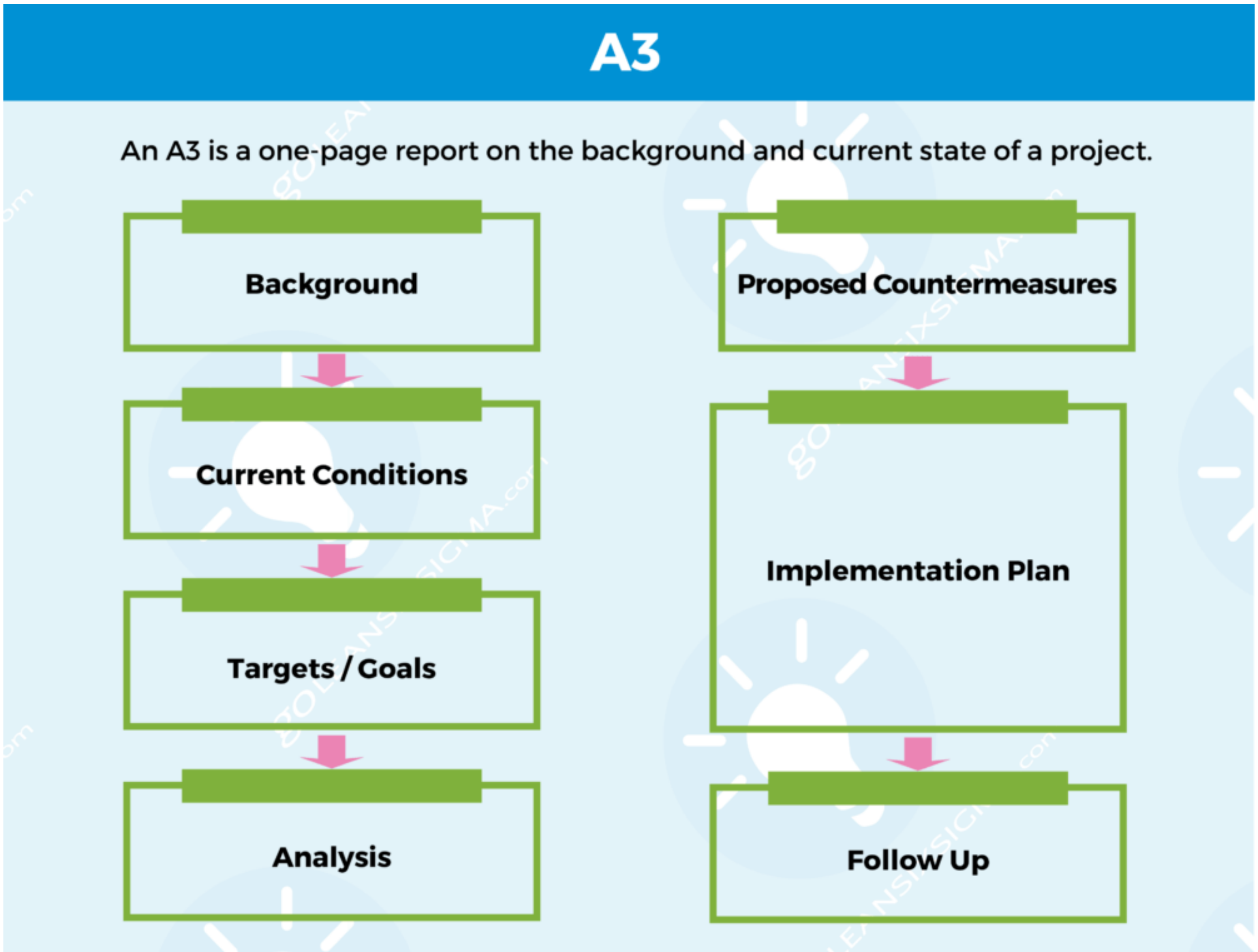


Figure 14: A3 Report Sample Template (GO Lean Six Sigma, n.d.)

Some teams may also find it useful to define a communication plan that defines stakeholders, their objectives, message content, delivery method, frequency, and target dates. Figure 15 provides an example of what teams can use to track their communication plans. For this project, communications and deadlines were tracked using Microsoft Outlook and a shared team calendar.

Communication Plan Examples				
Stakeholder/ Stakeholder Group	Objectives (Actions Desired)	Message Content	Delivery Method(s)/Venues	By When (Frequency)

Figure 15: Example Communication Plan (GO Lean Six Sigma, n.d.)

Teams may also find it helpful to document any perceived constraints and their corresponding assumptions as well as have a more structured way to define their goal statement. Figure 16 provides a sample template for defining constraints and assumptions; Figure 17 provides a methodology for defining a goal statement including the various parts of speech.

Constraints & Assumptions

Constraints	Assumptions	Process

Figure 16: Constraints & Assumptions Template (GO Lean Six Sigma, n.d.)

Goal Statement Builder		
Project Title:		
[Project Name]		
		Build the Goal Statement
The goal of this project is to: [VERB]	*Enter "Decrease or Increase"*	
The unit being measured in this project is: [UNIT]	*Enter unit*	
The measure for this unit is: [UNIT MEASURE]	*Enter Y Unit Measure*	
The estimated baseline of this measure is: [BASELINE]	*Enter Project Y Baseline*	
The estimated target of this measure is: [TARGET]	*Enter Project Y Target*	
The estimated due date for improvement is: [TARGET DATE	*Enter Estimated Target Date*	
Final Goal Statement:		
the from to by 1/0/1900		

Figure 17: Goal Statement Builder (GO Lean Six Sigma, n.d.)

During the Define Phase especially, teams will have many meetings and interviews as they coordinate with the various stakeholders. Tools can include a template for stakeholder analysis, team meeting agenda, meeting evaluation, and status report. Examples of each of these can be found in Figure 18, Figure 19, Figure 20, and Figure 21.

Stakeholder Analysis (Advanced)					
Stakeholder/ Stakeholder Group	Impact Level	Level of Support	Reason for Resistance or Support	Action(s) to Address This Stakeholder Group	Contact

Figure 18: Stakeholder Analysis Template (GO Lean Six Sigma, n.d.)

Team Meeting Agenda

Meeting For: ______________________ **Date & Time:** ______________

Facilitator: ____

Invitees: ____

Agenda Topics (Potential)	Time Alotted	Who	Notes
Ground Rules	*10 min*	*Sean*	*Update Original*

Figure 19: Team Meeting Agenda Sample (GO Lean Six Sigma, n.d.)

Meeting Evaluation					
Criteria	1 (Low)	2	3	4	5 (High)
Stick to the Agenda					
Manage Time					
Follow Ground Rules					
Balance of Participation					
Listen to Each Other					
Make Progress					
Other					

Figure 20: Meeting Evaluation Template (GO Lean Six Sigma, n.d.)

Status Report

Project	
Date:	
Contact:	
Prepared By:	

Status: (Red/Yellow/Green)	
Overall Status:	Green
Explanation:	

Key Accomplishments
List key tasks and milestones achieved in this period, along with any key changes

ID #	Key Issues	Action Plan	Status
			Open

Critical Milestones	Health	Due Date	Comment/Status
	G		
	G		
	G		
	G		
	G		
	G		
	G		
	G		

Figure 21: Status Report Template (GO Lean Six Sigma, n.d.)

To further assess how stakeholders may be impacted by the initiative, project teams may decide to define a RACI matrix to establish who is responsible, accountable, consulted, and informed for each task. They may also decide to develop a relationship map. An example of a relationship map can be seen in Figure 22.

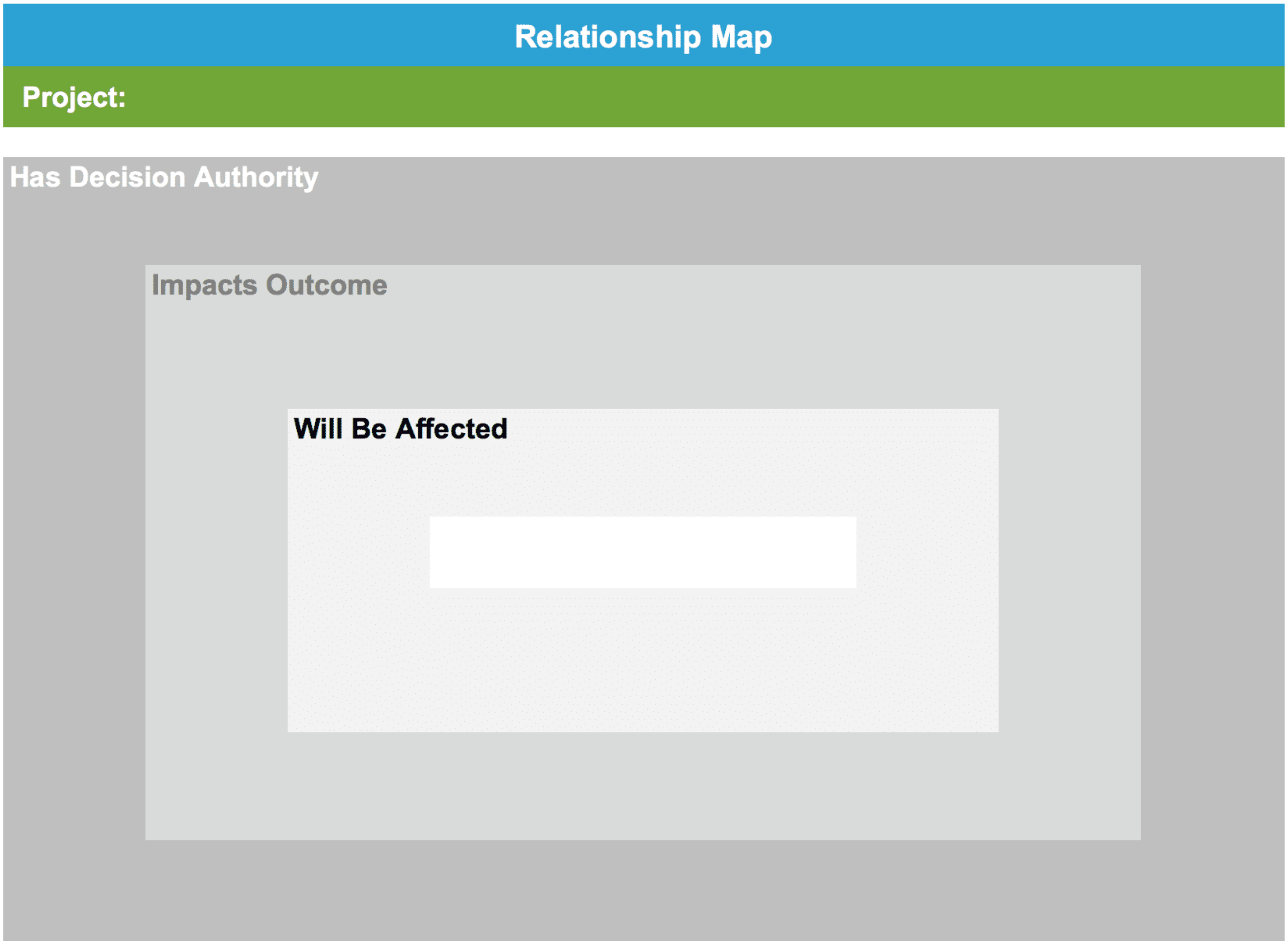
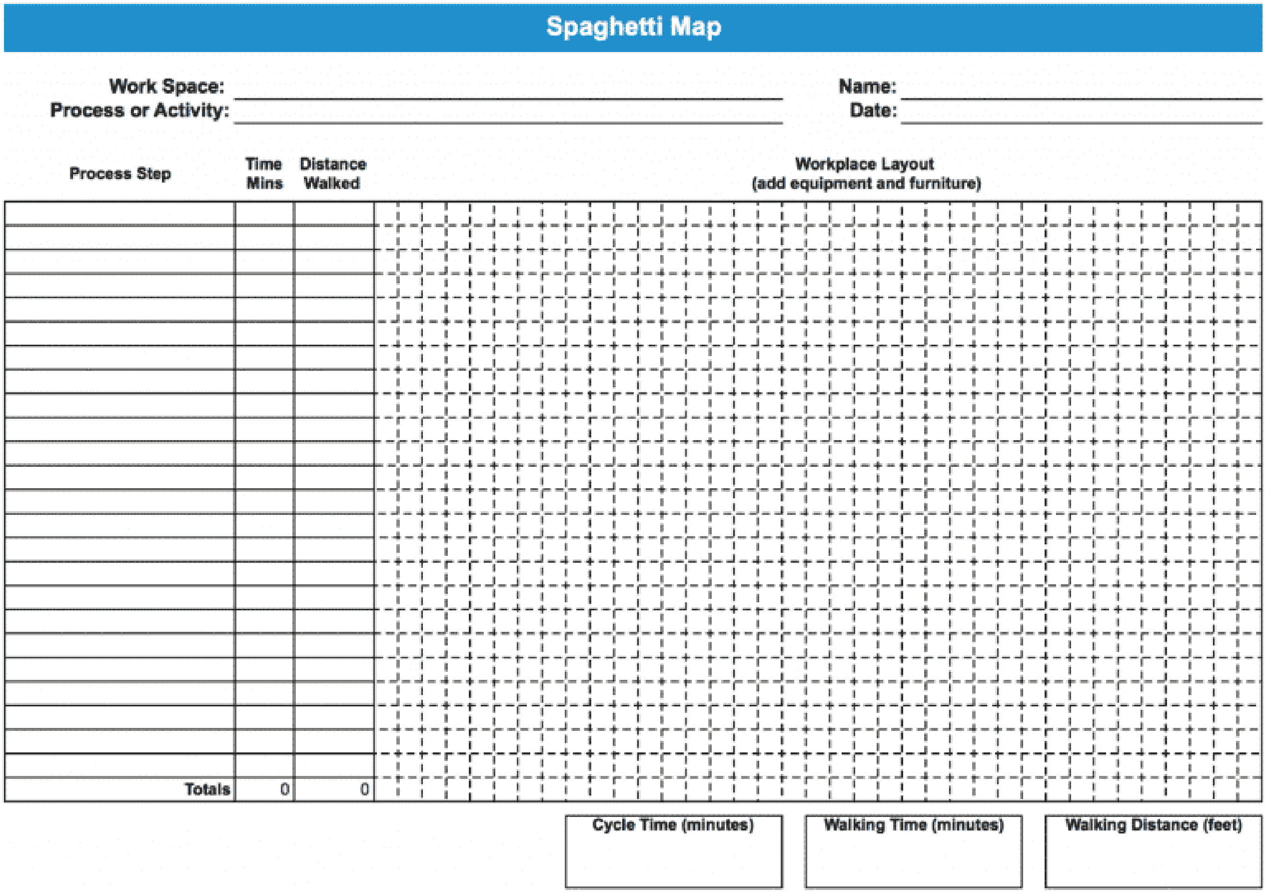


Figure 22: Relationship Map Example (GO Lean Six Sigma, n.d.)

Finally, teams may find it helpful to prepare a spaghetti map, a SIPOC, or develop a Threats and Opportunities Matrix. A SIPOC diagram maps out processes using the framework of Suppliers, Inputs, Processes, Outputs, and Customers. Alternatively, a spaghetti map is focused on mapping the movement of workers during the completion of a process. An example of a spaghetti map can be found in Figure 23 and an example of a Threats and Opportunities Matrix can be found in Figure 24.


Spaghetti Map

Work Space: ____________ Name: ____________

Process or Activity: ____________ Date: ____________

Process Step	Time Mins	Distance Walked	Workplace Layout (add equipment and furniture)
Totals	0	0	

Cycle Time (minutes)	Walking Time (minutes)	Walking Distance (feet)

Figure 23: Spaghetti Map (GO Lean Six Sigma, n.d.)

Threats & Opportunities Matrix

Project Objective: ______________________

Project Sponsor: ______________________

Project Stakeholder: ______________________

	Threats If We Don't Do Something	**Opportunities** If We Do Take Action
Short Term (Less Than 6 Months)		
Long Term (More Than 6 Months)		

Figure 24: Threats & Opportunities Matrix (GO Lean Six Sigma, n.d.)

3.5 Measure Phase

3.5.1 Measure Phase Overview

The measure phase is designed to capture the current state of the problem and quantity the current process or system's performance. In other words, it answers the questions "How does the process currently perform?" as well as "What is the magnitude of the problem?" Although Measure is its own dedicated phase of Lean Six Sigma projects, appropriately measuring data and progress is both crucial and critical throughout the life of not only the process improvement initiative but the on-going process once it is in a sustainable, optimized state.

As teams collect data, it is important to focus on what the different customers identified during the Define phase care about. Typically, this results in two focus areas – reducing lead-time and improving quality. In this particular project, emphasis was also placed on reducing costs, which is tied to lead-time.

In summary, the main objective of the measure phase is to collect data that can measure the problem so that there is a starting baseline that will be particularly necessary for the Improve phase. This phase focuses not only on collecting data, measuring historical performance, and arguable most excitingly beginning to identify waste. In short, the measure phase answers the question "How do you know it's a problem?" and results in the following deliverables:

- Detailed process "swimlane" map showing tasks, owners, and initial observations on waste
- Data collection plan and collected data
- Results of Measurement system analysis to validate collected data

- Graphical analysis of data to visualize historical performance and trends
- Process capability and sigma baseline

3.5.2 Swimlane Map

A swimlane map is an illustration of a process broken down into “swimlanes” for each different owner of a process step. A swimlane map was created for the new case setup, data processing, and post-migration workflows. Figure 25 shows the swimlane created for the new case setup process. Red stars indicate initial observations of waste identified in the current process.

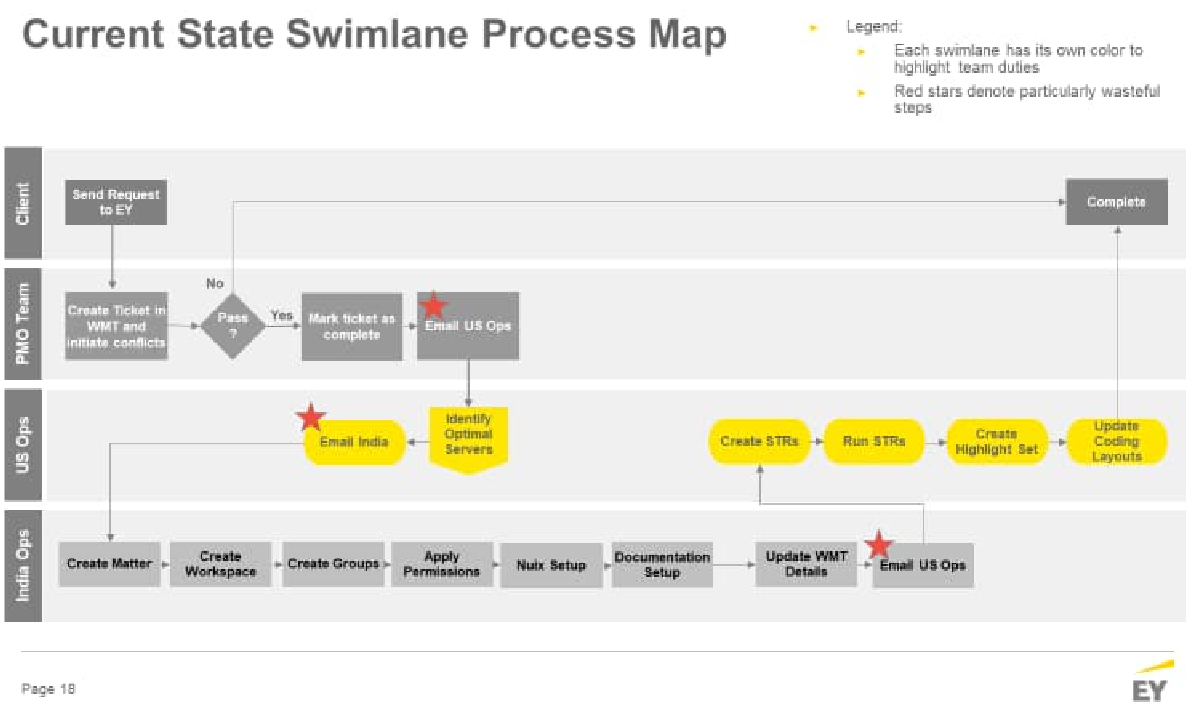


Figure 25: New Case Setup Swimlane

The swimlane process map is designed to visually show the inputs, actions, and outputs of a process clearly defined by responsible resource. It should clearly indicate the relationship between inputs (X) and outputs (Y) as well as the CTQ factors in the process. It provides an opportunity to identify waste in the process as well as the critical path.

