

# **DATA MANAGEMENT PROJECT PROPOSAL**

By Patrick Garbart

# TABLE OF CONTENTS

EXECUTIVE SUMMARY .....	1
INFORMATION PAPER	
I.    Introduction.....	2
II.   Defining Your Data.....	3
III.  Origin of the Data .....	4
IV.   Data Quality .....	5
V.    Using the Data.....	5
VI.   Monitor the Data .....	6
VII.  Business Process Map.....	6
VIII. Development of Standards.....	7
IX.   Monitor Performance and Goals .....	8
X.    Development of Standard Operating Procedures (SOPs) .....	10
XI.   Quality Assurance (QA) Committees and Reviews.....	11
XII.  Follow-Up Briefing for Project.....	14
XIII. Pre-Decisional Involvement (PDI) with the Union .....	14
XIV.  Strategic Timeline for Data Management Branch .....	15
APPENDIX 1 – Pre-Brief Template for Data Management Project .....	18
APPENDIX 2A – Supplement for Using Tableau.....	19
APPENDIX 2B – Suggested Steps to Get Knowledge of Tableau .....	21
APPENDIX 3 – Sample Process Map .....	23
APPENDIX 4 – Sample SOP .....	24
APPENDIX 5 – Sample Composition of QA Committee .....	26
APPENDIX 6 – Sample QA Committee Charter .....	27
APPENDIX 7 – Sample Checklist for QA Committee Activities .....	29
APPENDIX 8 – Follow-Up Briefing Checklist for Project Proposal.....	30
APPENDIX 9 – Sample Position Descriptions .....	31
APPENDIX 10 – Sample Course Syllabus - Data Management and Data Quality.....	40
APPENDIX 11 – Footnotes .....	42

## **EXECUTIVE SUMMARY**

The purpose of this project proposal is to 1) inform senior management of data management challenges and 2) get approval to plan and implement the data management project proposal. A table of contents is provided to summarize all parts of the proposal. The proposal begins with an information paper, covering the importance of data management. This includes important concepts related to data quality (validity). It's followed by a strategic plan to create a Data Management Section/Directorate. Your consideration for this proposal is greatly appreciated.

# INFORMATION PAPER

## I. INTRODUCTION

In the modern age, federal organizations are tasked with analyzing large amounts of data. The US Census Bureau reviews information from various populations in the United States. The Department of Labor analyzes jobless claims from thousands of records at various cities across the nation. During disaster recovery efforts, the Federal Emergency Management Agency (FEMA) reviews large volumes of data to reconcile expenses for emergency operations. The Department of Veterans Affairs (VA) manages thousands of patient records in 23 regions across the country. Many federal organizations analyze thousands of rows of Excel spreadsheets for fleet vehicle management, specifically for mileage reporting, maintenance costs, and delivery schedules. Each federal organization is required to monitor and reconcile expenditures and budgets, which can easily add up to thousands of financial transactions. These are a few examples of challenging data management projects. It is critical to adopt a defined, professional approach for these data management projects. The sections throughout the information paper represent a list of suggested areas that federal organizations (and the private sector) should review for their data management initiatives.

It can be challenging relying on computer systems and staff to ensure data is reviewed and reported accurately, to include trend analysis and forecasting. Check-and-balance systems are necessary to ensure data quality is not compromised. Without the presence of check-and-balance systems, specifically written protocols, process maps, and reviews from an independent third party, an organization risks having inaccurate data.

Different numbers can be reported by two individuals for the same reporting period. Disagreements can arise on how to accurately report data. Senior managers can become confused on seeing relationships between data sets. These are signs that data quality is compromised.

This is an *introductory* proposal. More specific information can be presented (and plans created) based on the goals of management. The sections in the information paper outline a suggested plan to review and improve data management efforts, with the goal of increasing data quality (validity). Please contact me if you have any questions about this proposal, or if you would like assistance with planning for data management initiatives. I can help organizations create positions/sections for analytics/quality assurance teams; develop materials for data management courses; initiate a QA committee program; and assist with strategic planning for organizational goals.

*Throughout this paper, some comments are annotated with a number in superscript. Please refer to **Appendix 11(Footnotes)** for specific information about these comments.*

## II. DEFINING YOUR DATA

Organizations can be confused on what type of data is necessary for creating various reports. Defining your data is a critical step in any data management project. There are some questions that need to be asked when attempting to define the data.