In addition to swimlane process maps, some Lean Six Sigma projects elect to perform value stream mapping which provides a visualization of value across the organization or enterprise. It provides the transformation of raw materials into the final product available for customers. Because this project does not use raw materials, it was deemed out of scope for this initiative. Additionally, spaghetti diagrams, or spaghetti charts, map distances travelled by people and materials. This type of analysis was also not applicable to this initiative.

3.5.3 Data Collection Plan

A Data Collection plan is designed to achieve context for the data that will be analyzed. As part of the plan, the data collection goals should be identified as well as operational definitions. It can also include a sampling plan and should include a validation plan. It can reference a data collection form as well as a data collection check sheet. An example of a data collection plan can be found in Figure 26.

Data Collection Plan					
Measure Title	**Data Type (Continuous or Discrete)**	**Operational Definition**	**Stratification Factors (By who/what/ where/when)**	**Sampling Notes (Time Frame, etc.)**	**Who and How (Person responsible and method - Check Sheet?)**
Order Lead Time	*Minutes - Continuous*	*The amount of time (in minutes) it takes from the moment the patron places the order to the moment the order is delivered.*	*By Time of Day By Server*	*Sample every 4th customer from 11 - 2 for the next 6 weeks starting 1/30*	*Host will check the time stamp on the security video*
Incomplete Orders	*Discrete*	*Any pick-up order missing the correct supplies including napkins, hot sauce, forks, or knives*	*None*	*Sampling all orders for 6 weeks to check for incomplete items*	*Cashier to fill in the check sheet for incomplete orders*

Figure 26: Sample Data Collection Plan (GO Lean Six Sigma, n.d.)

For this project, three different sets of data were collected and subsequently analyzed:

1. **Time Entry Data –** The primary objective of collecting and analyzing this data set was to measure the costs associated with the process.
2. **Workflow Tool –** This dataset tracks various requests including when they were received and when they were released to the client. Each request receives a unique identifier which is referenced in the description of the time entry data.
3. **Matter Databases –** Finally, individual matter databases were analyzed to identify when certain activities were completed and if the appropriate quality measures were achieved.

Together, these three data sources help analyze all three objectives for this project – quality, speed, and cost.

3.5.4 Historical Performance

To analyze historical performance, histograms were used for each process and for each data source. Figure 27 shows the histogram created for new case setup time.

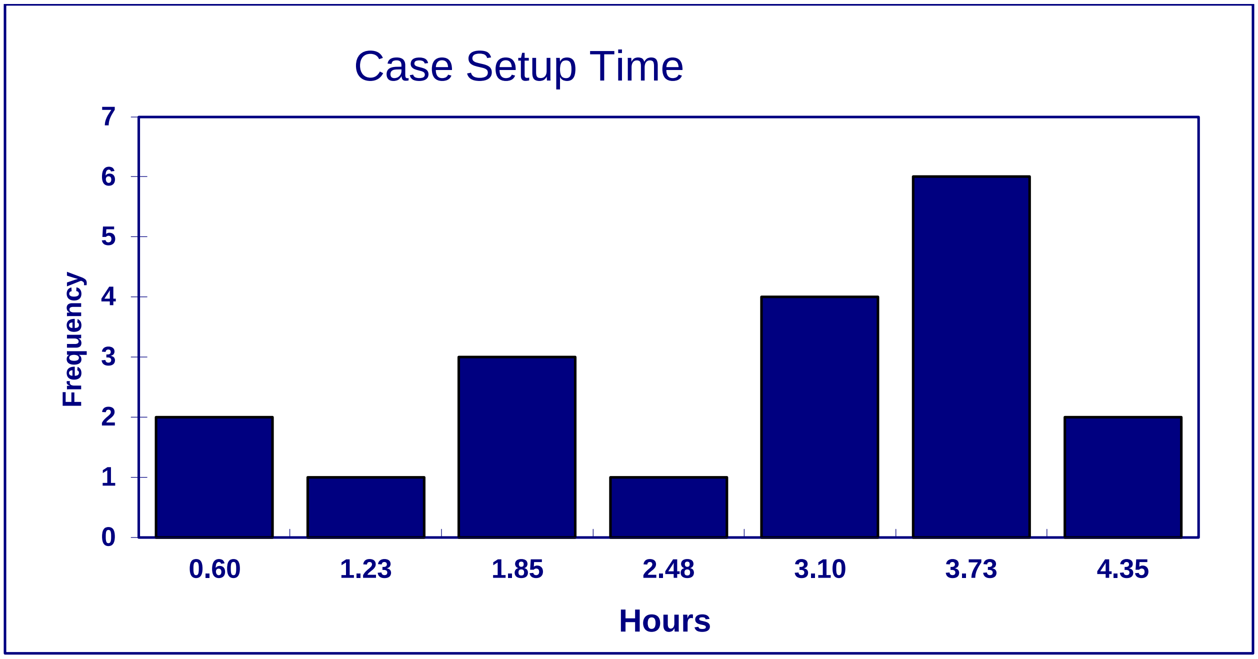


Figure 27: New Case Setup Time Histogram

Chapter 4: Results provides additional details comparing the baseline metrics identified in this phase to the final results of the program. Additional data visualizations could include control charts, frequency plots, box plots, main effects plots, scatter plots, and pareto charts. Because this process was immediately recognized as an opportunity for automation, the visualizations described were not prepared to justify and measure the baseline problem.

3.5.5 Sigma Baseline

Lean Six Sigma projects are defined by measuring defects per million opportunities. To be compliant with Lean Six Sigma, there need to be fewer than 3.4 defects per million opportunities (DPMO). Defects for this project were defined as requests that did not complete within the Service Level Agreement (SLA) time which can result in the engagement team facing a financial penalty. Fortunately, this project did not have any missed SLAs for the year prior to the initiative beginning so the project was already considered to have met the quality threshold.

While this may imply that a Lean Six Sigma project could be a wasted effort for this initiative, it is important to note that the sigma level only measures defects and does not indicate costs. This initiative is a good example of a project that while it had appropriate quality measures and compliance, it was not financially viable to continue operating at the current costs. We will explore the financial impact further in chapter 4.

3.5.6 Measure Phase Summary

As discussed, the Measure Phase aims to answer the question “How do you know it’s a problem?” Figure 28 summarizes EY’s requirements for this phase.

Critical Questions	Standard Tasks	Recommended Tools
Describe how you collected data to help you understand the problem?	Collect Data	Data Collection Action Plan
Is the data and insight that you've gained reliable?	Evaluate data trustworthiness (source, how it is collected, who, etc)	N/A – Coach Should Help Evaluate
What are your observationsof the historical performance of the process?	Graphs / Lean tools (VSM, 5S), Calculate baseline performance	Graphical Tools Should be Used Here
What is the current state process?	Process mapping	Swimlane Process Map
What forms of waste are present?	Define Defect and Operational Definitions	8 Wastes, “Starbursts” on Process Map, Timing of Tasks, Value Analysis

Figure 28: Measure Phase EY Requirements

Additional Lean Six Sigma Tools could include the following:

- Check Sheet – A check Sheet systematically collects data and its frequency of occurrence, particularly defects, in a simple tally sheet. An example of a check sheet can be found in Figure 29.

Standard Event Occurrence Check Sheet

Check Sheet Name: Late Lunch Orders

Time Frame: Week of February 2nd

#	Date	Time	Food Order	Total Lead Time (min)	Server	Prep Cook	Notes
1	*2/2*	*1:14 PM*	*Conch Salad*	*35*	*Tracy*	*John*	*New Prep Cook on the Conch Salad*

Figure 29: Sample Check Sheet (GO Lean Six Sigma, n.d.)

- Cost of Poor Quality (COPQ) Calculator – It can also be helpful to quantify the cost of defects throughout a process and in particular examine prevention, appraisal, and failure costs. Failure costs could include both internal and external failures.
- Efficiency & Effectiveness Matrix – These matrices map input, process, and output measures to their effectiveness and efficiency. An example can be found in.

Efficiency & Effectiveness Matrix

	Input Measures	Process Measures	Output Measures
Effectiveness Measures			
Efficiency Measures			

Figure 30: Sample Efficiency & Effectiveness Matrix (GO Lean Six Sigma, n.d.)

- Matching X&Y Measures – This template maps which X input processes impact each Y output measure. This type of report assists in determining which datasets are relevant to initiatives and are helpful in performing root cause analysis. Figure 31 below provides a sample of what this type of matrix can look like.

Matching X & Y Measures

X_1	X_2	X_3	X_4	X_5	Y

Figure 31: Matching X & Y Measures (GO Lean Six Sigma, n.d.)

3.6 Analyze Phase

3.6.1 Analyze Phase Overview

The Analyze phase aims to use the data collected during the Measure phase to identify the root cause of the problem identified during the Define phase. It is critical during the analyze phase to identify the true root causes and not just straight to forming solutions which can waste time, consume resources, create more variation, and even cause new problems. The main objective of this phase is to identify the root cause of problems including inefficiencies, quality concerns, and also opportunities for improvement. Said simply, the analyze phase answers the question “What is the root cause?”

3.6.2 Waste Analysis

Different Lean Six Sigma resources define different categories of waste but can typically be rolled up under the “3M:” Muda (waste), Muri (overburden), and Mura (unevenness) (Hessing, Classic Wastes, n.d.). At EY, the acronym DOWNTIME is used to examine the different types of waste. Figure 32 shows the DOWNTIME categories and definitions.

D	*Defects*	Failure to meet the requirements of internal or external customers
O	*Over-production*	Creating more materials or information than needed, or producing it before the customer wants it.
W	*Waiting*	Waiting for materials or information to be processed; results in idle time.
N	*Not utilizing Talent*	People seen as source of labor only and are told what to do rather than asked to think. Not consulted for improvement ideas.
T	*Transportation*	Moving materials or information unnecessarily. More opportunity for error.
I	*Inventory*	More materials or information produced than needed to satisfy the customer.
M	*Motion*	Inefficiencies in process require additional clicks or moving materials or information back and forth between people.
E	*Excessive Processing*	Re-work being performed downstream because it wasn’t done (or done properly) at the source.

Figure 32: Types of Waste: DOWNTIME

The following types of waste were identified for this project:

- **Defects –** Because there are so many repetitive, manual tasks, the process is open to human error specifically in the privilege screen creation and permission group creation and application.
- **Over-production –** NA: Requests are only worked on once initiated by the client.
- **Waiting –** There is lost time in the hand offs between teams – an estimated ~12 hours throughout the process.
- **Not Utilizing Talent –** The tasks are necessary, however mindless. They are primed for automation to allow labor to work on more stimulating and rewarding tasks.
- **Transportation –** The process has unnecessary handoffs which allow more room for typos/misinformation to be passed along.
- **Inventory Excess –** NA: Only the necessary steps are performed.
- **Motion Waste –** While there isn't waste per se, the process does require lots of manual clicks which are prime for automation.
- **Excessive Processing –** NA: Only the necessary steps are performed.

3.6.3 Cause Identification

Lean Six Sigma has several different tools to help identify causes for the waste identified during the waste analysis:

- **Brainstorming:** While the most simplistic tool, brainstorming can be one of the most efficient for quickly gathering multiple ideas. For this initiative in particular, the areas that came to mind during brainstorming were the areas ripe for automation.
- **5 Whys:** The "5 Whys" are exactly what they sound like – asking a question of why waste exists, asking why that is, and repeating 3 more times with the goal to identify the

true root cause of the problem. This exercise was completed repeatedly throughout the process.

- **Fishbone Diagram:** A fishbown diagram is a type of diagram designed to visualize causes and effects, an example of which can be found in Figure 33 (Hessing, Cause and Effect Diagram (aka Ishikawa, Fishbone), n.d.). A fishbone diagram was not prepared for this initiative.

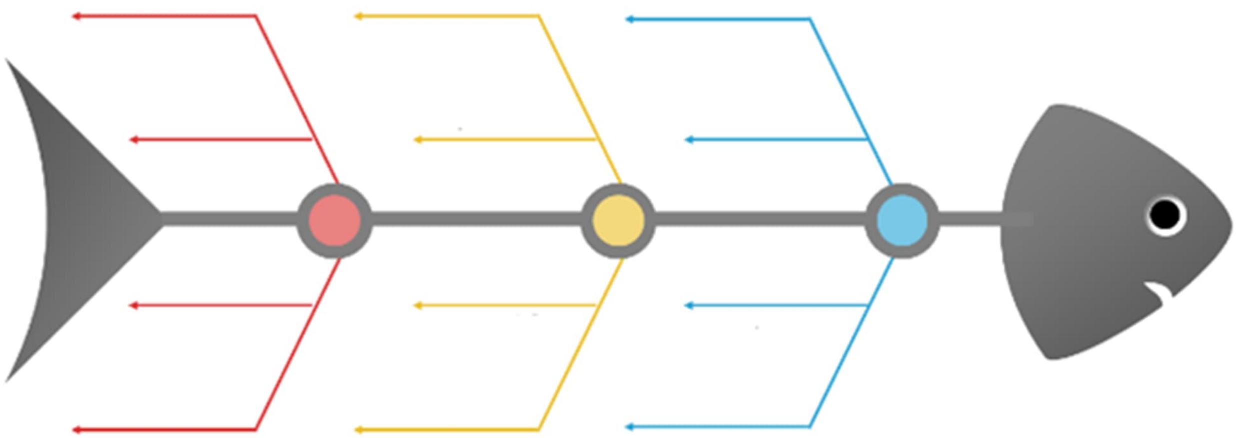


Figure 33: Fishbone Diagram (GO Lean Six Sigma, n.d.)

- **Fault Tree Diagram:** A Fault Tree Diagram is a graphical tool to perform Fault Tree Analysis (FTA). Fault Tree Analysis is designed to visualize various level of components to identify the appropriate level of failures and can be seen in Figure 34. A Fault Tree Diagram was deemed unnecessary for this initiative since it is a linear process without individual components.

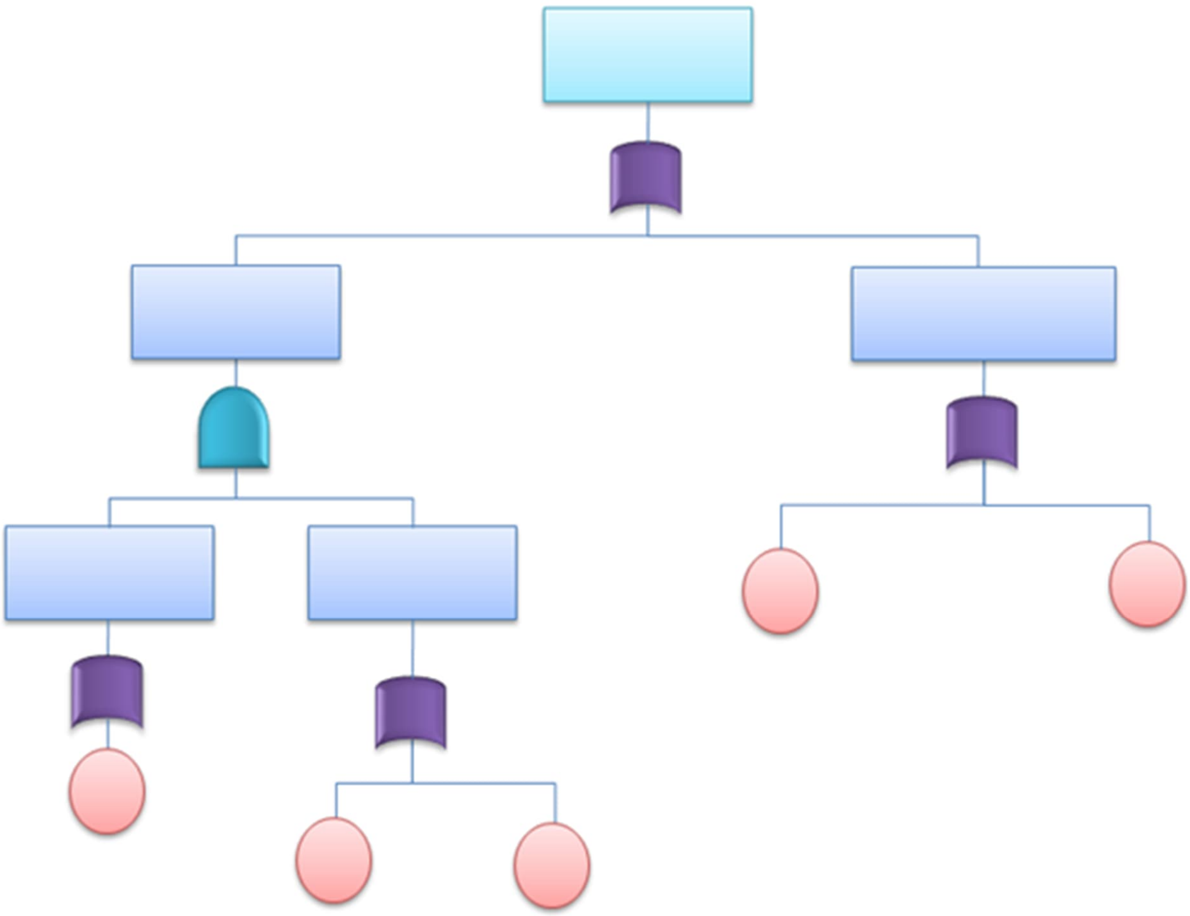


Figure 34: Fault Tree Diagram (GO Lean Six Sigma, n.d.)

- **Affinity Diagram (K-J Method):** An affinity diagram, also known as the K-J Method, is helpful to use with unfamiliar problemd. Because this project was intimately familiar, an affinity diagram was deemed unnecessary.
- **Tree Diagram:** Similar to the Fault Tree Diagram, a tree diagram is useful to break down complex processes into smaller parts. While not prepared for this project, it would be helpful for other automation initiatives to eventually build requirements for automated processes.

- **Interrelationship Diagram:** An interrelationship digraph, also known as a network diagram, ties relationships and influences between several concepts. It can identify critical issues and key drivers of problems but was not prepared for this initiative.
- **Failure Mode Effects Analysis (FMEA):** FMEA a tool that aims to anticipate what might go wrong with a product or process and identify the possible causes and probabilities of failures. Because this process does not often have failures, it was not deemed necessary for this engagement.

3.6.4 Cause Identification

During the cause identification phase, it is often necessary to pull in statistical analysis methods to analyze the different inputs and outputs of processing. Keeping in mind Lean Six Sigma is designed for manufacturing processes, it is often necessary to look at the distribution of key metrics for various components. It is necessary to understand the various distributions including normal, lognormal, F, Chi Squared, Exponential and Student's T for continuous measures and binomial probability, poisson, and hypergeometric for discrete distributions.

It can also be necessary during this phase to complete at least some form of hypothesis testing with an emphasis on design of experiments (DOE). DOE aims to find the right settings for key process input variables via factorial designs, fractional factorial designs, Taguchi loss function, Evolutionary Operations (EVOP), Latin Square, Graeci-Latin Square Design, or Hyper-Graeco-Latin Square Design. While not completed during the Analyze phase of this initiative, these concepts will be explored further in chapter 4 when discussing the results of this initiative.

3.6.5 Analyze Phase Summary

The Analyze Phase is the first instance where the consideration of RPA becomes necessary. For Lean Six Sigma, there is an emphasis on fixing the "problems" but with RPA, there is a world of opportunity that should be considered. For this initiative, it was more of an optimization opportunity than solving a particularly troubled program. Figure 35 provides a summary of the steps necessary for EY's program.

Critical Questions	Standard Tasks	Recommended Tools
Describe the relationship between inputs and the output (x's and Y)?	Complete Process Analysis	Coach's Discretion
Describe the significant process inputs & root causes	Identify waste, Root cause analysis, Cause-Effect Diagram	Fishbone, 5 Whys
What are the primary root causes that you want to attack? How did you prioritize?	Prioritize root causes	Prioritization Matrix

Figure 35: EY Analyze Phase Summary

3.7 Improve Phase

3.7.1 Improve Phase Introduction

Now that the problem has been clearly defined and the data gathered and analyze, teams are ready to begin identifying potential solutions to the problems identified during the analyze phase. The Improve Phase is where the team brainstorms solutions, pilots process changes, implements solutions and lastly, collects data to confirm there is a measurable improvement. By the end of this phase, we should be able to answer the question "How will you fix the problem?" For this initiative, we know at least part of the answer is RPA but we also want to determine the way to implement, test and operationalize. Process owners are consulted, and an action plan is circulated to relevant stakeholders as a part of the change management process.

3.7.2 Identifying Solutions

Similar to the Analyze Phase, brainstorming and affinity diagrams can be useful to help identify solutions. Additionally, Six Sigma promotes the themes of the 5S (sort, straighten, shine, standardize, and sustain) when considering solutions as well as Poka-Yoke which is the act of error-proofing a process through great design. Takt time is used to align production with sales which is not applicable to this process since the process is only initiated after a request is created.

For this project, a prioritization matrix was created to prioritize the steps to automate. One of the goals of this automation initiative was to automate in a way where intermittent gains could be achieved throughout the process.

3.7.3 Solution Finalization

Finally, it's time to finalize which solutions will be implemented. This includes cost-benefit analysis. Cost-benefit analysis with automation can be more complex than typical Return on Investment (ROI) calculations because of the variety of benefits that automation can bring including cost savings, quality improvements and speed improvements that can lead to higher customer satisfaction and improved employee morale via the reduction of tedious tasks which can hopefully improve employee retention. Further complicating matters, the different benefits can have different importance for different stakeholders. Because of these complexities, it can be helpful to use optimization models to prioritize which automations to implement and in which order.

Letting x_{ij} represent a binary variable indicating if the ith automation for the jth workstream is automated, minimizing the cost of implementing each automation can provide a basic model for selecting each solution.

Minimize:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}\, x_{ij}$$

Subject To:

Supply (i) : $\sum_{j=1}^{m} x_{ij} = 1$ where, i = 1,2, …, n

Demand (j) : $\sum_{i=1}^{n} x_{ij} = 1$ where, j = 1,2, …, m

Cost (i,j) : $c_{ij}x_{ij} \geq 0$ where i = 1,2,…,n and j = 1,2,…, m

While simplistic, this model has its shortcomings. Firstly, it assumes the cost for each automation component is known. Secondly, it assumes that all automation initiatives are for an individual component, when in actuality many components are grouped together. Finally, it assumes cost is the only determinant when selecting the solution. For this reason, it can be more beneficial to maximize satisfaction (variable s) with desired weights (w) amongst the stakeholders within given constraints:

Maximize:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} w_j s_{ij}\, x_{ij}$$

Subject To:

Supply (i) : $\sum_{j=1}^{m} x_{ij} = 1$ where, i = 1,2, …, n

Demand (j) : $\sum_{i=1}^{n} x_{ij} = 1$ where, j = 1,2, …, m

Cost (i,j) : $\sum_{j=1}^{m} \sum_{i=1}^{n} c_{ij} x_{ij} \leq B$ 1 where, j = 1,2, …, m and i = 1,2,…,n, and B = budget

While better than the first model, the second is still simplistic in that it assumes satisfaction can be easily calculated. This is where a data-driven organization can benefit from having metrics available on the value of employee retention, the cost of various mistakes, etc. For this reason, weights are introduced to allow prioritization of importance of different stakeholders. Finally, additional constraints (such as time limitations) may be necessary and should be considered.

3.7.4 Improve Phase Summary

For this initiative, automation was deemed to have an overwhelmingly positive ROI because of the minimal development time partnered with the increasing demand of data to be processed. The results of these improvements will be discussed in detail in Chapter 4. Figure 36 provides the details of the steps required to complete the Improve Phase in EY's program.

Critical Questions	Standard Tasks	Recommended Tools
Describe how you identified and evaluated all potential solutions?	Generate Solutions	'Brainstorming' & 'Narrowing Down The List of Ideas'
How did you prioritize your potential solutions?	Select Solutions	Solution Prioritization Matrix
How did you validate that your improvement solved the root cause?	Test/pilot solutions; Recalculate Baseline Metrics	Before vs After Evaluation – Coach's Discretion
How did the team plan and track actions?	Action Plan	Action Plan
What does the future state process look like?	Update process documentation	Future State Process Map / VSM

Figure 36: EY Improve Phase Requirements

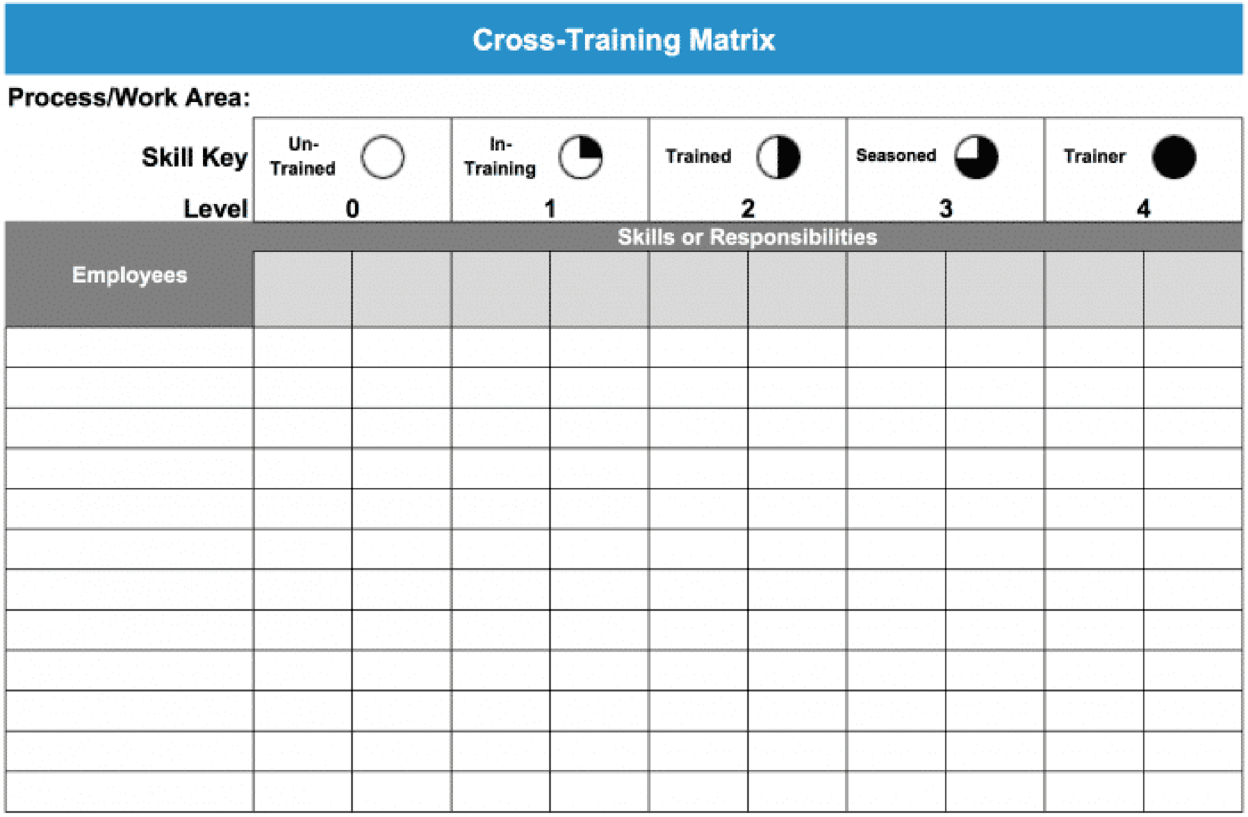
Additional tools that could assist project teams in the Improve Phase could include a 5S Manufacturing assessment, 5S transactional assessment, Cross-Training Matrix, FMEA, Idea Funneling Guide, Implementation Plan, Pilot Checklist, Solution Selection Matrix, or Weighted Criteria Matrix. Examples of each of these resources can be found below.

5S Manufacturing Assessment								
Work Area:			**Key:**	Use sheet to rate work area 5 times (note each date) 1 = "non-existent", - 3 = "average" and 5 = "excellent"				
5S Phase	**Definition**	**Standards To Be Met**	**Ratings**					**Next Steps**
		Date of Assessment						
Sort (Seiri)	The right materials are available and anything unnecessary is removed	- Unused parts, tools and equipment removed						
		- There is nothing in the hallways impeding flow						
		- There is no excess inventory being stowed away						
		- There is no out of date signage on the walls						
Set in Order (Seiton)	There is place for everything and everything is in its place	- Shelving has clear labels or pictures for parts						
		- Floors are taped to indicate equipment locations						
		- Tool locations are marked or shadow-boarded						
		- Things are not put down, they are put away						
Shine (Seiso)	Everything is clean and in working order	- All equipment is clean and painted to show leaks						
		- Cables are bundled and there are no loose wires						
		- Cleaning tools and supplies are readily available						
		- All surfaces are dirt and grime free						
Standardize (Seiketsu)	Guidelines and practices are established to maintain first three steps	- 5S activities and locations are clearly outlined						
		- Audit forms and checklists are being used						
		- There is a 5S schedule & responsibilities are clear						
		- Quantities and limits are clearly marked						
Sustain (Shitsuke)	5S is a habit that people incorporate into their daily practice	- Leadership enforces daily 5S habits						
		- There is accountability for ongoing 5S practice						
		- 5S results are prominently displayed						
		- Employees are 5S-trained and recognized						
		Total Score	0	0	0	0	0	

Figure 37: 5S Manufacturing Assessment Template (GO Lean Six Sigma, n.d.)

5S Transactional Assessment								
Work Area:			**Key:**	Use sheet to rate work area 5 times (note each date) 1 = "non-existent", - 3 = "average" and 5 = "excellent"				
5S Phase	**Definition**	**Standards To Be Met**	**Ratings**					**Next Steps**
		Date of Assessment						
Sort (Seiri)	The right materials are available and anything unnecessary is removed	- No unused items are stored						
		- No unneeded materials, forms or supplies						
		- There are no out-of-date posters on the wall						
		- Excess supplies are reallocated						
Set in Order (Seiton)	There is place for everything and everything is in its place	- It's clear where working vs archive files belong						
		- The shared drive is easy to navigate						
		- Signage & naming conventions are clear						
		- Equipment and supply areas are clearly labeled						
Shine (Seiso)	Everything is clean and in working order	- IT conducts regular maintenance						
		- Licenses are renewed and updated on schedule						
		- Systems suffer minimal downtime						
		- Employees have access to the right applications						
Standardize (Seiketsu)	Guidelines and practices are established to maintain first three steps	- 5S activities and locations are clearly outlined						
		- Audit forms and checklists exist						
		- There is a 5S schedule & responsibilities are clear						
		- Quantities and limits are clearly marked						
Sustain (Shitsuke)	5S is a habit that people incorporate into their daily practice	- Leadership enforces 5S habits						
		- There is accountability for ongoing 5S practices						
		- 5S results are prominently displayed						
		- Employees are recognized for 5S practice						
		Total Score	0	0	0	0	0	

Figure 38: 5S Transactional Assessment Template (GO Lean Six Sigma, n.d.)



Cross-Training Matrix

Process/Work Area:

Skill Key	Un-Trained	In-Training	Trained	Seasoned	Trainer
Level	0	1	2	3	4

Employees	Skills or Responsibilities									

Figure 39: Cross-Training Matrix (GO Lean Six Sigma, n.d.)

FMEA Form

Process/Product Name: ______________________ Prepared By: ______________________

Responsible: ______________________ FMEA Date (Orig.): ____________ (Rev.): ____________

Process Step/Input	Potential Failure Mode	Potential Failure Effects	SEVERITY (1 - 10)	Potential Causes	OCCURRENCE (1 - 10)	Current Controls	DETECTION (1 - 10)	RPN	Action Recommended	Resp.	Actions Taken	SEVERITY (1 - 10)	OCCURRENCE (1 - 10)	DETECTION (1 - 10)	RPN
What is the process step, change or feature under investigation?	In what ways could the step, change or feature go wrong?	What is the impact on the customer if this failure is not prevented or corrected?		What causes the step, change or feature to go wrong? (how could it occur?)		What controls exist that either prevent or detect the failure?			What are the recommended actions for reducing the occurrence of the cause or improving detection?	Who is responsible for making sure the actions are completed?	What actions were completed (and when) with respect to the RPN?				
Fill carafe with water	*Wrong amount of water*	*Coffee to strong or weak*	*8*	*Faded level marks on carafe*	*4*	*Visual Inspection*	*4*	*128*	*Replace old carafes*	*Mel*	*Carafe replaced 9/15*	*8*	*1*	*3*	*24*
								0							0
								0							0
								0							0
								0							0
								0							0

Figure 40: FMEA Form (GO Lean Six Sigma, n.d.)

GoLeanSixSigma.com Idea Funneling		
1. Clarify the "How Can We" Statement	**Examples**	**Draft the "How Can We" Statement Below:**
A problem well-stated is a problem half-solved. Make sure the statement clarifies exactly what you want addressed. Test it to make sure people are clear on what you're asking for.	How can we reduce the time to submit an application?	
	How can we reduce the errors in production?	
	How can we spend more time with patients?	
2. Select the Problem Solvers	**Examples**	**List Potential Problem Solvers**
Now is your opportunity to involve people who know the process well, people who might be impacted by changes to the process and people who are unconnected–new eyes.	Process Participants	
	Stakeholders	
	People new to the problem/process	
3. Send the Problem to Individuals	**Details**	**Determine the Due Date & the Idea Collection Method**
Send the "How Can We" Question out to individuals, not a group. Make sure you ask each person separately, give them a deadline and provide link to where you want ideas captured.	Set a due date	Solution Ideas Due on:
	Create a link to an idea collection document	Shareable Link to Idea Capture:
4. Let People Think		**(Quiet–It's Theta Time)**
Give people time to sleep on it. Allow a few days so people can walk, bike, shower, sleep and dream on it. They've got a deadline so they know what to do. It's theta time!		
5. Assess Ideas as a Group	**Examples**	**Determine the Assessment Plan**
Decide on method or format to clarify, refine and expand the solution ideas	LCS: Likes, Concerns, Suggestions	
	Plus/Deltas	
6. Pick a Winner	**Examples**	**Enter the Selection Tool**
Use selected criteria to make final decisions on which solutions to test out	Solution Selection Matrix	
	Impact Effort Matrix	

Figure 41: Idea Funneling Template (GO Lean Six Sigma, n.d.)

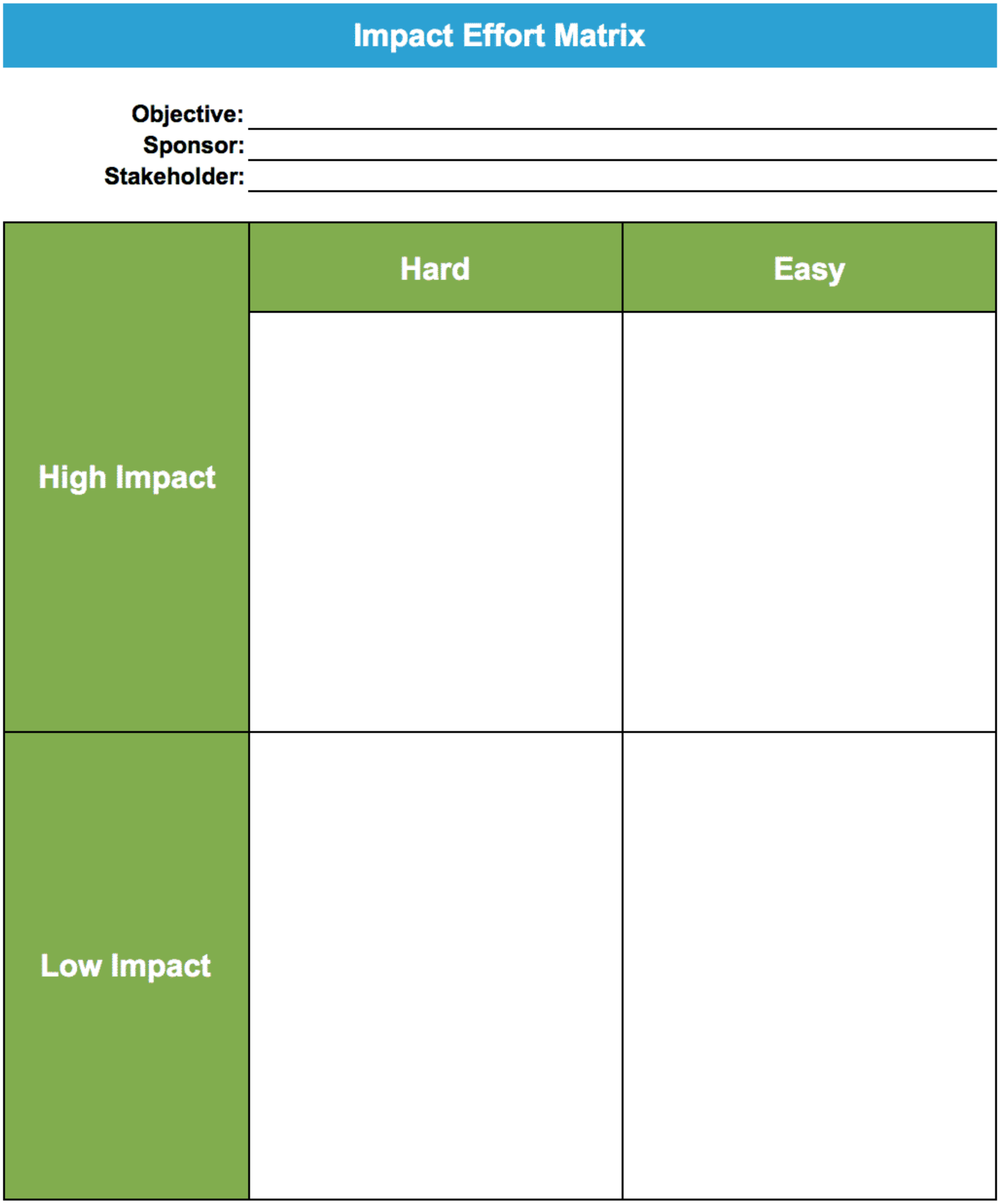


Figure 42: Impact Effort Matrix (GO Lean Six Sigma, n.d.)

Implementation Plan

Action Item (List steps required to implement solutions)	Responsible (List person(s) responsible for action steps)	Due Date (Indicate when action items must be completed)

Figure 43: Implementation Plan Template (GO Lean Six Sigma, n.d.)

Pilot Checklist

Pilot Effort: ______________________________

Pilot Strategy	Yes/No	Detail
Training Preparation		
Measurement Plan		
Adjustment Plan		

Figure 44: Pilot Checklist (GO Lean Six Sigma, n.d.)

Solution Selection Matrix

Project Goal

Enter Goal Statement below:
(As stated on Project Charter)

Please rank each solution for each criteria
by using the 1-5 Scale as indicated below

Very Low (less good)		Moderate		Very High (best)
1	2	3	4	5

Potential Solution (Provide Brief Description)	Potential to Meet Goal	Positive Customer Impact	Cost to Implement (1 = $$$ & 5 = $)	Stakeholder Buy-in	Time to Implement (1 = Long 5 = Quick)	Total Score	Implement? Yes/No
Weighted Criteria	10	9	8	7	5		
Don't include condiments with takeout	5	3	4	4	5	*162*	Yes

Figure 45: Solution Selection Matrix (GO Lean Six Sigma, n.d.)