**1. Why does an organization need data?** This question may be trivial, but needs to be asked because a senior manager might ask this question, especially if many labor hours are being spent on data collection efforts.

**2. Why do we want to analyze the data?** Write out the answer. It helps clarify ideas and strategies. If an organization can't answer this question then having a need for the data will be hard to prove.

**3. What questions do we want the *data* to answer?** More importantly, what types of questions will senior managers have? What problems are most important to stakeholders? Do you want the data to summarize several numbers, such as a total amount of orders placed for a year? Do you want your data to highlight a trend that occurs over the course of several months?

**4. Who is the target audience?** Is it senior managers? Internal customers? External customers? Co-workers? Contractors and vendors? Each of these groups will have a different need for the data. The presenter of the data must be ready to address the concerns of all stakeholders, and more specifically, *how* the data affects their workload and operations.

**5. Do you want to present both macro and micro-level information?** Macro-level analysis gives an analyst an overall view of a large population, or large number of records. The following sets of data are examples of macro-level analysis:

- one year of data
- all training records
- all employee records

Macro-level data analysis can put a huge demand on the IT infrastructure and software applications. Depending on the processing time of the computer/server, macro-level analysis takes longer to generate, compared with working on a small data set. Before proceeding with macro-level analysis, one should guarantee that the benefit of performing macro-level analysis will be greater than the cost of analyzing the data (e.g., labor hours, opportunity costs involved with delaying other projects, level of involvement from other personnel and sections, etc.). It is especially important to understand the data (what it says or the arguments it will support). Otherwise, a macro-level analysis project can be costly. Very few employees will have access to macro-level data so it becomes critical that specific boundaries be defined for data that is requested and used.

Micro-level analysis is a smaller scale of data analysis. The data is *usually* readily available, or at least involves a minimum amount of research time. These are examples of micro-level data analysis:

- How many personnel have completed Privacy Act training for 2017?
- How many personnel actions have been processed within the last month?

Micro-level analysis is characterized by 1) minimal response time to produce the results and 2) using data mining processes that are clearly understood and have been used by several personnel. Micro-level analysis is a good way to test a research model or a software application. For example, if an organization purchases a new training database, it may be beneficial to use the database to perform an analysis of a small sample (10 employees) and test the capabilities, and accuracy, before investing the time to perform *macro*-level analysis.

### III. ORIGIN OF THE DATA

If the source of the data is unknown then it will be hard to get. Which process, system, or personnel have access to the data that is needed? Sometimes understanding the business process of how the data is created can help identify who has access to that data. Are there any special stipulations with getting the data? Is special approval needed for access? Will this approval be a one-time request, or will it involve permanent access to a database? Do you have to be a certain grade level, or senior manager, to access the data? Is the data considered sensitive, which requires the organization to filter information (redact). Are there any processes in place for filtering that data, or will it be a major task that could involve several meetings with IT managers? One of the most critical pieces of a data project is to have official approval from senior managers. When approval is given, it makes it that much easier to have groups of personnel work together to compile the data.

Making sure that the data is formatted correctly is a critical step that is often skipped in new projects, *before* starting a project *and briefing* management officials. What format is the data? If the data is coming from a centralized database does it need to be converted to an Excel spreadsheet? Does it need to be filtered, modified, and then distributed? If the data cannot be reasonably downloaded from an application then it will be difficult to complete the project.

There are two important questions that should be asked before a commitment is reached to report the data.

- 1) Is there a general understanding that the data source can be used to provide the information for the project?
- 2) Can the data be transferred to a compatible source (Excel)?

An introductory briefing is sometimes necessary for senior management to get an idea of what type of data is being compiled *before* too many labor hours are committed. There are several topics a project manager should cover when conducting a pre-brief for a data management project:

- time commitment
- occupational series of personnel needed

- number of personnel needed
- organizations used
- type of data that will be presented
- problems that the data will help identify and resolve

Please refer to Appendix 1 for a template that can be used for a data management project pre-brief.

## IV. DATA QUALITY

Data quality is arguably the most important component in any data management project. Data quality can refer to how *consistent* the data is on a long-term basis. It can also refer to how *relevant* the data is, to answer a question, or to monitor a performance metric.