Weighted Criteria Matrix

Prioritization Criteria	Value	Option 1	Score		Option 2	Score		Option 3	Score		Option 4	Score	
Time to implement	*6*	*Just a phone call*	*9*	*54*	*4 months interviewing & screening*	*1*	*6*	*Change scripts & order reprints*	*3*	*18*			*0*
				0			0			0			0
				0			0			0			0
				0			0			0			0
				0			0			0			0
				0			0			0			0
				0			0			0			0
				0			0			0			0
Totals			54			6			18			0	

Figure 46: Weighted Criteria Matrix (GO Lean Six Sigma, n.d.)

3.8 Control – Maintain the Solution

Finally, we reach the control phase which aims to sustain the improvements previously implemented. Now that the process problem is fixed and improvements are in place, the team must facilitate proper change management to ensure that the process maintains the gains. Control phases typically include a monitoring plan to ensure the gains are met and a response plan in case of an adverse event. Once finalized, these plans are turned over to the appropriate stakeholder. In short, the Control phase aims to answer the question "How do you know it will stay fixed?"

3.8.1 Change Management

For this initiative, several aspects of change management were completed. Firstly, documentation guides were created for the operations team including detailed step-by-step instructions demonstrating what the automation was doing behind-the scenes. The goal was that every team member could use this guide in case something in the automation failed. One of the downsides of automation can be when resources do not know how to perform tasks because they become reliant on the automation. Secondly, stakeholders were engaged throughout the process and several meetings were held to ensure a smooth transition. Finally, when the automation was deployed in production, additional QC measures were put in place the first 3 weeks to ensure there were not any surprises or unexpected mistakes.

3.8.2 Control Phase Summary

For this particular initiative, the control phase was the easiest to complete because the teams accountable for executing the work are under my position’s domain. However, this is not typically what I see with my clients where change management can ultimately play a bigger part than the actual process improvement. A communication plan cannot be emphasized enough. Figure 47 shows the EY requirements for the control phase with a clear emphasis on planning and documentation.

Critical Questions	Standard Tasks	Recommended Tools
Describe how your project meet the intended goal as listeed in Define?	Validate Savings on Improvements	Business case
Describe how the project has been handed-off to the process owner?	Document Standard Process	'Best Practices and Lessons Learned' & 'Standardized Work - Documenting Process Changes' & Project handoff Plan
What is the response plan if there are problems in the future?	Develop plan to address problems	Reaction Plan, Control plan
What were your lessons learned?	Document Key-Learnings	Lessons Learned
Can your project be considered a best practice? Describe where the improvement can be replicated in other areas of EY?	Document Standard Process	Scalability Slide
Attach your project summary and upload this deck to your project in Traction	Document Standard Process	One Pager Template

Figure 47: Control Phase EY Requirements

Other Lean Six Sigma Tools for the define phase include the following:

- Control Chart – A control chart is often in the form of a decision tree that aids stakeholders in the decision-making process to ensure the new process is followed.
- Executive Summary – An executive summary is designed to provide decision-makers easily consumable data points to understand the value of initiatives in a way that can be easily and succinctly articulated. Figure 48 provides a template for completing an executive summary.

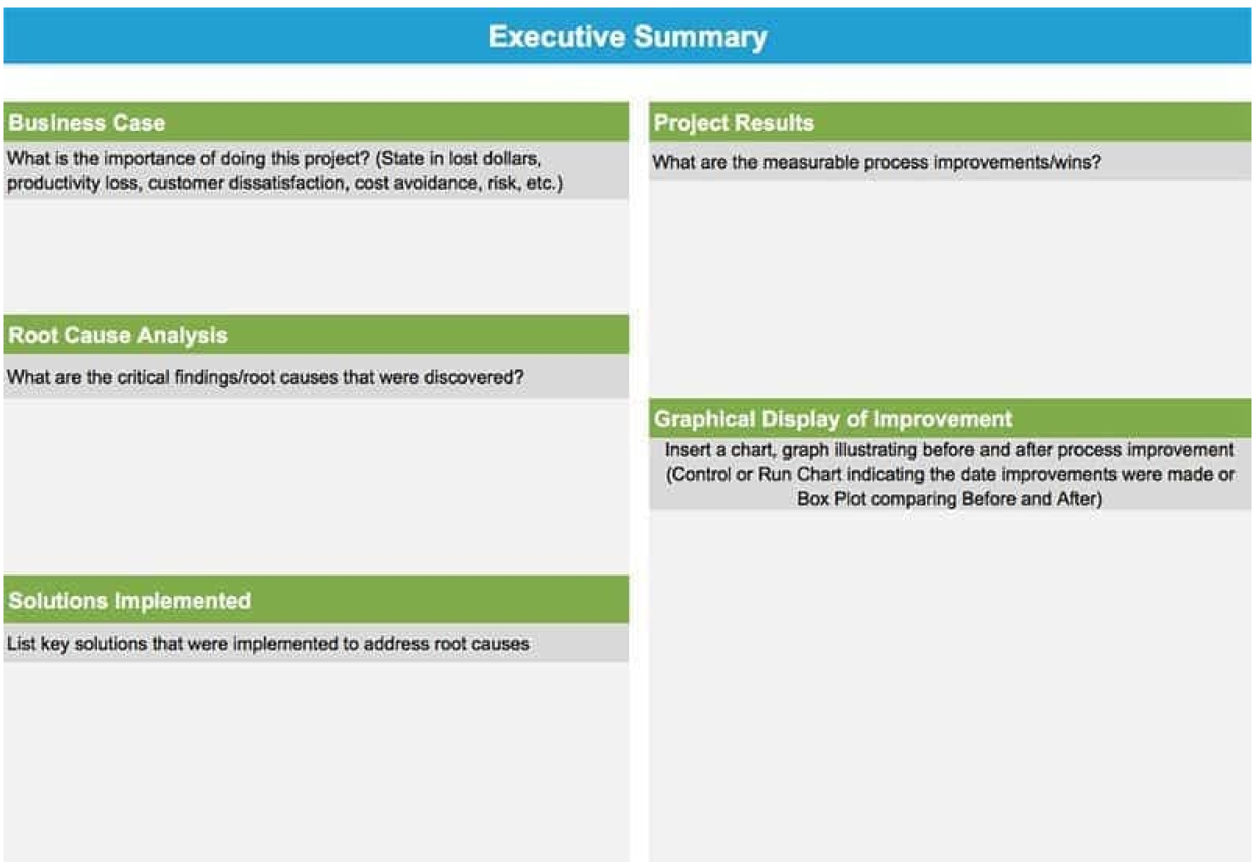


Figure 48: Executive Summary Template (GO Lean Six Sigma, n.d.)

- Innovation Transfer Opportunities – This worksheet structures how changes to the process can benefit other departments, business units, or work teams in the spirt of knowledge sharing. Figure 49 provides an example of this worksheet. While this worksheet was not completed for this initiative, knowledge sharing was very important to this team and was a top priority for leadership. This knowledge transfer happened more in meetings than written communication and ultimately in the merging of multiple teams and workflows.

Innovation Transfer Opportunities		
Process Innovation/Addition/Change/Removal (Describe the implemented solution)	Process Benefits (Describe the both measurable and intangible benefits of the change)	Area/Department/Business Unit (Indicate the process area that could benefit from innovation transfer)

Figure 49: Innovation Transfer Opportunities Template (GO Lean Six Sigma, n.d.)

- Monitoring – It's important for teams to have a monitoring plan in place as well as a response plan. Examples of both of these can be found in Figure 50 and Figure 51.

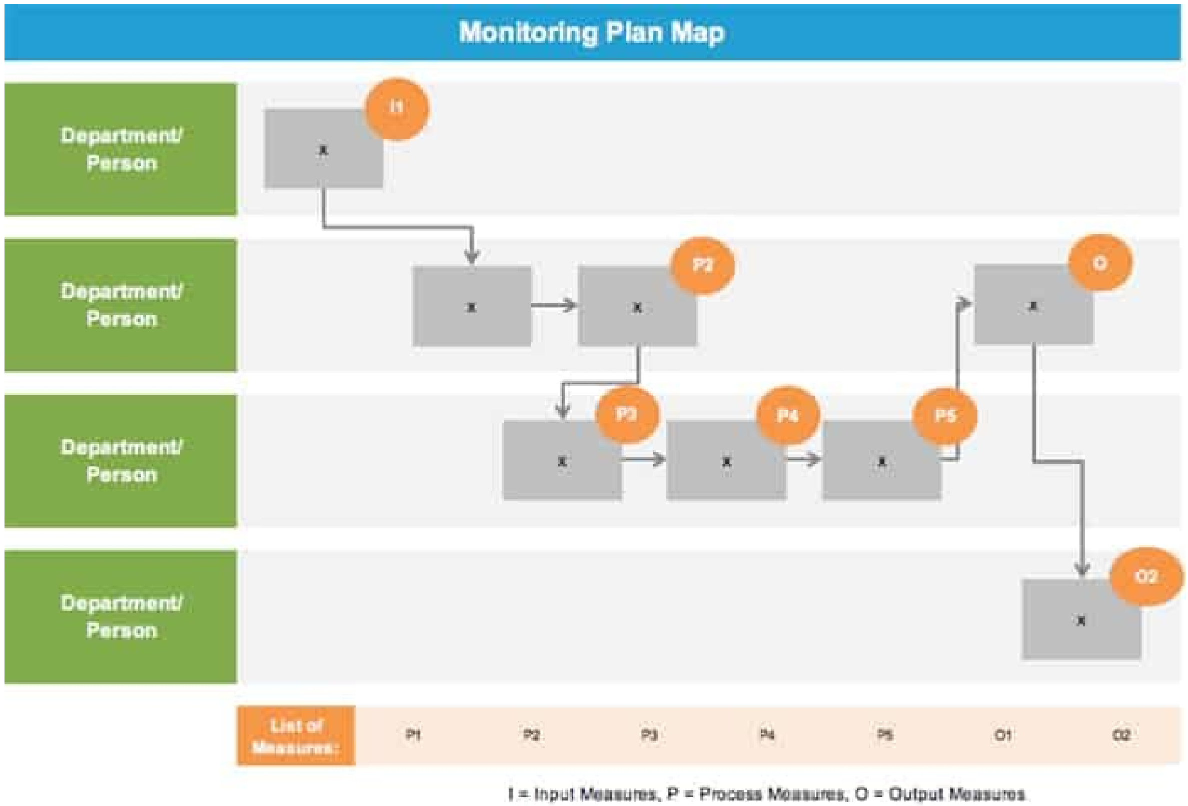


Figure 50: Monitoring Plan Map Example (GO Lean Six Sigma, n.d.)

Monitoring Plan						Response Plan		
Name of the Measure	Input, Process or Output?	What is the Target?	Method of Data Capture	Checking Frequency	Person Responsible	Upper/Lower Trigger Point	Who Will Respond?	Reaction Plan
O1: Order Lead Time	*O*	*less than 16 minutes for cold food; Less than 20 minutes for hot food*	*Time stamp, in and out*	*Daily*	*Server*	*Over 18 minutes for cold food; Over 22 minutes for hot food*	*Manager*	*Observe the process to see why it's taking longer. Make the corrections. Are the orders still being processed in FIFO order? Are Servers turning orders into the kitchen immediately after taking them? Are we stocked at point of use through peak hours?*

Figure 51: Monitoring and Response Plan (GO Lean Six Sigma, n.d.)

- New Procedure Audit – Figure 52 illustrates a sample of a new procedure audit which is designed to report on adoption. It was unnecessary for this initiative since team members did not have the option to use the former methodology.

New Procedure Audit

Procedure Name	Observed Levels								
	Awareness Level/Date			Level of Use/Date			Level of Enforcement/Date		

Figure 52: New Procedure Audit (GO Lean Six Sigma, n.d.)

Finally, Figure 53 provides an example of a template that can be used for project closure. As with previous resources, there is an emphasis on communication and articulating the value including savings over time.

Project Closure

Lessons Learned	Customer Impact
Do's and Don'ts for Future Efforts	Positive Impacts on External Customer

Final Calculations of Savings or Gains	
Hard Savings/Profit Increase	Soft Savings - Cost or Time

Process Owner Hand-off		Sign-off From Project Sponsor
Has been informed of process changes:	Yes / No	
Agrees to continued monitoring of new process:	Yes / No	
Has received new process documentation:	Yes / No	

Figure 53: Project Closure (GO Lean Six Sigma, n.d.)

3.9 Results Methodology

Finally, the results of this initiative were analyzed using the hypothesis-testing procedures taught in EMIS 7377: Statistical Design and Analysis of Experiments. From "Applied Statistics and Probability for Scientists and Engineers," the seven steps below were used as the hypothesis-testing methodology (Runger, 2011):

1. **Parameter of Interest:** From the problem context, identify the parameter of interest.
2. **Null Hypothesis, H_0:** State the null hypothesis, H_0.

3. **Alternative hypothesis, H_1:** Specify an appropriate alternative hypothesis, H_1.

4. **Test statistic:** Determine an appropriate test statistic.

5. **Reject H_0 if:** State the rejection criteria for the null hypothesis.

6. **Computations:** Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute that value.

7. **Draw conclusions:** Decide whether or not H_0 should be rejected and report that in the problem context.

These results are discussed in the next chapter.

CHAPTER 4

RESULTS

4.1 Outline

This chapter describes the process to assess the impact of this initiative and thus our hypothesis. As discussed in Chapter 3, the results of this project are assessed using the hypothesis-testing. This chapter is organized by initially providing the results of the data exploration and analysis of our sample datasets. Next, it uses the 7 steps in the hypothesis-testing methodology as defined in Chapter 3. Finally, it discusses the nuances between statistical and practical significance and highlights the impact the initiative made.

4.2 Data Exploration

In order to properly analyze the results of our process, it is first important to understand the datasets available. For this analysis, we have two sample sets – the time and expense data for requests charged in November 2018 as well as in November 2020. In the legal industry, the cycle of work can be very cyclical depending on the time of year. For that reason, it was important to select the same month to explore between sample sets. Figure 54: Sample Data Sets visualizes both data sets with November 2018 data shown in red and November 2020 costs shown in blue.

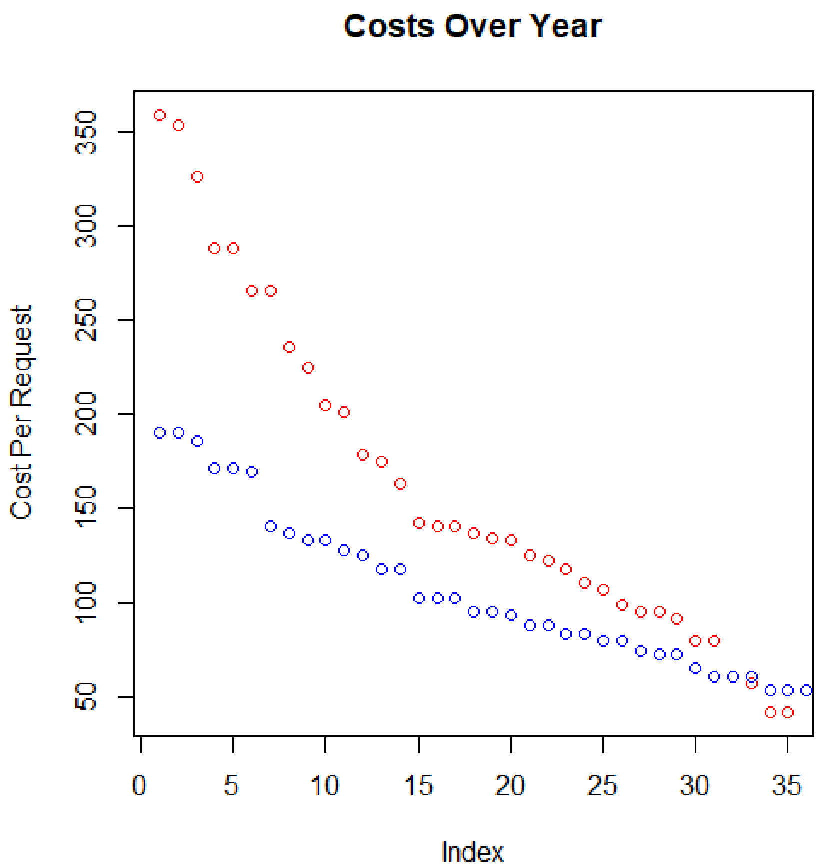


Figure 54: Sample Data Sets

Using R, Figure 55 and Figure 56 show an initial summary of the two datasets. Initial observations include a reduction in the mean costs from $207.40 in 2018 to $95.88 in 2020. Additionally, we see a reduction in the median from $140.60 in 2018 to $76.95 in 2020. However, while we can immediately see a reduction in costs between the two sets, further analysis is needed to test if the difference is statistically significant.

```
 ï..Request              Cost
Length:39          Min.   : 41.8
Class :character   1st Qu.:102.6
Mode  :character   Median :140.6
                   Mean   :207.4
                   3rd Qu.:266.0
                   Max.   :777.1
```

Figure 55: November 2018 Data Summary

```
 ï..Row.Labels        Sum.of.Total
 Length:54          Min.   : 17.10
 Class :character   1st Qu.: 41.80
 Mode  :character   Median : 76.95
                    Mean   : 95.88
                    3rd Qu.:117.80
                    Max.   :805.60
```

Figure 56: November 2020 Data Summary

4.2.1 Data Cleansing

With any data analysis, including statistical analysis, it is important to analyze and validate data sets to ensure that outliers don’t skew results. To begin this analysis, we prepared a histogram using R to see if we could visually detect any outliers in both the 2018 and 2020 datasets.

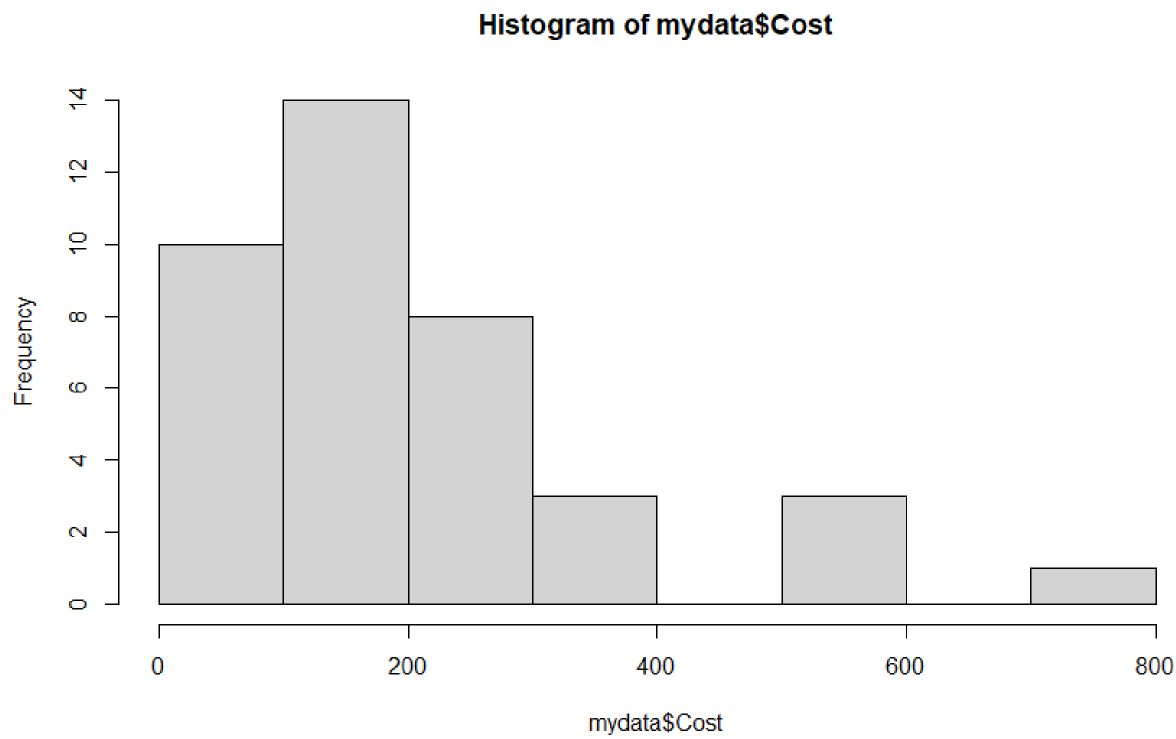


Figure 57: November 2018 Cost Histogram

Figure 57 visualizes are data from the 2018 dataset and shows that we have 4 requests that appear to be outliers. We confirm this in Figure 58 by using R to create a boxplot of the same dataset which identifies the same 4 data points as outliers.

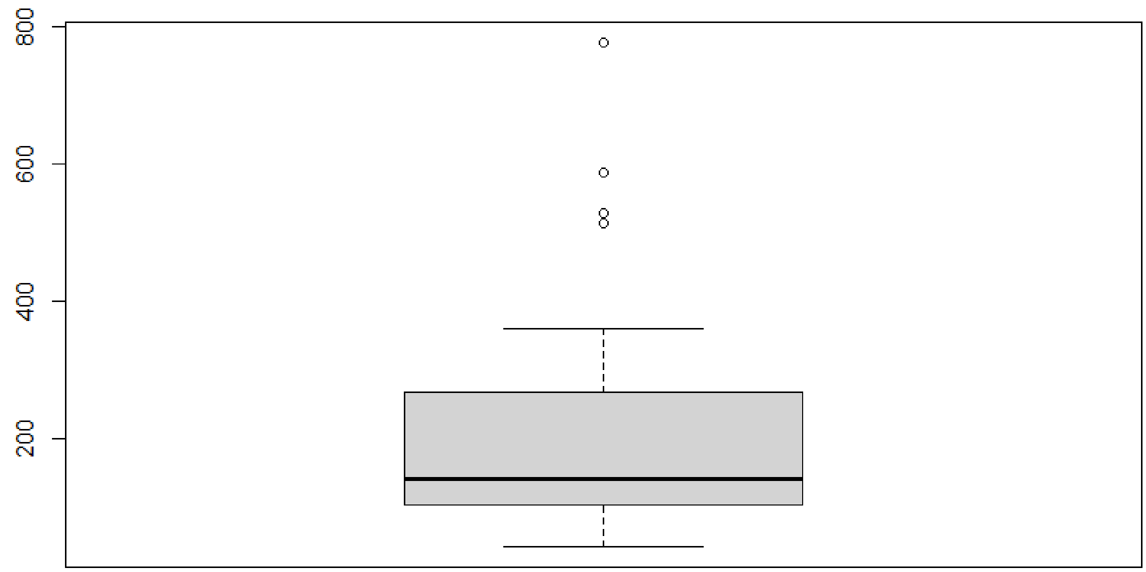


Figure 58: 2018 Cost Box Plot

After removing these outliers, the median cost is reduced from $140.60 to $136.80 in addition to the mean reducing from $207.40 to $162.40. This is calculated using the summary function in R which is provided in Figure 59.

```
 ï..Request             Cost
 Length:35          Min.   : 41.8
 Class :character   1st Qu.: 96.9
 Mode  :character   Median :136.8
                    Mean   :162.4
                    3rd Qu.:214.9
                    Max.   :359.1
```

Figure 59: November 2018 Clean Data Summary

Repeating the same process for the 2020 dataset, the histogram in Figure 60 also appears to have an outlier which we confirm in the boxplot in Figure 61.

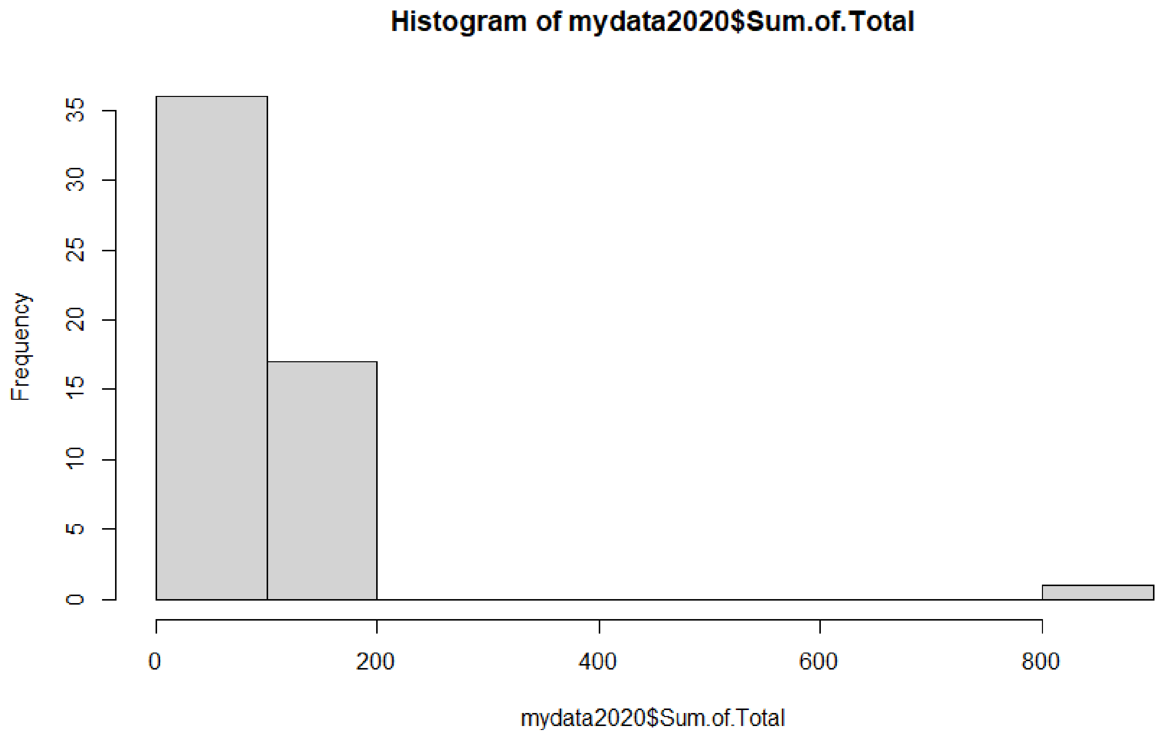


Figure 60: November 2020 Cost Histogram

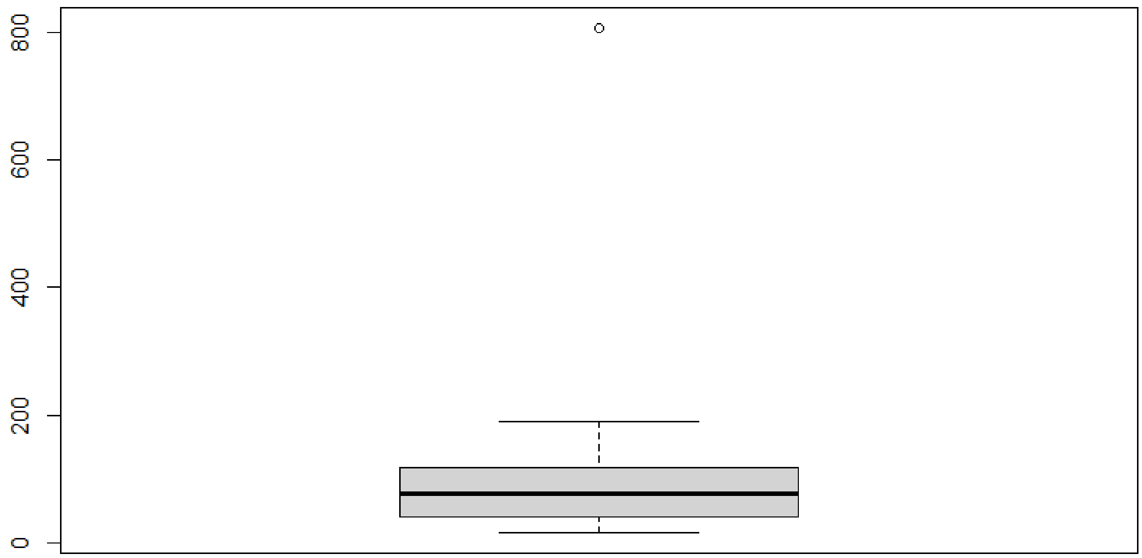


Figure 61: November 2020 Box Plot

After removing these outliers, we summarize the dataset in R again in Figure 62. Using these results, we can see the mean cost reduce from $95.88 to $82.49. The remaining testing below was performed against the sample sets less the outliers identified in this analysis.

```
ï..Row.Labels           Cost
Length:53          Min.   : 17.10
Class :character   1st Qu.: 41.80
Mode  :character   Median : 74.10
                   Mean   : 82.49
                   3rd Qu.:117.80
                   Max.   :190.00
```

Figure 62: November 2020 Clean Dataset Summary

4.3 Parameter of Interest and Hypothesis

While we have discussed the various benefits of RPA throughout this report, we have designed this experiment to use cost as the parameter of interest. For this analysis, we will study cost as the number of hours charged multiplied by the resource's hourly rate. In actuality, there are IT costs associated with this work, but they are fixed and not affected by this initiative.

As discussed in Chapter 1, we can represent this practical problem as a general statistical problem with a fully inclusive null and alternative hypothesis where the burden of proof is showing that improvement was really achieved:

Ho: There was no improvement to the output of the process.

Ha: There was improvement in the output of the process.

4.4 Test Statistic

In statistics, there are both parametric and non-parametric tests. Parametric tests perform analysis on the mean of sample sets and non-parametric tests are performed on the median value of sample sets. Parametric tests include the 1-Sample T-Test, 2-Sample T-Test, One-Way ANOVA, and MANOVA analysis. 1-Sample T-tests are designed to test how a sample compares to the whole group, 2-Sample T-tests are designed to test 2 samples against each other, ANOVA test compare 3 or more samples, and, finally, MANOVA analysis compares 3 or more samples against multiple variables. Non-parametric tests perform similar tests as the parametric tests while using the median values. Table 4 shows each parametric test and its corresponding non-parametric test(s).

Table 4: Parametric and Non-Parametric Tests

Parametric tests (means)	Nonparametric tests (medians)
1-sample t test	1-sample Sign, 1-sample Wilcoxon
2-sample t test	Mann-Whitney test
One-Way ANOVA	Kruskal-Wallis, Mood's median test
Factorial DOE with one factor and one blocking variable	Friedman test

Because we have 2 sample sets and are not comparing multiple variables, the appropriate parametric test would be the 2-Sample T-test and the corresponding non-parametric test is the Mann-Whitney test. Because parametric tests typically hold more statistical power than nonparametric tests, we will use the 2-sample T-test to test our hypothesis. It is important to note that T-tests assume that the data follows a normal distribution. However, the central limit theorem states that if a sample size is large enough, it can be assumed that the population follows a normal distribution. It is generally accepted that a sample size of 30 members is large enough for the central limit theorem. While both of our sample sets have a sample size of more than 30 data points, Figure 63 and Figure 64 provide a visualization of how our sample sets fit within a normal distribution.

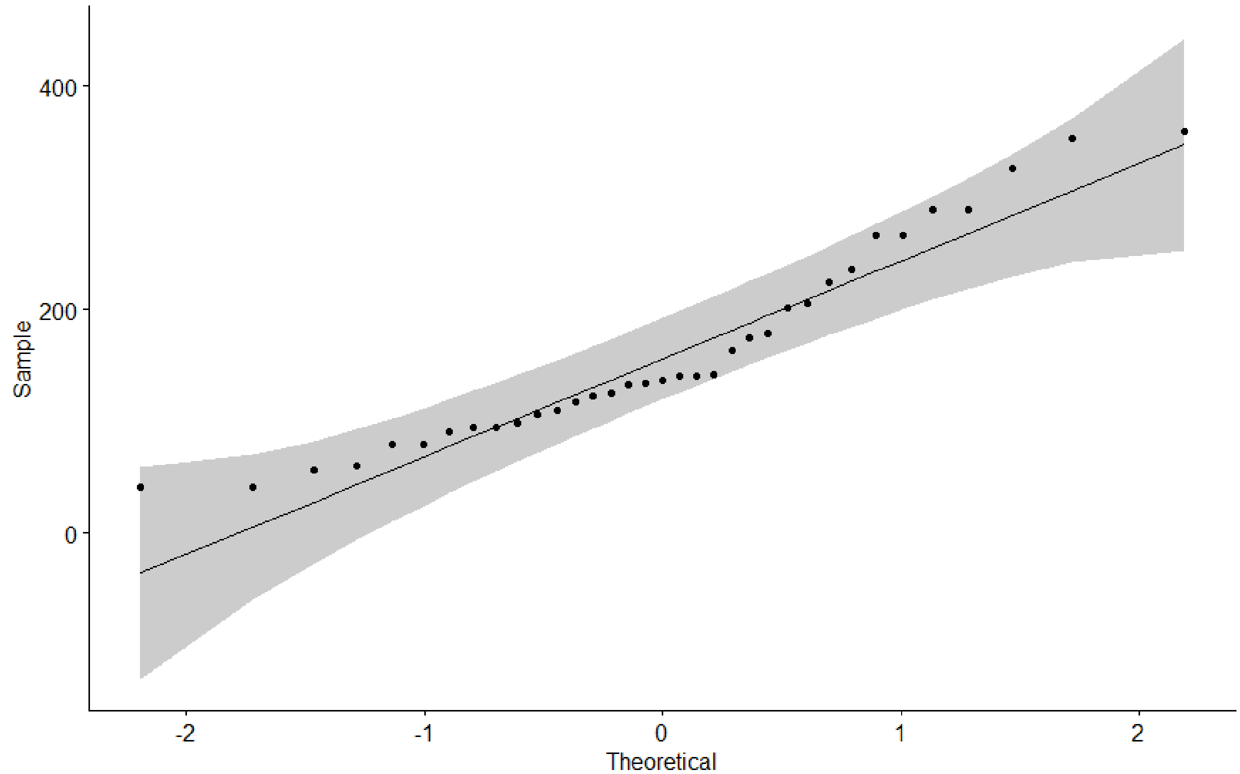


Figure 63: 2018 Data Quantile-Quantile (QQ) Plot

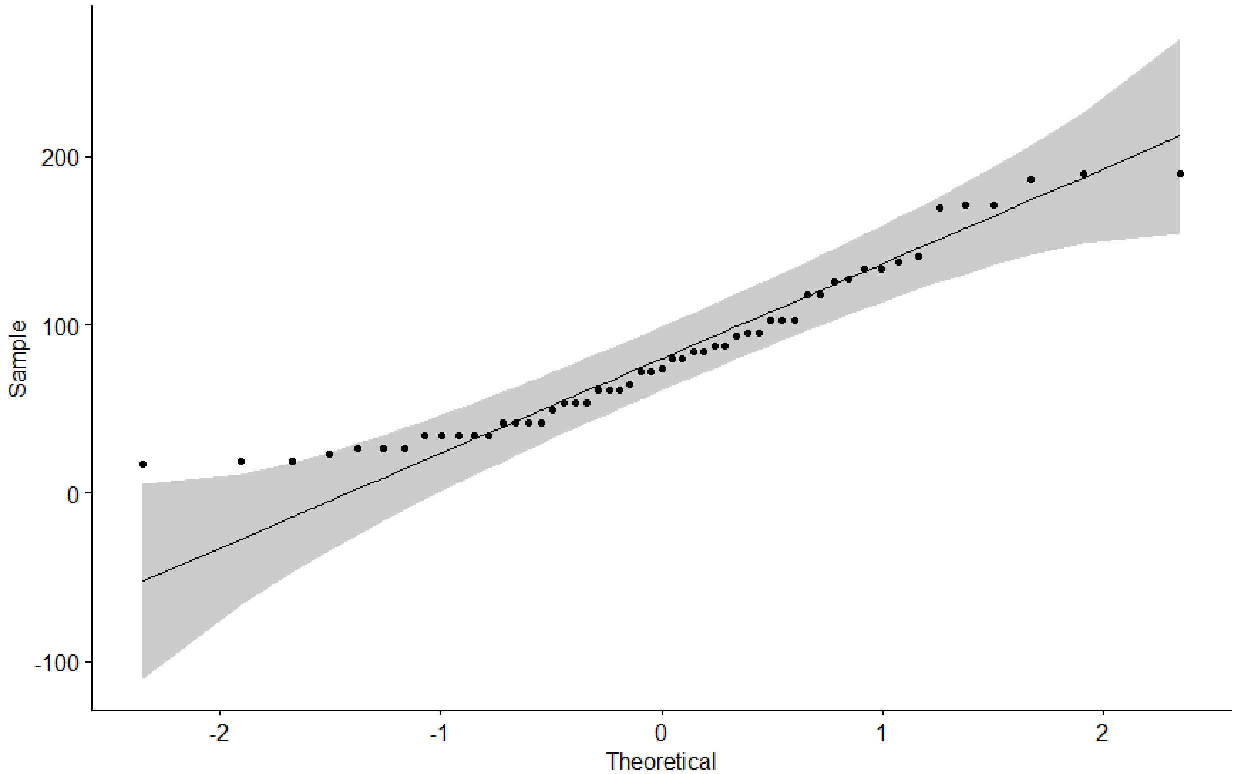


Figure 64: 2020 Data Quantile-Quantile (QQ) Plot

While Figure 63 appears to show that the 2018 sample data set is normally distributed, a Shapiro-Wilk's test was performed in R that returned a W value of 0.92047 and a p score of 0.01471. Similarly, the 2020 dataset was calculated to have a 0.003184 p score which we can see visualized in the histograms in Figure 65 and Figure 66 which we would expect to follow more of a bell-curve if normally distributed.