Data quality can be a sensitive interpersonal issue. Not trusting the person in charge of the data can be a first impression when having a third party<sup>1</sup> check the same data. But without the presence of check-and-balance systems, specifically having third parties conduct random checks of the data, an organization adopts the risk of having poor data quality. The questions below can be used to ensure a project has adequate controls in place for data quality.

- a. What check-and-balance systems are in place to ensure accurate data?
- b. How often are the check-and-balance systems used?
- c. Can the data be verified easily, and without special coordination?
- d. Are the check-and-balance systems recognized by others, and easily understood through written Standard Operating Procedures (SOPs)?
- e. Is there a governing body for accreditation<sup>2</sup>? If so, do the accreditation standards support the organization's protocols for data quality? Have you received approval by the governing body for the check-and-balance systems, either through an official request, or an accreditation review?

## V. USING THE DATA

At this point, the assumption is that the organization has received approval for getting the data. All necessary departments (e.g., IT) have been involved and informed of future data requests. Initial check-and-balance systems have been put in place to test data quality. The next step involves studying the data and how software applications interact with the data (on a larger scale).

Which software application will be used to analyze and package the data (e.g., Tableau)? Are personnel familiar with these applications? If not, is there a plan in place, both for training and getting a robust understanding of the software capabilities? Does the software require periodic updates? If so, it could affect the timeliness of presenting data to those users who do not have access to the most recent edition.

Is the platform/application new? Are there learning curves with using the platform/application? Personnel involved with the data management project should test a data sample. The results should be verified with a third party. Please refer to Appendix 2A and 2B for information about using Tableau.

## **VI. MONITOR THE DATA**

Periodic monitoring of a completed process is a step often skipped by organizations. All data management processes should be monitored with a periodic audit. Audit processes should not be casual and should include the following:

- use of a third party to conduct audits
- audit approval process by both management and an accreditation organization (if applicable)
- SOP on the audit process
- periodic briefings for senior management

Data quality can be affected by a number of things. Once an organization has set up a successful data management process, it may encounter errors due to software updates or learning curves of new personnel using the process. These things can certainly increase the risk of poor data quality. An organization should take steps to make sure all software updates are timely and all personnel receive initial and periodic training on software systems. This includes summarizing audit processes for senior management.

## **VII. BUSINESS PROCESS MAP**

A chart, whether it be a flow chart or process map, should show all major steps in a data management process, commonly referred to as workflow. This includes critical decisions and internal controls used. See Appendix 3 for a sample process map.

There are various charts that can be used to demonstrate workflow. The important thing is not to restrict the organization to a certain format, but use a visual that will spell out all major steps that can easily be understood by a general audience.



The following list of steps is designed to strengthen an organization's process map:

1. Create an initial process map.
2. Meet with Subject Matter Experts (SMEs) to confirm process steps and internal controls.
3. Modify the process map if necessary, based on feedback from SMEs.
4. Have a third party go through the process, from start to end while taking notes.
5. Have the third party turn in notes and brief out their interpretation of the steps.
6. Based on the feedback above, identify any gaps.
7. Discuss gaps with SMEs and create, or adjust, internal controls.
8. Adjust the process map.
9. Reach a consensus on the final draft.
10. Execute the new process map.
11. Monitor the new process map, to account for updates in workflow, technological improvements, and other things that may change the process.

## VIII. DEVELOPMENT OF STANDARDS

Once data quality processes have been mapped and confirmed, an organization should proceed with developing *standards*. A standard is the executed delivery of a product or service, with an expected outcome. Consumers who purchase items from fast-food franchises expect the same product quality and delivery, regardless of the restaurant location. This can be understood as a standard. Within the personnel arena, getting approval for and processing personnel actions within 72 hours can be regarded as a generally recognized standard of service within HR departments at various organizations. A standard can simply be a generally recognized way of doing business within an organization that demonstrates a consistent delivery of a product or service.

An absence of standards can confuse stakeholders. It can also confuse employees within the work unit. For example, let's say there are two employees who process the same type of personnel actions. The first employee usually completes actions within two days. The other employee takes four days to complete actions. The recipients of the actions (customers) get confused on why it takes twice as long to get actions processed from the second employee. In this situation, all stakeholders can be confused on the appropriate expected delivery of a service—should it be two days or four days? The development and execution of standards will help ensure business processes are followed correctly, and within the correct time period.