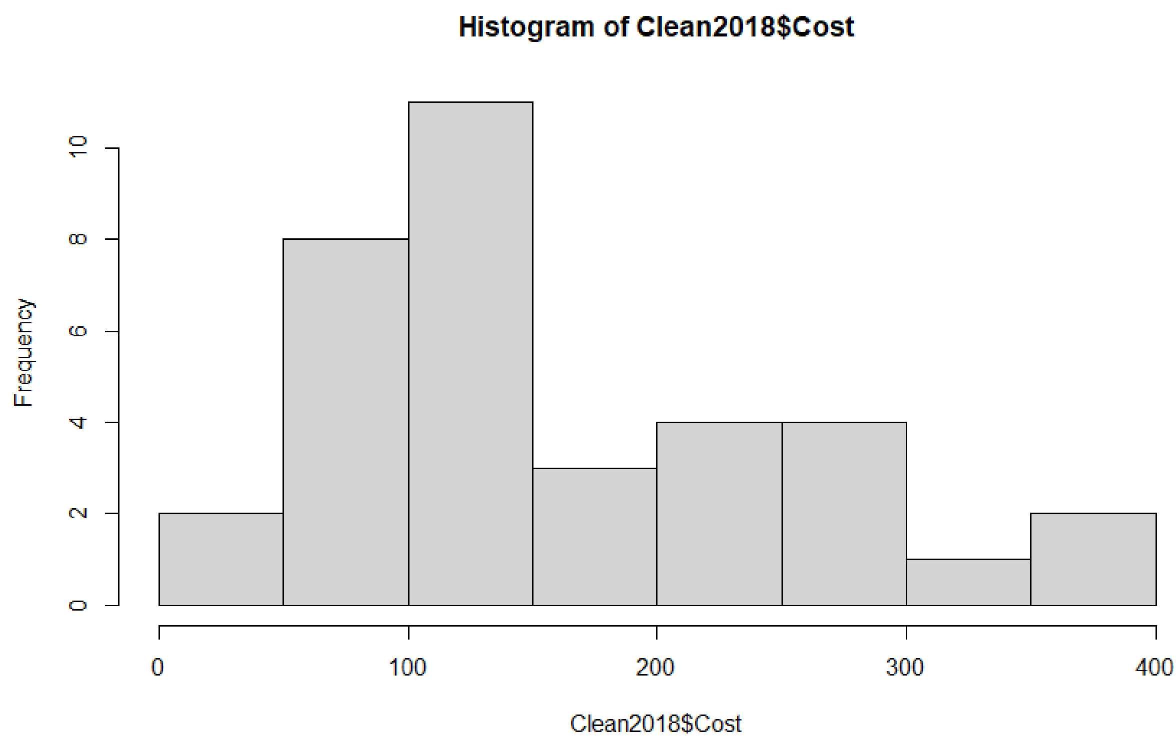


Figure 65: November 2018 Costs Less Outliers

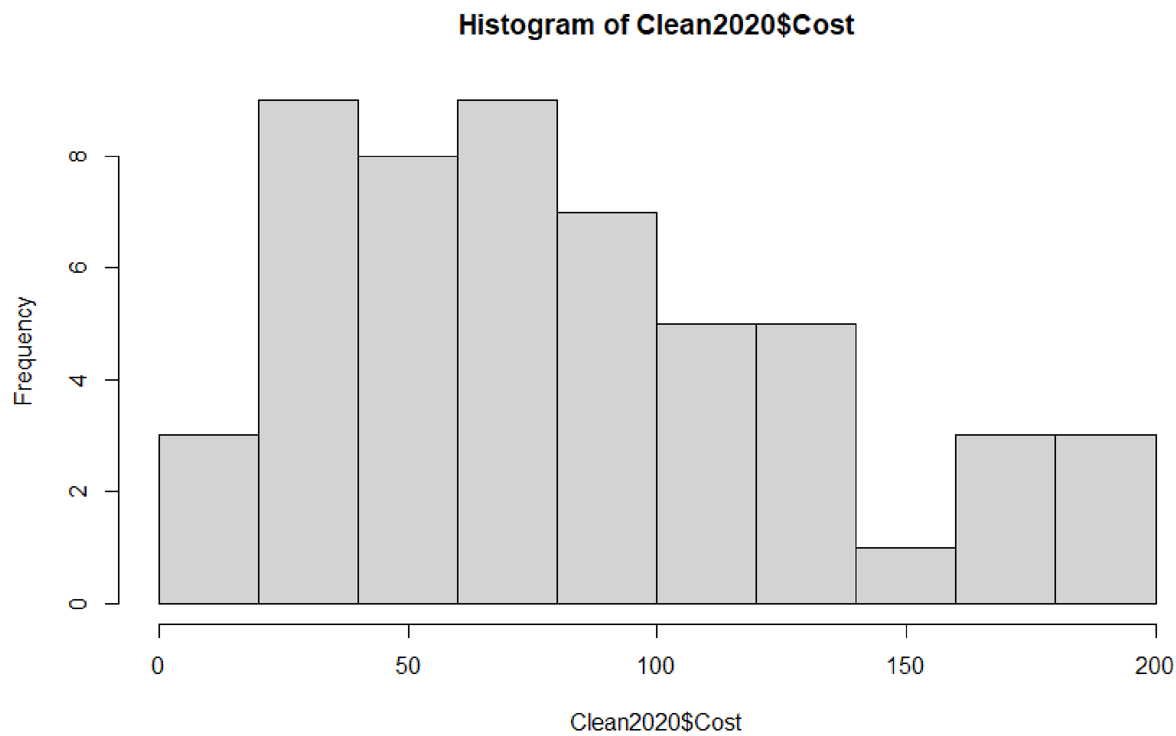


Figure 66: November 2020 Costs Less Outliers

Because both of our sample sets have more than 30 data points and the T-test holds more power, we will still rely on the 2-Sample T-test as our test statistic. However, we will also perform the corresponding Mann-Whitney non-parametric test to further validate our results in section 4.8.

4.5 Rejection Criteria

Because we are testing that the new process is cheaper than the original, this is a lower-tailed test and thus, we reject H_0 if $z_0 > z_{a/2}$ and we fail to reject H_0 if $- z_{a/2} \leq z_0 \leq z_{a/2}$.

4.6 Computations

It should be noted that the different types of T-tests have different equations to calculate the T-value. The equations below are specific to the 2-Sample T-test.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$DF = n_1 + n_2 - 2$$

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Using R, the t-value was calculated to be 4.872 with a p-value of 6.22×10^{-6}.

4.7 Conclusion

Using the p-value calculated in section 4.6, we can assume there is a less than 0.01% chance that the reduction in cost is due to a statistical error and we can therefore reject the null hypothesis that there is no improvement to the process. While these measures allow us to conclude there is a statistically significant improvement in the process, it is more impactful to understand the practical significance or “engineering significance” as the text denotes it. We will discuss this further in section 4.9.

4.8 Mann-Whitney Test

The Mann-Whitney Test, also known as the Wilcoxon rank-sum test, is designed to address if the probability of one sample exceeding a second sample is equal to the probability of the second sample exceeding the first. It can be calculated using the equation below:

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} S(X_i, Y_j),$$

with

$$S(X,Y) = \begin{cases} 1, & \text{if } X > Y, \\ \frac{1}{2}, & \text{if } Y = X, \\ 0, & \text{if } X < Y. \end{cases}$$

For our datasets, the p-value is calculated to be 1.828 x 10-6 which means there is less than a 0.01% chance that the difference in means of the 2 sample sets is due to statistical error.

4.9 Engineering Significance

While statistical significance can show a change between two datasets, the text refers to what is called "engineering significance" to refer to the significance a statistical difference has on operations. This section provides more details on the actual benefits realized through this initiative.

In addition to the reduction in cost from the direct labor resources, we were also able to realize an additional reduction in costs in executive oversight. From the US operations team alone, we were able to reduce the number of Full-Time Employees (FTEs) from 15 resources, including 6 executives with higher cost rates down to 3 FTEs by the end of 2018 which is shown in Figure 67.

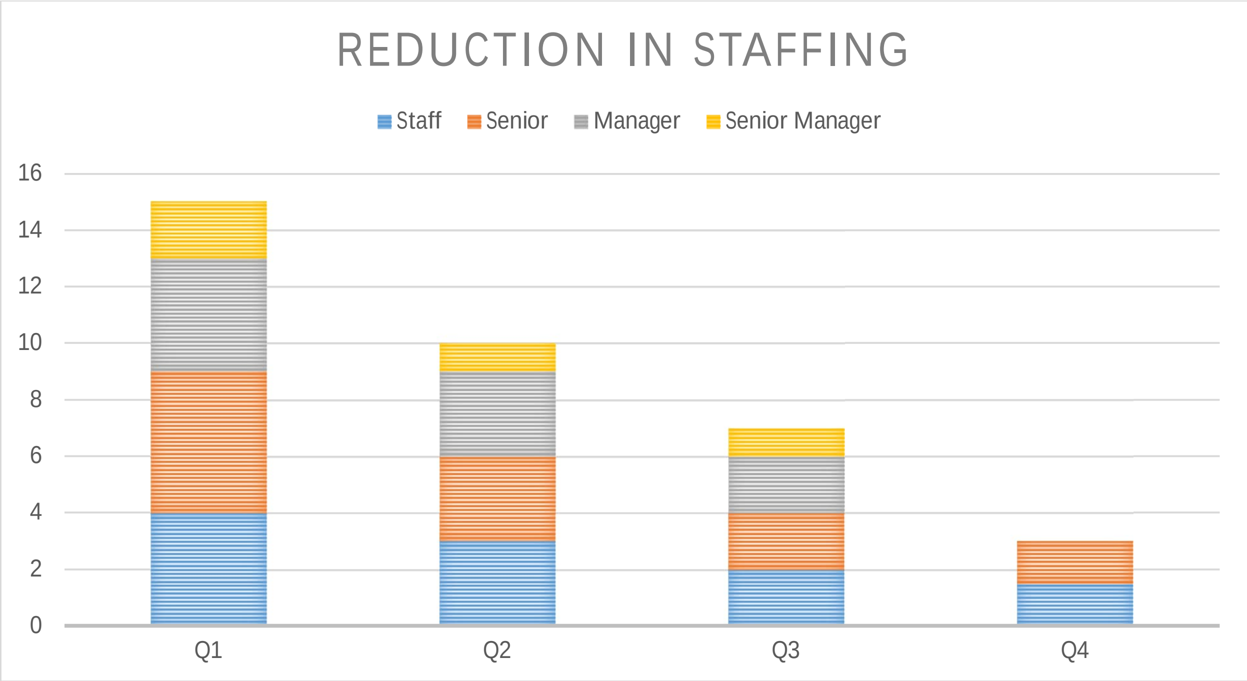


Figure 67: Reduction in Staffing Chart

4.9.1 Financial Impact Using Time Value of Money

In 2018, a total of $2,971,170 USD in savings was recognized based on actual resource time charged and cost rates. Using the Time Value of Money function taught in EMIS 8361 and EMIS 8363, we can convert that to today's dollars using the formula below:

FV = PV x [1 + (i / n)] (n x t)

Where:

- FV = Future value of money
- PV = Present value of money
- i = interest rate
- n = number of compounding periods per year
- t = number of years

Using this formula and assuming an interest rate of 3% for inflation, we can calculate the 2018 savings to be equivalent to $3,246,677.68 in today's dollars.

4.9.2 Speed Impact Using Queuing Theory

In EMIS 8372 – Queuing Theory, Little's Law is taught to calculate the average number of items in a queue (QueuingTheoryTextbook). This is a significant metric to have because it can be used to understand how much hardware needs to be available to have an appropriately sized infrastructure and IT environment. Little's Law states $L = \lambda W$, where L is equal to the average number of items in the queue, λ is the average arrival time for items in the queue, and W is the average amount of time the items spend in the system. Little's Law allows us to recognize that the amount of time items spend getting processed is directly related to the number of items in the queue. This translates into the number of IT resources needed to manage the queue as well as the amount of time spent in what is commonly known as "queue management." This reduction is illustrated in Figure 67 where the reduction in executive oversight time can be assumed is at least in part due to a reduction in time needed for queue management.

CHAPTER 5

SUMMARY AND DISCUSSION

5.1 Outline

Finally, this chapter concludes this research by highlighting this initiative's impact. It is organized using the four Student Learning Objectives (SLO) of the DE EM program:

- SLO1: Student demonstrates expert knowledge of the literature in a sub-area of Operations Research or Engineering Management.
- SLO2: Student demonstrates the ability to clearly explain and document his/her research in the discipline of Engineering Management.
- SLO3: Student demonstrates the ability to identify research directions independently in the discipline of Engineering Management.
- SLO4: Student demonstrates the potential economic impact of his/her praxis in the discipline of Engineering Management

5.2 Literature Review

Based on over 55 reviewed pieces of lieterature, there appears to be an opportunity to fill a gap in the research by applying Lean Six Sigma principles to RPA initiatives. Furthermore, it is exciting to provide an alternative solution to the publications that cite the need for data to be structured to be a candidate for an RPA solution.

5.3 Engineering Management Contribution

While designed for heuristic models, Dr. Barr's publication "Designing and Reporting on Computational Experiments with Heuristic Methods" provides a framework for evaluating research contribution: fast, accurate, robust, simple, high-impact, generalizable, and innovative.

- **Fast:** While the problem statement of this research was focused on cost, there was also significant time savings when it came to the actual implementations. For example, one process that previously took 3 full-time resources an average of 34 days to complete now completes in 2 hours with only 15 minutes of human input time.
- **Accurate:** In addition to significant cost and time savings, we have not encountered any quality issues from steps in our workflow that have been automated.
- **Robust and Simple:** As alluded to in Chapter 4, this work was specifically designed to be scalable and customizable which allowed it to be easily deployed across multiple geographies.
- **High-Impact:** Even at a firm as large as Ernst & Young, $15 million in savings is a significant impact to firm margin.
- **Innovative:** Finally, RPA is still an emerging technology and the use of it on unstructured data can certainly be considered innovative. Unsurprisingly, this work is presented on regularly in sessions focused on legal innovation and technology.

5.2.1 External Recognition

In addition to being presented to the praxis supervisory committee, this research has also been presented externally and received recognition:

- **RelativityFest 2018 Innovation Award Finalist –** The industry-leading review platform is named Relativity and they host a conference every year named RelativityFest that includes innovation awards in different categories. This research was presented in 2018 under the submission Relativity Automation Tool (RAT). The submission was named one of the top three finalists which means it was presented at an awards ceremony that included over 6,000 attendees.
- **Robotic Process Automation (RPA): Turning Buzzwords into Action –** This research was also prepared into a presentation on RPA that has been presented in multiple industry groups including the Association of Certified Fraud Examiners (ACFE), the Association of Certified eDiscovery Specialists (ACEDS), Summit on Legal Innovation and Disruption (SOLID), and Women in eDiscovery (WiE) in multiple cities including Washington DC, Dallas, and Austin. The presentation focuses on the applications within this research initiative and then invites participants to discuss potential applications and opportunities within their organizations. The abstract is presented as follows:

  Robotic Process Automation in eDiscovery: Putting Buzzwords into Action

  In the professional services industry, it is hard to go a day without hearing buzzwords like AI (Artificial Intelligence) and RPA (Robotics Process Automation). Research shows that while many organizations have included plans to implement these technologies in their roadmaps, very few have actually done so. This session is designed to highlight successful implementations of RPA in the eDiscovery process, including how they are implemented and the benefits they provide. Our speakers will cover the key factors required for successful planning and implementation of RPA in eDiscovery and the common misconceptions and pitfalls that hold many organizations back.

- **RelativityFest 2020 Innovation Award Finalist –** In 2020, again RPA using the Relativity platform was submitted – this time for an application marketed as Rebound which provided a centralized location for organizations to respond to the COVID-19 pandemic. What started as a web scraper to monitor guidance issued by the US government quickly escalated into a full-blown COVID-19 Command Center. The platform triangulates over 15,000 international data sources to monitor government regulations, tax guidance, travel advisories, as well as case and testing tracking. Additionally, the platform creates a repository to enable enterprise responses including internal and external communications, insurance policies, contracts, disaster recovery plans, recovery assessments, and other types of structured and unstructured data. Figure 68 provides a screenshot of the application's front interface, Figure 69 shows a visualization of the types of regulations and information the platform tracker, and Figure 70 provides a visualization of the timeline visual prepared. Having a timeline view was especially helpful given how frequently the regulations changed.

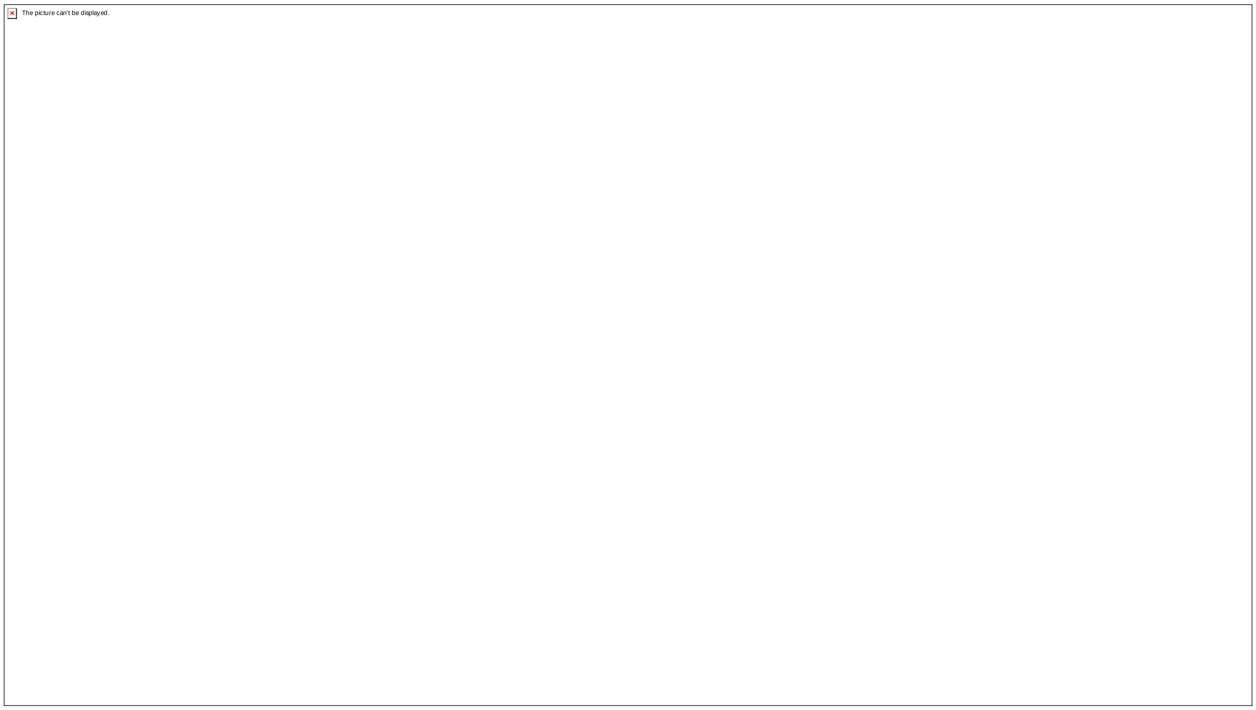


Figure 68: EY Rebound Screenshot

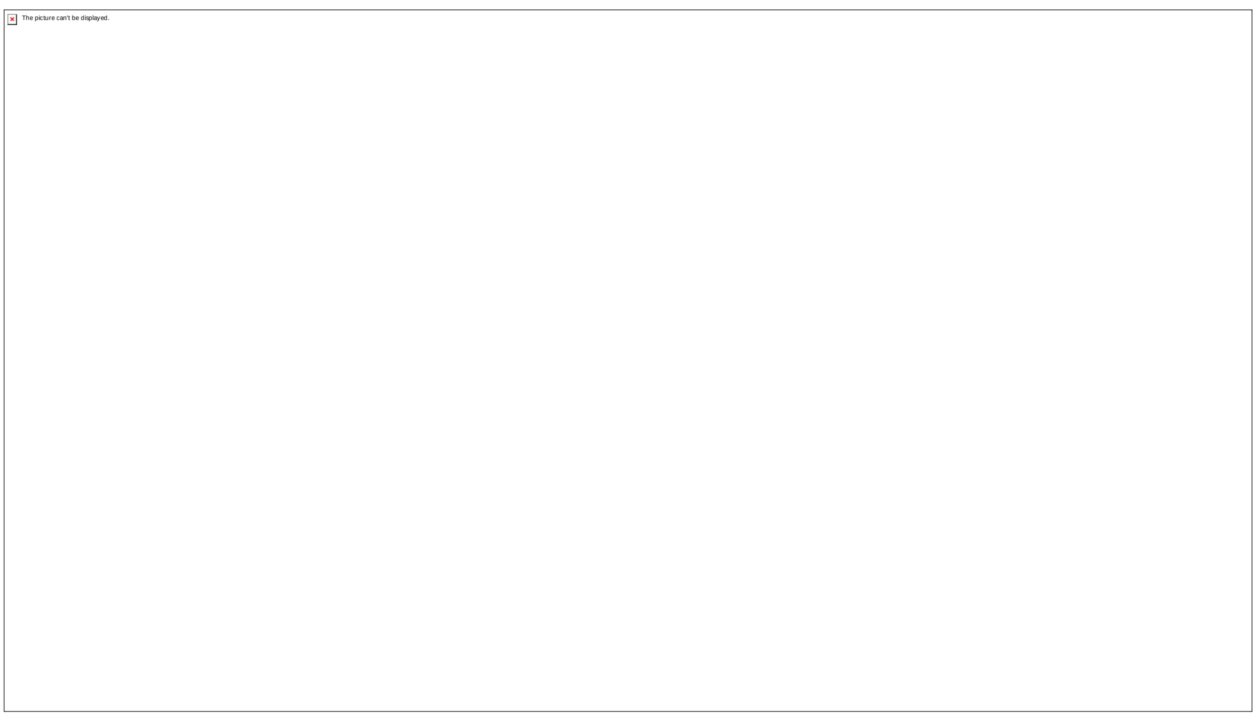


Figure 69: EY Rebound Screenshot

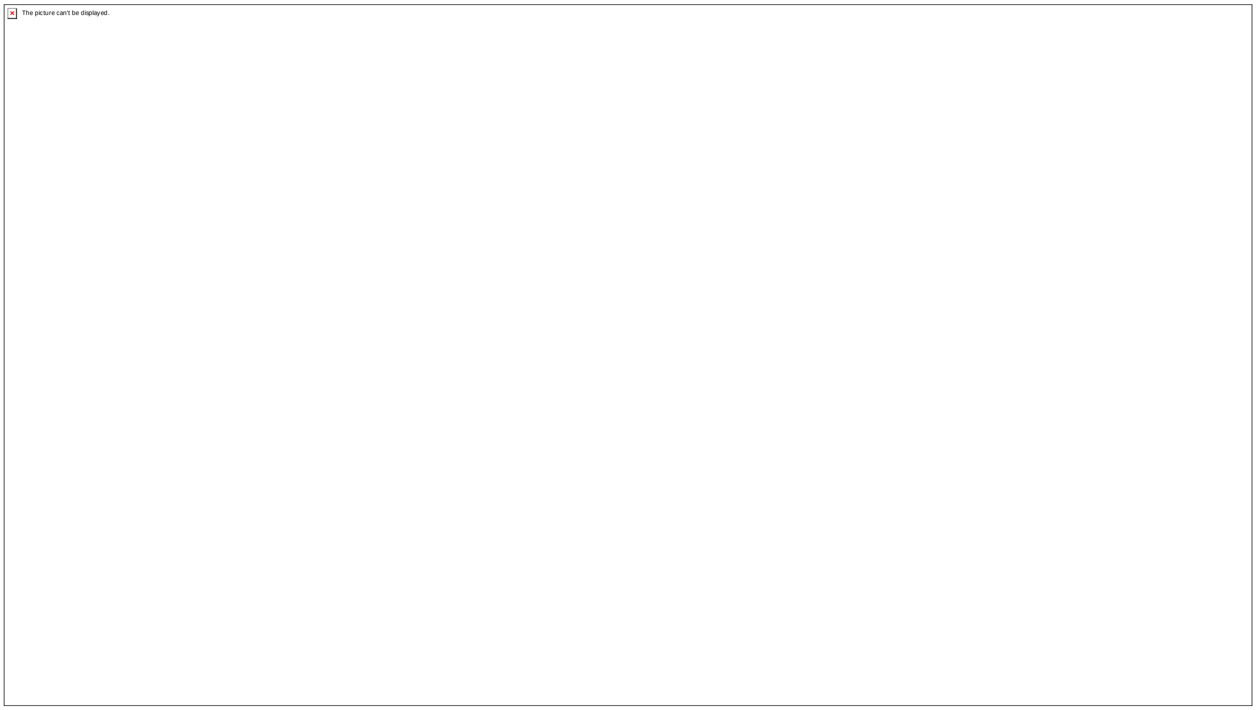


Figure 70: EY Rebound Screenshot

Once again, this submission was selected as a top 3 finalist and was presented to an audience of 14,000+ attendees.

- **The Artificially Intelligent Investigation –** This research contributed to a presentation to the Dallas ACFE chapter in September 2021. The presentation was promoted using its abstract below:

    The (Artificially) Intelligent Investigation

    - Demystify the different types of AI and how they can be used to not only identify fraudulent activity but automate the subsequent investigation
    - Learn how to get started using the 3 A's: Analytics, AI, and Automation

    - Walk through a tale of two investigations to understand the impact these technologies can have on an investigations

- **Summit On Legal Innovation and Disruption West–** This research directly attributed to a presentation given at SOLID West given in February 2022 titled Legal Tech: the movement to bring technology in-house. This presentation was promoted using the abstract below:

  There is an ongoing trend of companies buying added technology licenses from various vendors. As more technology moves in-house, there are many considerations that companies will need to address. Such considerations include:

    - Are we fully articulating the Legal Tech ROI to the broader business?
    - How should we manage to work with disparate systems and integration ideas for a robust view of the business?
    - Are we considering a holistic technology approach? Are we appropriately connected with the compliance function for a "proactive" and "reactive" view of the business?

Finally, this research applied for and was selected to be presented at the Institute for Operations Research and Management Science (INFORMS) Business Analytics Conference held in Houston, Texas from April 3 – April 5, 2022.

5.3 Research Directions

One of the most exciting things about this research, is the idea that it can be used to help unstructured data become structured. Data Volumes have increased roughly 5,000% since 2010 and an estimated 80 – 90% of data is unstructured. To make more of this data easier to consume can have significant impacts across multiple industries and sectors. Additionally, it would be interesting to study how the academic literature aligns to marketing materials promoted by companies. The assumption would be that academic publications do not include all of the innovation completed outside of academia but that professional marketing materials are not as vetted and tested (and therefore trusted) as publications for an academic audience.

Additionally, automation lends itself to having more audit history and more data points captured for analysis. For example, EY now has tracked information on when tasks are completed. This has allowed us to begin pulling metrics and drawing insights across matters into what we call portfolio insights. Historically, matters were reviewed and analyzed on an individual basis but with automation captureing and tracking this new level of detail, there is now an opportunity to look at more macro-level trends which we expect to shape the industry.

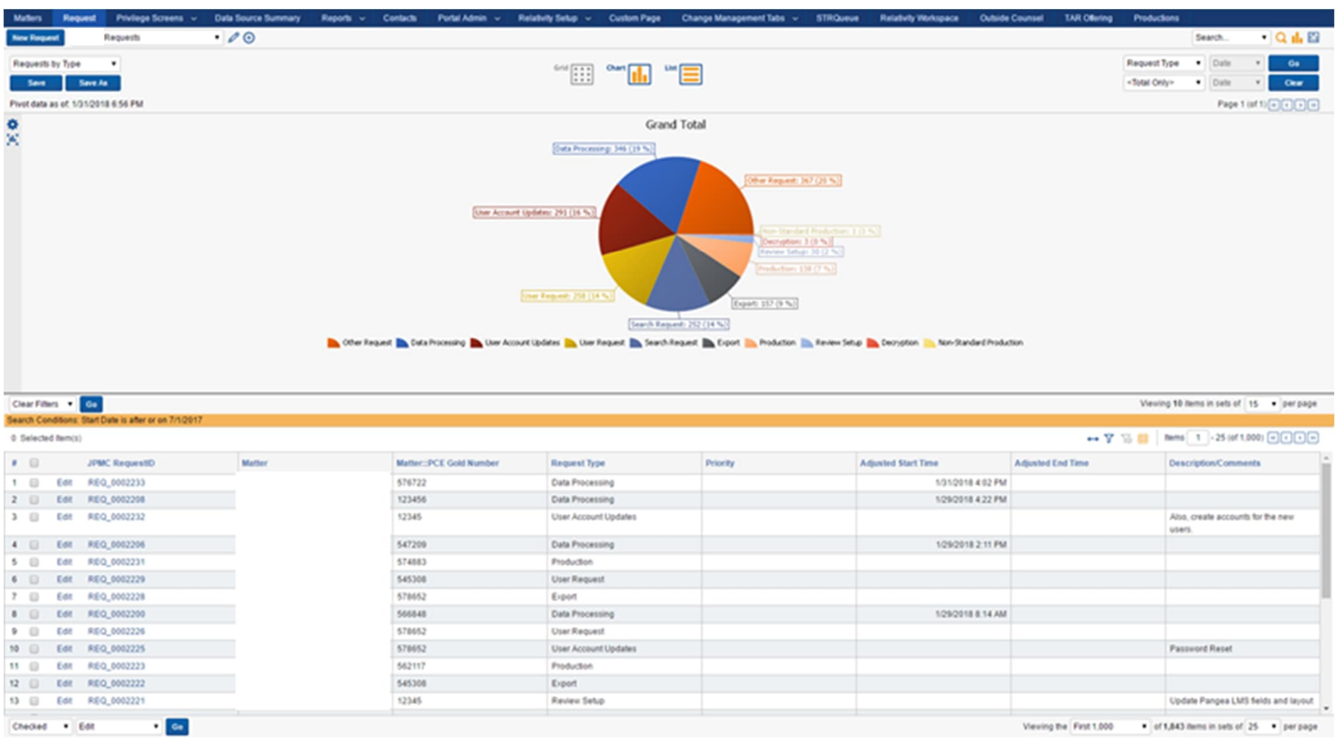


Figure 71: Activity Across Portfolio

For example, Figure 71 is a dashboard we created within our platform to show the types of activity across a portfolio of matters for the same client. This allows case managers to better appreciate activity at a wholistic level to better understand how their portfolio is performing. For example, this illustration shows that the majority of work is spent on data processing requests and now many production requests. This likely means that the portfolio is over-collecting data and paying for it to be processed and reviewed but not actually produced. This presents an opportunity for cost savings.

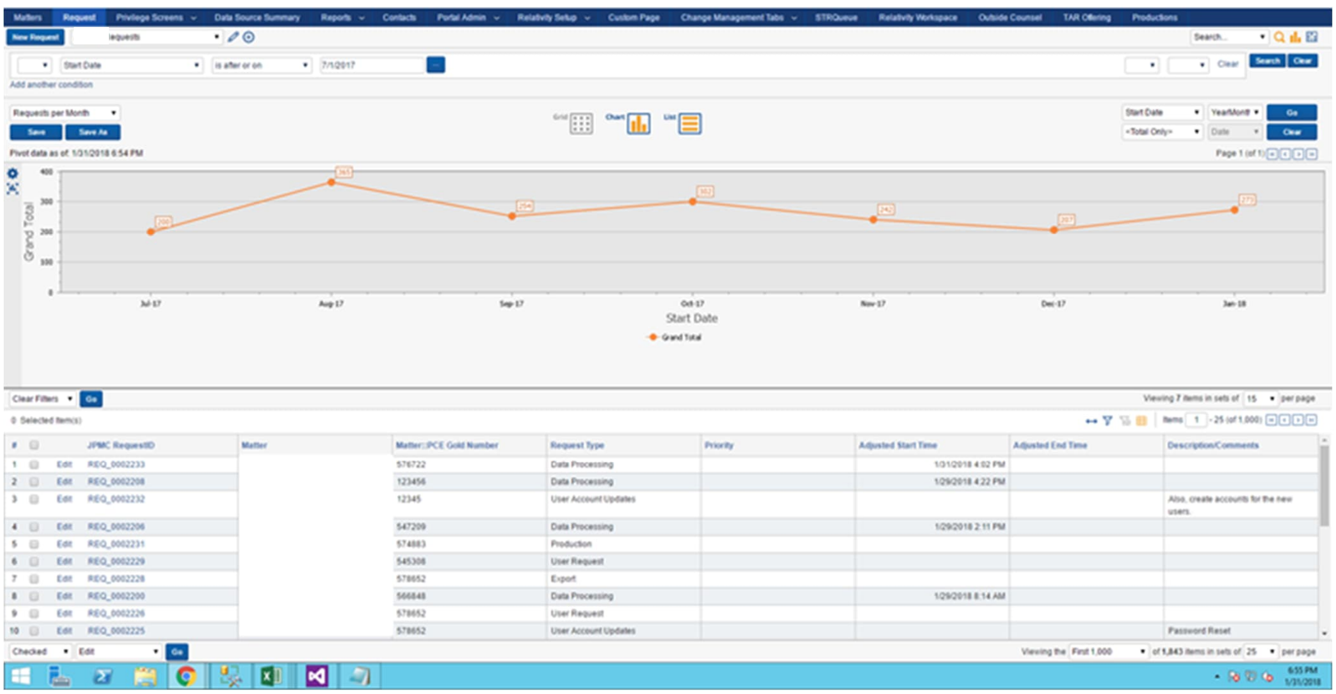


Figure 72: Activity Summary for Portfolio

Similar to Figure 71, Figure 72shows all activity within a queue regardless of matter. This allows case managers to understand activity across a portfolio and best manage the queue.

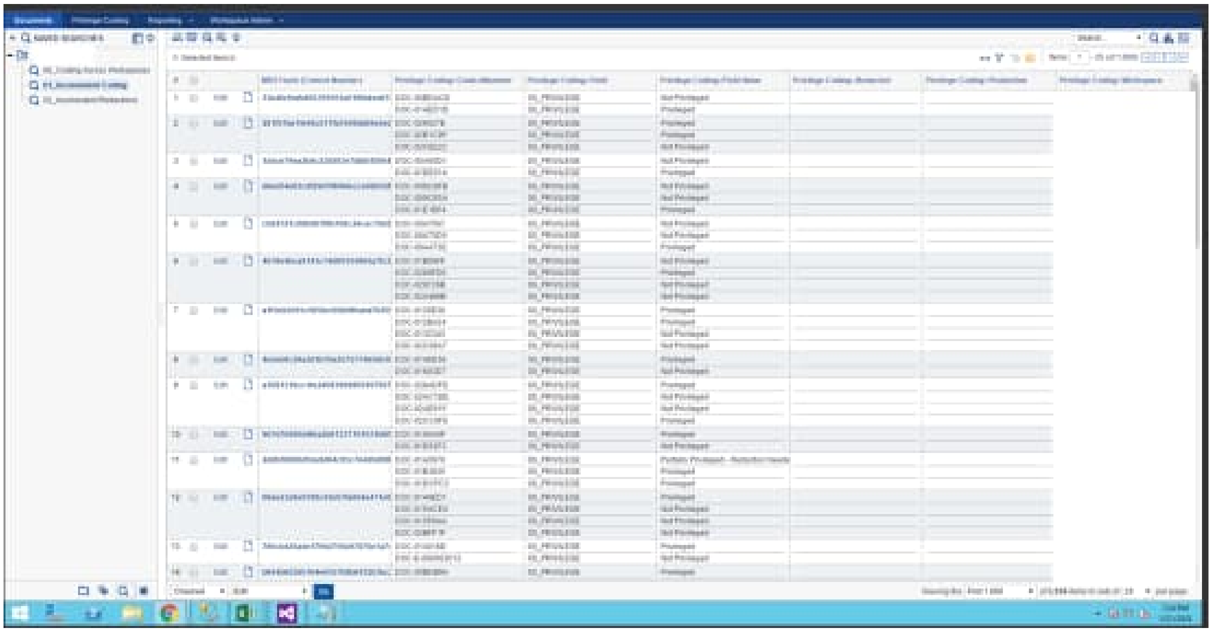


Figure 73: Privilege Bank

Figure 73 shows an example of tracking privilege information across an entire portfolio of matters. This is an interesting application of insights because it can potentially reduce the amount of documents needed to be reviewed by outside counsel which is typically the highest cost of litigation.

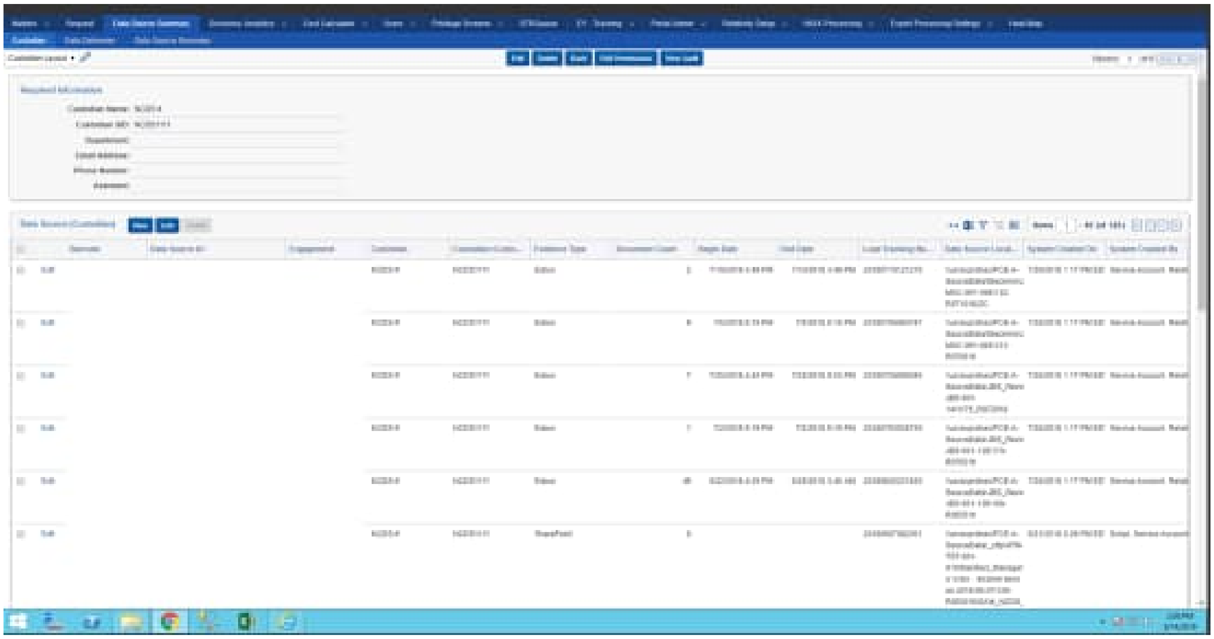


Figure 74: Custodian Summary

Figure 74 illustrates a view that was created to show the amount of data across a portfolio of matters for the same custodian. Leveraging data that has been previously collected, processed, and even reviewed or produced can provide significant time and cost savings.

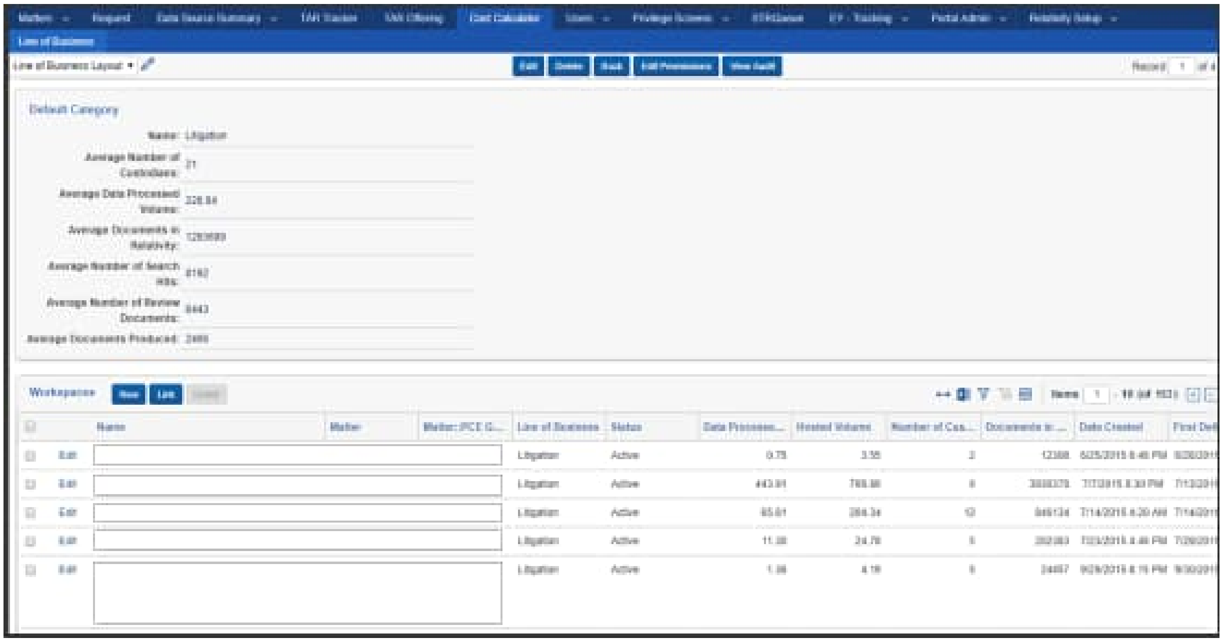


Figure 75: Cost Calculator

Figure 75 illustrates a view we created that pulls various metric points now available to help predict and project costs for matters as well as associate them back to a client's particular line of business.