A process map defines *what steps* will be followed in a process. A standard will define *how* the process will be followed. With regards to data management standards, there are two important components that should be considered.

- 1) **Time** - How frequently should a data report be pulled? When should this report be sent to other individuals? It's important to indicate specific expectations for time periods.
- 2) **Who** - Who will receive the reports? Which employees are in the chain of custody (how the reports get transferred from one individual to the next)?

The two components above are reflected in the sample standard below:

*The Data Technician (DT) will run a report of the data the first working day of every month. The DT will send the Management Analyst (MA) the report the same day. The MA will have two working days to 1) conduct data checks, 2) identify any anomalies, and 3) produce analytics charts. The MA will then forward the charts to the Program Manager (PM). The PM will have three working days to review the reports, conduct random data quality checks, summarize data, and provide feedback to senior management.*

The most important thing about a standard is that it must be an expectation of senior management, and advertised as such. If senior management does not support, expect, and market the standards then the standards have little value.

## IX. MONITOR PERFORMANCE AND GOALS

After data management processes and standards have been developed, it is time to plan for expected outcomes. There are various ways to track performance. Targets, Key Performance Indicators (KPIs), performance metrics, industry benchmarks, and Balanced Scorecards are all tools that can be used to monitor performance and goals. Whatever tool is used it should be easy to understand and use for data management processes. Below, are some basic descriptions of each tool.

- a) **Targets.** Targets are usually goals represented by a number. For example, an organization may want to reduce expenditures by 5%. During the year, the organization tracks expenditures and takes precautions to cut costs. At the end of the year, the organization sees if it has met the target. Targets usually do not require lengthy briefings or explanations. Targets can provide a simple way to monitor performance when there are a lack of resources available to monitor that performance.
- b) **Balanced Scorecard.** A Balanced Scorecard is simply a list of financial and non-financial measures that an organization uses to compare against targets. These measures are usually tracked and discussed on a monthly basis. It makes sense to use Balanced Scorecard reporting when targets are easily understood and measured, with minimal cost of monitoring these measures.
- c) **Standards.** Standards are generally recognized ways of doing business, or delivering a product or service with specific expectations. In the case of a data management initiative, performing monthly random data checks (for accuracy) could be understood as a standard of service. Standards are more than just a tool. They are agreed upon expectations. A good standard is one that has been tested, acknowledged, and part of the organizational culture.
- d) **Key Performance Indicators (KPIs)** can be used to measure progress towards strategic goals, as in the case of managing a Lean Six Sigma or Lean Kaizen project. These measures are usually tied to Balance Scorecard management, and can show the health of

an organization. For example, let's say a large organization is trying to evaluate the cost of using laptops for staff. In the past, laptops have been replaced every three years, due to required software upgrades. The organization would like to decrease the workload of IT staff and purchase laptops that can be used for five years. So the organization creates an IT Equipment Lifecycle KPI. All future laptop acquisitions are measured with the KPI.

In general, KPIs should be:

1. Specific
2. Measurable
3. Achievable
4. Relevant
5. Time-phased (shown for a relevant period)

- e) **Performance Metrics.** These are periodic snapshots of goals that have been met or not met. In many cases, metrics will not be adopted unless there is an understanding, and commitment, that more than one line of business (LOB) will be monitored. The more people who are involved with a data management initiative, especially at the national level, the more it makes sense to create performance metrics. Also, metrics are more in line with defined and frequent problems. Performance metrics are more centric to stakeholder needs. Key components of a metric include productivity and quality of service. Use of performance metrics make sense when improvements are needed with critical paths in workflow processes, or when an organization is trying to resolve a complex problem. An organization should understand that the use of performance metrics can be labor-intensive, as it involves thorough and lengthy reviews of large data sets.
- f) **Benchmarks.** Benchmarks can be used to grade an organization's demonstrated performance to industry standards. A benchmark can be a simple number. In the case of hospital operations, benchmarking can occur when a hospital attempts to compare its patient appointment wait times to those in similar sized hospitals across the country. A benchmark can also be complex, as in the case of measuring the quality and cost of a process used in a construction project. There are *usually* three pre-requisites for using benchmarks:
1. Best Practices have been established and recognized in an industry or trade.
  2. Standards have been adopted based on the Best Practices.
  3. An organization manages performance metrics, and often compares them to the standards and Best Practices.
- g) **Relative Value Units (RVUs).** RVUs are a way to measure the value of services provided when revenue is solely received from federal institutions, or when the value of employee services cannot be *directly* tied to revenues. Its use requires a long-term commitment because Independent Government Estimates (IGEs), salary surveys, and organizational research must be conducted to derive accurate monetary values of services provided, compared with employee salaries. Its use is controversial because some employees don't like having a monetary value tied directly to their performance.