537435	558050	561367	560067	565683	568095	558872	572816	571525	579669	583822	574499	589065	577700	598061	589633
555476	524595	549167	562211	565816	567971	553974	575327	576130	567118	547209	547209	581219	584215	422660	599204
544297	521457	430538	564609	504368	520507	547786	574633	565198	579576	544604	542221	575679	591937	597312	568825
554331	557525	558360	564430	566217	568563	571213	554572	576748	566685	547209	584092	588802	550627	596479	580859
545613	559345	561073	563930	567874	570101	566701	574955	577119	565457	584057	584723	589221	575364	552424	596479
548073	557438	545240	563548	552418	513632	563656	573979	552856	458782	580870	563143	588069	584191	582020	
555410	557441	561268	563811	557960	558532	572028	777777	576548	573075	582865	574784	587239	590718	586980	
549850_01	557434	561310	564698	567023	563436	558878	575636	575464	579098	581422	578349	579986	593900	595036	
548733_01	557440	535988	565149	565961	557251	566664	560690	575740	580272	579187	547991	582453	581966	590877	
535674	557429	562629	563844	552109	555862	543659	574094	564011	578153	583163	578345	569393	592400	555856	
546492	547223	562593	535183	546611	562602	546636	561726	576983	574469	582704	540984	588518	566300	584613	
553835	546767	453044	552636	525668	542727	570271	572468	573160	541503	583829	585854	575923	595042	583285	
545528	555940	549604	557112	567666	570262	570703	563393	575863	581673	581329	574883	536713	594109	552565	
538335	533053	582096	517829	466904	568325	572831	574345	491486	578291	581547	583280	589864	588649	596787	
554973	541302	562331	541302	566767	437589	572871	564641	575703	576722	580802	585906	582741	581149	596818	
549768	436741	558878	562867	566974	563461	556114	573641	546491	477450	582946	576303	576595	593604	596296	

Figure 76: Matter Heatmap

Finally, Figure 76 visualizes the activity across all matters within a portfolio. The darker colors represent a matter with more activity. Figure 76 clearly illustrates that while this is a large portfolio of matters, very few are actually active. This type of insight allows for better load-balancing across IT resources which improves the reviewer experience and efficiency which ultimately can lead to cost savings.

As more and more data points are tracked and stored, we expect to be able to pull more insights which ultimately can lead to more testing and data-driven decisions.

5.4 Financial Impact

And finally, the financial impact of this project is satisfying to calculate. Because of the cost savings recognized in 2018, practice leadership made the determination to create a combined Service Delivery Center (SDC) which has more than doubled the amount of data processed using this workflow and leads us to confidently say the firm has recognized more than $15,000,000 in savings due to these automations and workflow refinements.

In addition to the direct labor savings we have discussed, we routinely receive feedback that our more junior staff enjoy not having to do as much of the tedious work they used to. I have been told by multiple employees on my team that the only reason they have stayed at the firm as long as they have is to continue to work on my team and innovating. It would be fascinating to see if there are additional savings due to employee retention. Unfortunately, this information is considered confidential to the firm and unavailable to me.

BIBLIOGRAPHY

19 Sedona Conference. (2018). The Sedona Principles, Third Edition: Best Practices, Recommendations & Principles for Addressing Electronic Document Production. The Sedona Conference Journal, 118.

Aguirre, S., Rodriguez, A.: Automation of a business process using robotic process automation (RPA): a case study. In: Figueroa-García, J.C., López-Santana, E.R., Villa-Ramírez, J.L., Ferro-Escobar, R. (eds.) WEA 2017. CCIS, vol. 742, pp. 65–71. Springer, Cham (2017).

American Bar Association. (n.d.). Retrieved September 12, 2021, from https://www.americanbar.org/content/dam/aba-cms-dotorg/products/inv/book/214612/Chapter%201.pdf

Anagnoste, S.: Robotic automation process – the operating system for the digital enterprise. In: Proceedings of the International Conference on Business Excellence, vol. 12, no. 1, pp. 54–69, De Gruyter, Poland (2018)

Anagnoste, S.: Robotic automation process-the next major revolution in terms of back office operations improvement. In: Proceedings of the International Conference on Business Excellence, vol. 11, no. 1, pp. 676–686, De Gruyter Open (2017)

Asatiani, A., Penttinen, E.: Turning robotic process automation into commercial success – case OpusCapita. J. Inf. Technol. Teach. Cases 6, 67–74 (2016)

ASME Book Committee. (2010). Guide to the Engineering Management Body of Knowledge. American Society of Mechanical Engineers.

Bhutta, Z. (2020, October 22). RPA: Untapped Potential For Law Firms. Retrieved from Forbes: https://www.forbes.com/sites/forbestechcouncil/2020/10/22/rpa-untapped-potential-for-law-firms/?sh=789884923f5a

Bhutta, Z. (2020, October 22). RPA: Untapped Potential For Law Firms. Retrieved from Forbes: https://www.forbes.com/sites/forbestechcouncil/2020/10/22/rpa-untapped-potential-for-law-firms/?sh=789884923f5a

Blair, S. A. (2011, October). Using Lean Six Sigma And Predictive Coding To Confront Volume Problem. Retrieved from The Legal Intelligencer: http://www.evergreeneditions.com/article/Using+Lean+Six+Sigma+And+Predictive+Coding+To+Confront+Volume+Problem/871859/86147/article.html

Blair, S. A. (2011, October). Using Lean Six Sigma And Predictive Coding To Confront Volume Problem. Retrieved from The Legal Intelligencer: http://www.evergreeneditions.com/article/Using+Lean+Six+Sigma+And+Predictive+Coding+To+Confront+Volume+Problem/871859/86147/article.html

Bosilj Vukšić, V., Brkić, Lj., Tomičić-Pupek, K.: Understanding the success factors in adopting business process management software: case studies. Interdisc. Description Complex Syst. 16(2), 194–215 (2018)

Bourgouin, A., Leshob, A., Renard, L.: Towards a process analysis approach to adopt robotic process automation. In: ICEBE 2018 - 15th International Conference on e-Business Engineering, pp. 46–53. IEEE (2018)

Cewe, C., Koch, D., Mertens, R.: Minimal effort requirements engineering for robotic process automation with test driven development and screen recording. In: Teniente, E., Weidlich, M. (eds.) BPM 2017. LNBIP, vol. 308, pp. 642–648. Springer, Cham (2018).

Chappell, D.: Introducing blue prism: automating business processes with presentation integration (2010).

D. Baviskar, S. Ahirrao, V. Potdar and K. Kotecha, "Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence: A Systematic Literature Review and Future Directions," in IEEE Access, vol. 9, pp. 72894-72936, 2021, doi: 10.1109/ACCESS.2021.3072900.

Dunlap, R., Lacity, M.: Resolving tussles in service automation deployments: service automation at Blue Cross Blue Shield North Carolina (BCBSNC). J. Inf. Technol. Teach. Cases 7(1), 29–34 (2017)

EDRM. (2010, December 9). Review Guide. Retrieved September 12, 2021, from EDRM: https://edrm.net/resources/frameworks-and-standards/edrm-model/review-guide/

EDRM. (2010, November 3). Identification. Retrieved September 12, 2021, from EDRM: https://edrm.net/resources/frameworks-and-standards/edrm-model/identification/

EDRM. (2020). Preservation Guide. Retrieved September 12, 2021, from EDRM: https://edrm.net/resources/frameworks-and-standards/edrm-model/preservation/

EDRM. (2020). Processing Guide. Retrieved September 12, 2021, from EDRM: https://edrm.net/resources/frameworks-and-standards/edrm-model/processing/

Electronic Discovery Reference Model. (2020). EDRM Model. Retrieved September 11, 2021, from EDRM: https://edrm.net/wiki/edrm-model/

Encompass Contributor. (2017, October 12). An Introduction to Six Sigma. Retrieved from https://www.nelsonmullins.com/idea_exchange/blogs/encompass-blog/document_review/an-introduction-to-six-sigma

Encompass Contributor. (2017, October 12). An Introduction to Six Sigma. Retrieved from https://www.nelsonmullins.com/idea_exchange/blogs/encompass-blog/document_review/an-introduction-to-six-sigma

Ernst & Young. (2021). 2021 EY Law Survey. 3.

Everest Group: Defining Enterprise RPA (2018). https://www.uipath.com/company/rpa-analyst-reports/defining-enterprise-rpa-everest-research-report. Accessed 21 Sep 2021

Everest Group: Robotic process automation annual report 2018-creating business value in a digital-first word (2018). https://www2.everestgrp.com/reports/EGR-2018-38-R-2691. Accessed 21 Sep 2021

Evolution of Robotic Process Automation (RPA): The Path to Cognitive RPA | by AIMDek Technologies | Medium, Dec. 2020, [online] Available: https://medium.com/@AIMDekTech/evolution-ofrobotic-process-automation-the-path-to-cognitive-rpa-c3bd52c8b865.

Exigent. (2020, May 21). 6 Use Cases for AI and RPA in Law Firms and Legal Departments. Retrieved September 30, 2021, from Exigent: https://www.exigent-group.com/blog/use-cases-ai-rpa-law-firms-legal-departments/

Exigent. (2020, May 21). 6 Use Cases for AI and RPA in Law Firms and Legal Departments. Retrieved September 30, 2021, from Exigent: https://www.exigent-group.com/blog/use-cases-ai-rpa-law-firms-legal-departments/

Forrester: Building a center of expertise to support robotic automation: preparing for the life cycle of business change (2014). http://neoops.com/wp-content/uploads/2014/03/Forrester-RA-COE.pdf. Accessed 22 Sep 2021

Forrester: The Forrester Wave™: Robotic Process Automation, Q1 2017 - The 12 Providers That Matter Most and How They Stack Up, Forrester Research, Inc. (2017). https://www.forrester.com/report/The+Forrester+Wave+Robotic+Process+Automation+Q1+2017/-/E-RES131182. Accessed 5 Mar 2019

Gartner Glossary. (2021). Retrieved September 12, 2021, from Gartner: https://www.gartner.com/en/information-technology/glossary/information-governance

Gartner. (2018). Beynd Tactical RPA. Gartner.

Gartner. (2021). Robotic Process Automation Software Reviews and Ratings. Retrieved from Gartner: https://www.gartner.com/reviews/market/robotic-process-automation-software

Gejke, C.: A new season in the risk landscape: connecting the advancement in technology with changes in customer behaviour to enhance the way risk is measured and managed. J. Risk Manag. Finan. Inst. 11(2), 148–155 (2018)

GO Lean Six Sigma. (n.d.). Voice of the Customer (VOC) Tree Diagram. Retrieved December 19, 2021, from GO Lean Six Sigma: https://goleansixsigma.com/tree-diagram/

Greyer-Klingeberg, J., Nakladal, J., Baldauf, F.: Process mining and robotic process automation: a perfect match. In: 16th International Conference on Business Process Management, Sydney, Australia (2018)

Gupta, P., Fernandes, S.F., Jain, M.: Automation in recruitment: a new frontier. J. Inf. Technol. Teach. Cases 8(2), 118–125 (2018)

Hallikainen, P., Bekkhus, R., Pan, S.L.: How OpusCapita used internal RPA capabilities to offer services to clients. MIS Q. Exec. 17(1), 41–52 (2018)

Harmon, P.: The State of the BPM market – 2018 (2018). https://www.redhat.com/cms/managed-files/mi-bptrends-state-of-bpm-2018-survey-analyst-paper-201803-en.pdf. Accessed 21 Sep 2021

Hessing, T. (n.d.). Cause and Effect Diagram (aka Ishikawa, Fishbone). Retrieved January 16, 2022, from Six Sigma Study Guide: https://sixsigmastudyguide.com/cause-effect-diagram-aka-ishikawa-fishbone-herringbone-fishikawa/

Hessing, T. (n.d.). Classic Wastes. Retrieved January 16, 2022, from Six Sigma Studying Guide: https://sixsigmastudyguide.com/classic-wastes/

Institute for Robotic Process Automation: Introduction to Robotic Process Automation (2015). https://irpaai.com/wp-content/uploads/2015/05/Robotic-Process-Automation-June2015.pdf. Accessed 21 Sep 2021

Issac, R., Muni, R., Desai, K.: Delineated analysis of robotic process automation tools. In: ICAECC 2018 - Second International Conference on Advances in Electronics, Computers and Communications, pp. 1–5. IEEE (2018)

J. Siderska, "Robotic process automation—A driver of digital transformation?", Eng. Manage. Prod. Services, vol. 12, no. 2, pp. 21-31, Jun. 2020.

Kirchmer, M.: Robotic process automation - pragmatic solution or dangerous illusion?. Business Transformation & Operational Excellence World Summit (BTOES) (2017)

Kobayashi, T., Nakashima, R., Uchida, R., Arai, K.: SNS door phone as robotic process automation. In: Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces, pp. 457–460. ACM (2018)

Kulbacki, M., et al.: Survey of drones for agriculture automation from planting to harvest. In: INES 2018 22nd International Conference on Intelligent Engineering Systems, pp. 353–358. IEEE (2018)

Lacity, M., Willcocks, L.P.: Robotic process automation at telefónica O2. MIS Q. Exec. 15, 21–35 (2016)

Lacity, M., Willcocks, L.P.: Robotic process automation: the next transformation lever for shared services. In: The Outsourcing Unit Working Research Paper Series, Paper 16/01 (2016). http://www.umsl.edu/~lacitym/OUWP1601.pdf. Accessed 21 Sep 2021

Leno, V., Dumas, M., Maggi, F.M., La Rosa, M.: Multi-perspective process model discovery for robotic process automation. In: CEUR Workshop Proceedings, vol. 2114, pp. 37–45 (2018)

Leopold, H., van der Aa, H., Reijers, H.A.: Identifying candidate tasks for robotic process automation in textual process descriptions. In: Gulden, J., Reinhartz-Berger, I., Schmidt, R., Guerreiro, S., Guédria, W., Bera, P. (eds.) BPMDS/EMMSAD -2018. LNBIP, vol. 318, pp. 67–81. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91704-7_5

Lin, S.C., Shih, L.H., Yang, D., Lin, J., Kung, J.F.: Apply RPA (Robotic Process Automation) in Semiconductor Smart Manufacturing. In: 2018 e-Manufacturing and Design Collaboration Symposium (eMDC), pp. 1–3. IEEE (2018)

McLean, R., Antony, J.: Why continuous improvement initiatives fail in manufacturing environments? a systematic review of the evidence. Int. J. Prod. Perform. Manag. 63(3), 370–376 (2014)

Mendling, J., Decker, G., Hull, R., Reijers, H.A., Weber, I.: How do machine learning, robotic process automation, and blockchains affect the human factor in business process management? Commun. Assoc. Inf. Syst. 43, 297–320 (2018)

Ojala, A., Helander, N.: Value creation and evolution of a value network: a longitudinal case study on a platform-as-a-service provider. In: 47th Hawaii International Conference on System Sciences (HICSS). IEEE (2014)

Okoli, C.: A guide to conducting a standalone systematic literature review. Commun. Assoc. Inf. Syst. 37(43), 879–910 (2015)

Ovum: Robotic Process Automation: Adding to the Process Transformation Toolkit - The role that RPA can play within service providers and enterprises (2015). http://research.globalriskcommunity.com/content76913. Accessed 21 Sep 2021

P. Martins, F. Sa, F. Morgado and C. Cunha, "Using machine learning for cognitive robotic process automation (RPA)", Proc. 15th Iberian Conf. Inf. Syst. Technol. (CISTI), pp. 1-6, Jun. 2020.

Penttinen, E., Kasslin, H., Asatiani, A.: How to choose between robotic process automation and back-end system automation?. In: 26th European Conference on Information Systems (2018)

Petrov, C. (2021, October 2). 25+ Impressive Big Data Statistics for 2021. Retrieved from Tech Jury: https://techjury.net/blog/big-data-statistics/#gref

Production Guide. (2010, November 4). Retrieved September 12, 2021, from EDRM: https://edrm.net/resources/frameworks-and-standards/edrm-model/production/

Purdue University. (2021, May 28). DMAIC vs. DMADV. Retrieved from Purdue University: https://www.purdue.edu/leansixsigmaonline/blog/dmaic-vs-dmadv/

Rajesh, K.V.N., Ramesh, K.V.N.: Robotic process automation: a death knell to dead-end jobs? CSI Commun. Knowl. Dig. IT Commun. 42(3), 10–14 (2018)

Ratia, M., Myllärniemi, J., Helander, N.: Robotic process automation - creating value by digitalizing work in the private healthcare. In: ACM International Conference Proceeding Series, International Academic Mindtrek Conference (2018)

Rehan Syeda, S. S. (2019). Robotic Process Automation: Contemporary themes and challenges. Computers in Industry, 1,9.

Relativity. (2018, September 19). Relativity. Retrieved from Innovation Awards Technology Submissions: https://www.relativity.com/blog/innovation-awards-technology-submissions-ii/

Relativity. (2018, September 19). Relativity. Retrieved from Innovation Awards Technology Submissions: https://www.relativity.com/blog/innovation-awards-technology-submissions-ii/

Runger, D. C. (2011). Probability and Statistics for Scientists and Engineers (Fifth ed.). Hoboken, New Jersey, United States of America: John Wiley & Sons, Inc.

S. Burnett, D. Analyst, A. Verma, S. Analyst and P. Srinivasan, Unstructured data process automation a deep dive into the role of artificial intelligence (AI) in automating content-centric processes, Dallas, TX, USA, 2019.

Scheer, A.W.: Performancesteigerung durch Automatisierung von Geschäftsprozessen, AWS Institut fur digitale Produkte und Prozesse gGmbH (AWSi) (2017). https://www.aws-institut.de/wp-content/uploads/2017/11/031117_GPPerformance_44seiten_final_300dpi_2Aufl_einzel.pdf. Accessed 5 Mar 2019

Schmider, J., Kumar, K., LaForest, C., Swankoski, B., Naim, K., Caubel, P.M.: Innovation in pharmacovigilance: use of artificial intelligence in adverse event case processing. Clin. Pharmacol. Ther. 105(4), 954–961 (2019)

Schmitz, M., Dietze, C., Czarnecki, C.: Enabling digital transformation through robotic process automation at Deutsche Telekom. In: Urbach, N., Röglinger, M. (eds.) Digitalization Cases. MP, pp. 15–33. Springer, Cham (2019).

Suri, V.K., Elia, M., van Hillegersberg, J.: Software bots - the next frontier for shared services and functional excellence. In: Oshri, I., Kotlarsky, J., Willcocks, Leslie P. (eds.) Global Sourcing 2017. LNBIP, vol. 306, pp. 81–94. Springer, Cham (2017).

TCDI. (2018, June 15). Legal Spend Analytics: Manage Your Costs Using Lean Six Sigma. Retrieved from TCDI: https://www.tcdi.com/legal-spend-analytics/

TCDI. (2018, June 15). Legal Spend Analytics: Manage Your Costs Using Lean Six Sigma. Retrieved from TCDI: https://www.tcdi.com/legal-spend-analytics/

The Basics: What is e-Discovery? (n.d.). Retrieved September 11, 2021, from Complete Discovery Source: https://cdslegal.com/knowledge/the-basics-what-is-e-discovery/

Tranfield, D., Denyer, D., Smart, P.: Towards a methodology for developing evidence informed management knowledge by means of systematic review. Br. J. Manag. 14(3), 207–222 (2003)

Tsaih, R., Hsu, C. C.: Artificial intelligence in smart tourism: a conceptual framework. In: Proceedings of The 18th International Conference on Electronic Business, ICEB, Guilin, China, 2–6 December, pp. 124–133 (2018)

Valdes-Faura, M.: RPA and BPM, it's all about the process! (2018). https://www.bonitasoft.com/news/rpa_bpm_all_about_process. Accessed 21 Sep 2021

Van Belkum, S., Brun, N., Cleve, S., McGovern, P., Lumpkin, M., Schaeffer, P.E., Pauli, T., Trethowan, J., Netzer, T.: Artificial intelligence in clinical development and regulatory affairs – preparing for the future. Regul. Rapp. 15(10), 17–21 (2018)

Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a literature review. MIS Q. 26(2), 13–23 (2002)

Wilcock, L., Lacity, M., Craig, A.: The IT function and robotic process automation. In: LSE Research Online Documents on Economics from London School of Economics and Political Science, LSE Library (2015). http://eprints.lse.ac.uk/64519/ Accessed 21 Sep 2021

Wilcocks, L., Lacity, M., Craig, A.: Robotic process automation: strategic transformation lever for global business services? J. Inf. Technol. Teach. Cases 7(1), 17–28 (2017)

Zhang, N., Liu, B.: The key factors affecting RPA-business alignment. In: Proceedings of the 3rd International Conference on Crowd Science and Engineering, p. 10. ACM (2018)