However, RVUs are a good fit if there is an organizational directive to manage profit and loss statements.

## **X. DEVELOPMENT OF STANDARD OPERATING PROCEDURES (SOPs)**

Let's assume an organization has created and verified process maps for data quality management. Standards have been developed, and approved, for each of the processes. Tools are now being used to monitor performance and goals. The next step is to create SOPs. An SOP is a written step-by-step procedure that ensures a business process is followed correctly. A well-written SOP will allow someone who is unfamiliar with the process to read the SOP and perform all steps correctly, without interruption. To ensure stakeholders have a clear understanding of data management processes, it becomes necessary to create SOPs that define all unique steps, to include those involved with data quality (validity). Other important topics to include are the definition of each internal control, and the individual steps involved to make sure each control is followed correctly. Please reference Appendix 4 for a sample SOP.

The SOP is an important component of any accreditation inspection. During these types of inspections auditors can ask to see copies of SOPs. Some might ask other employees if they have knowledge of the SOPs, and where to locate them. If there are SOPs in place, it helps ensure staff are aware of following processes correctly, especially during an inspection or review.

Below, is a list of steps that can be used to develop an SOP for data management processes:

### **Step 1**

Identify the process map. If there is none, then one should be created. A good process map will list all major steps (in a sequence) and include internal controls.

### **Step 2**

Begin crafting your SOP. Start with listing each item in the process map. This will be your outline, and also major topics within the SOP.

### **Step 3**

Write out how you perform each step, in detail. It is important to remember that several personnel may be unfamiliar with your data management processes. So all basic steps should be included to make sure the procedures are being followed correctly. Let's say we are including a step about reviewing Excel spreadsheets for accuracy. An organization has designed a check-and-balance system for checking the accuracy of the spreadsheets and will include this information in an SOP. Here are two ways this can be written.

#### **Description 1**

*The Data Management Section will receive the spreadsheets. The Program Analyst (PA) will review for accuracy by making random checks, and then contact the Data Technician (DT) if there are any problems.*

## Description 2

*The Program Analyst (PA) will receive the spreadsheets from the Data Technician (DT). The PA will review the first five and the last five transactions of the month, which are included in the spreadsheet. Once these 10 transactions have been identified, each row (in the spreadsheet) of these transactions will be reviewed. Specifically, each cell (data element) in these rows will be compared with the information listed on page two of the profile screen, for each customer, in the Mark V software application. If there are any errors, then contact the DT to reconcile, then check the data again. If there are further errors, report them to the Program Manager. If there are no errors, then continue to the next step.*

Description 2 is much more detailed than description 1. Description 1 gives the wrong impression that it's a simple step. Also, there is a large margin of error, because the reader of the SOP may not be performing the step correctly when all the information in description 2 is missing.

## Step 4

Have a volunteer use the SOP draft, follow it, and try to complete all the data management processes. If there is confusion, then the steps must be re-written. Information will need to be added, or clarified. The SOP needs to have clear and simple language.

The last important thing to remember about SOPs is that they should be reviewed and tested every three years. Programs, requirements, and software applications change on a regular basis. In addition, it may be a good idea to have a Quality Assurance (QA) committee review SOPs for accuracy, and ease of use. The QA committee concept is discussed in detail below.

# XI. QUALITY ASSURANCE COMMITTEES AND REVIEWS

## BACKGROUND

To ensure procedures, controls, and standards of service are being followed correctly, and objectively, it is necessary to create a QA committee. At times, organizations encounter issues with data quality. Examples include data not coming from the correct source; new personnel handling data sets who are not familiar with data collection processes; and personnel not paying attention to detail when preparing data reports. All these things are data quality problems and lead to increases in labor costs (to correct deficiencies) and loss of credibility with a person, or an organization.

A much more complex problem can present itself when there are two employees in the same section presenting the same data set and the results are different. In this situation, it is difficult for the manager to believe which employee is correct. These types of situations create a need to have an independent third party (does not work for, or with, the section being reviewed) to review the data sets in an objective manner. The creation of a QA committee can fulfill the role of a third party used to review data sets and processes, in a fair and objective manner.

A QA committee can be used to identify areas of concern before they become major problems. Various public and private organizations have created working groups and committees that

discuss, monitor, and plan for the resolution of data quality issues<sup>3</sup>. For example, QA committees are used in the Department of Veterans Affairs (VA) to monitor KPIs and metrics. At the same time, the committee also identifies anomalies with data sets and initiates discussions, and studies, to resolve problem areas. In the case of the VA, the QA committee is used as an independent party to ensure procedures are being followed correctly; data sets are carefully reviewed, organized, and reported; and standards of service are being met. QA committees can be a helpful addition to any data management initiative by helping CBP organizations with quality control for its data management processes.

### **COMPOSITION OF QA COMMITTEE**

The suggested composition of a QA committee will be five volunteers serving on a detail, each with different occupational series and grade levels. A senior manager, GS-15, will serve as a reviewer, to review all final QA packages, and to communicate with other senior managers about findings and trends. Each individual will serve on a one year detail. This detail will not be full-time, but rather a commitment for the individual to use 20% of duty time as a committee member. Please refer to Appendix 5 for a sample composition of a QA committee. A QA analyst (permanent position referenced in Appendix 9) will serve as the facilitator, trainer, and librarian for all QA committee projects.

### **APPLICATION PROCESS**

The application process for this detail will include a signed statement from the supervisors of volunteers acknowledging that a time commitment will be needed. The detail will be announced and open for a one month period. Volunteers may be selected from different DHS components in an effort to ensure a diverse background in occupational series and levels of responsibility. A panel board of senior managers will be used to select volunteers for the QA committee. This is an effort to ensure only employees who demonstrate the highest professionalism (with both bearing and work ethic) are selected.

### **CHARTER OF QA COMMITTEE**

During the initial application and selection process, a QA committee charter will be created. It will define roles, responsibilities, and expectations. This charter will be a formal document, authorized and signed by two senior managers, to include one Assistant Commissioner. Key components of the charter will include the following:

- understanding that QA committee members will have access to records for QA reviews
- periodic training requirements
- understanding that all reviews will consist of work products that can be easily transferred and analyzed, either from PDF documents, or access to software applications
- authority to carry out reviews
- No QA committee member will be allowed to work on a project from their own section/directorate.

A sample charter can be found in Appendix 6.

## **SEQUENCE OF EVENTS FOR QA COMMITTEES**

The QA committee will be activated after members have been selected and have received an initial orientation. Senior management will be informed that the QA committee is ready to conduct reviews. The following sequence of events will be implemented:

1. CBP organizations, or DHS components, will be solicited to provide areas in need of review-those areas that have challenges with data quality.
2. Senior management will review the list of data quality challenges from various organizations and will select one for a QA committee review.
3. The QA committee will contact the organization and make arrangements to review a process.
4. The QA committee members will divide the workload into the following parts:
  - a. Defining the reason for the review. The organization, or senior management, should be able to specify the problems involved with an organization's data quality and processes before the QA committee begins work. If a problem cannot be identified, then it cannot be solved.
  - b. Given the definition of the problems, a QA committee member will review all regulations, policies, guidance, standards, and SOPs, and highlight areas that are relevant to the data quality problems.
  - c. Another committee member will begin review of all data sets (e.g., Excel spreadsheets, reports, etc.), identify any anomalies that are easily recognizable, and use existing data validation check-and-balance systems to determine if they are effective.
  - d. A third member will trace the workflow process by following an SOP, flow charts, and process maps. The individual will document all steps, while looking for anomalies with process flow. This includes adequate use of internal controls. The committee member will interview all employees involved with the process to evaluate their interpretation of data management processes, and to provide an opportunity for feedback about process improvement.
  - e. A fourth member will research other public and private organizations to see if there are any recognized Best Practices in the industry. The same person will document any Best Practices and list points of contact for those practices (e.g., Department of the Interior, Stanford University, FedEx, etc.). At a later date, arrangements can be made for a few committee members to visit the organizations and receive a full briefing for those Best Practices. The same committee member will serve as the recorder for QA committee meetings.
  - f. A fifth member, a senior manager, will serve as the official reviewer for all final QA committee packages and briefings. This member will also brief senior management of findings, and recommendations to correct any deficiencies.
  - g. Documents will be submitted to the QA analyst on a periodic basis for proper filing.
  - h. All members will meet periodically to discuss findings and recommendations.

## XII. FOLLOW-UP BRIEFING FOR PROJECT

A follow-up briefing is an opportunity for senior management to understand a data quality/data management project, including the planning phases, completed milestones, and actions that have yet to be delivered. The briefing is a chance to demonstrate the success of the project. It also allows an opportunity to identify room for improvement and any activities that should be added to the project. Appendix 8 includes a list of items that should be used in any follow-up project briefing.

## XIII. PRE-DECISIONAL INVOLVEMENT (PDI) WITH UNION

PDI with the Union can help the project planning process by informing Union members early about the change. Accurate, timely, and thorough communication will help address any concerns from the Union. It must be understood, and communicated, that the primary purpose of the project is not to find ways to penalize employees, but to modernize operations, equivalent to processes found in other public and private organizations. The outline below is a recommended sequence of steps for involving the Union at major stages of any data management project.

### Step 1 – Outline the purpose of the meeting

An outline of the meeting should be given to the Union in advance, preferably one month before the scheduled meeting. By creating an outline of the meeting, it gives Union officials a chance to address any concerns *before* the meeting, *not while* the meeting is being held. This streamlines facilitation of future meetings. A good notification memorandum should include the following:

- a. **Purpose.** The purpose of the meeting should be clear and reasonable.
- b. **Background.** A short paragraph should cover the series of events that have led up to the meeting, to include challenges and problems with data quality.
- c. **Goals of the meeting.** The goals should be specific and reasonable, items that are within the authority of management. Too many goals can drag out meetings and confuse attendees. It is suggested that only a few goals be introduced during any meeting.
- d. **References.** A references section should be used to include any official documents that supports agenda items (e.g., regulations, studies, mission statements, etc.).

### Step 2 – Send the Union notification

Once the memorandum is finalized it should be delivered to the Union for review and comment, to include acknowledgment of notification. This should be immediate. However, a suggested review time of 30 days can be given so that the Union may provide feedback before the meeting.

### Step 3 – Hold the meeting

When the meeting takes place, it is suggested that a professional facilitator be present (independent of the Union and management) to start the meeting on time, present ground rules



up front, to keep the meeting on track, and make sure all concerns are addressed. A recorder should be identified to take minutes for the meeting. During the discussion, both Union and non-Union members should be allowed to participate. Concerns that cannot be resolved during the meeting should be recorded in the minutes, with an understanding that it will be addressed at a later date.

#### **Step 4 – Produce minutes**

After the meeting, the recorder will produce a draft of the meeting minutes and send to all participants. Participants will have one week to make comments, to ensure the meeting minutes are accurate. The recorder will then make edits to the minutes and sent out a final draft for approval. Meeting participants will have one week to review and approve the meeting minutes.

#### **Special note:**

If a *reasonable* amount of time is given before scheduling meetings then all parties should be expected to provide representatives as scheduled, with the understanding that a backup may be needed for primary representatives when scheduling conflicts arise. It is the responsibility of the Union to designate a backup, not have all parties wait (and delay all processes) until a representative can be found. The same expectation should apply to other attendees of the meeting.

## **XIV. STRATEGIC TIMELINE FOR DATA MANAGEMENT BRANCH**

The project proposal includes the creation of a data management branch (a directorate if demand increases). This branch is designed to educate DHS and CBP organizations about data quality, analytics, and quality assurance. The branch can be used as a basis to begin a robust quality assurance program by assigning full-time QA expertise to assist with various QA initiatives; managing the QA committee program; and teaching organizations about the fundamentals and processes of quality assurance. The branch also offers various means of data mining and analytics, such as summarizing large volumes of data in operational components; conducting trend analysis for large sets of data; creating robust cost and benefit analyses (profit and loss statements); and streamlining data collection efforts. Below is a suggested timeline to create a data management branch.

#### **Phase 1 – Month 1** (starting December 2017)

Produce an information paper and share with senior managers. The goal of the information paper is to inform senior managers of the proposal, and allow a grace period for feedback.

#### **Phase 2 – Month 3**

Brief senior management on a multi-year plan to develop a data management branch. Based on approval and feedback from senior management, edit the multi-year plan.

#### **Phase 3 – Month 5**

Create a position description (PD) for a Program Manager (PM), GS-14 (reference draft in Appendix 9) Edit the PD based on feedback from senior managers and/or HR. Receive approval

from management and HR channels to hire the PM. Define roles, responsibilities, and planning milestones of the PM. There must be workload up front to ensure the PM is gainfully employed. The PM will have the following initial duties:

- manage a strategic planning document
- yearly surveys of data management/data quality needs
- collateral duty of Document Control Manager (at least on temporary basis), which will help support Freedom of Information Act (FOIA) requests or litigation hold
- Self-Inspection Program (SIP) management processes and deadlines
- meet with IT about new challenges with data management
- create training materials on data quality management

#### **Phase 4 – Month 6**

Advertise the job announcement for the PM.

#### **Phase 5 – Month 8**

Select the PM.

#### **Phase 6 – Month 9**

The PM will create a strategic planning document for FY19 which lists specific goals, milestones, and deliverables. The PM will begin initiating quality assurance and data analytics programs. This begins with surveying senior leaders of their data management and data quality needs. The PM will create a strategic planning document listing specific roles and responsibilities for future positions, along with program milestones. It will incorporate feedback and goals from other offices, based on survey responses.

#### **Phase 7 – Month 11**

The PM will present a proposal to hire staff to support growing data management needs. This proposal will include the establishment of the following two new sections (or only two personnel), which will report to the PM:

##### **1. Quality Assurance**

A QA analyst, GS-13, will have the following responsibilities:

- audit and compliance
- coordinate QA committee events
- manage QA committee documents and deadlines
- brief managers about QA findings and reviews

##### **2. Analytics**

A data analytics/management analyst, GS-13, will have the following responsibilities:

- manage all data requests (for large metadata)
- train others how to use software programs
- brief management on dashboards

**Phase 8 – Month 12**

Advertise job vacancy announcements for the QA analyst and data analytics/management analyst positions.

**Phase 9 – Month 14**

Make selections for the new positions.

**Phase 10 – Month 15**

New staff begin work. They will be given roles, responsibilities, and a strategic plan to help modernize data management programs within the CBP and DHS.

*If the demand for branch services increases significantly, it may be necessary to create a data management directorate. Additional phases are listed below to create a directorate:*

**Phase 11 – Month 21**

Create a data management director (GS-15) PD.

**Phase 12 – Month 23**

Hire a data management director.

**Phase 13 – Month 25**

After the director has been hired, the Data Management Directorate (DMD) will stand on its own and exist as a new organization within the Office of Facilities and Asset Management (OFAM). It will be charged with providing support to all CBP offices, to include the following:

- QA reviews for multiple programs
- robust QA training
- data mining
- implementation of applications (Tableau Server) throughout several programs
- customer support for dashboard management
- conducting a year-long open house for the new directorate

**Phase 14 – Month 27**

Create a marketing and strategic plan for 2020.

**Special Note:**

The DMD will report to senior management. However, additional oversight is needed to ensure the directorate is operating within industry standards (e.g., how audits are managed). A generally recognized governing body (accreditation organization) will be used to conduct yearly audits on the activities of the DMD.

## APPENDIX 1 - PRE-BRIEF TEMPLATE FOR DATA MANAGEMENT PROJECT

**Define the problem.** Tell your audience in a clear and concise manner why the organization should pursue the data management project.

**Define the source of the data.** Be prepared to brief an example of any data sets used to help clarify the problem. What is the source of the data?

**Cover the steps used in gathering and analyzing the data.** This helps your audience understand the relationship between data sets and stakeholders.

**Who are the stakeholders?** Who are the personnel and organizations that will be involved? Who are the customers that the project intends to serve?

**What is the time and labor commitment?** Define the number of people and occupational series used to get the data. A listing of personnel and organizations who will be involved with the project would be helpful. It may be a good idea to include a dollar value (number of hours x \$50/hour).

Example:

*Five personnel are needed to compile, analyze, and brief the data. This will be a monthly requirement of two hours for each person. This amounts to 10 hours per month, or \$500 in labor costs per month (\$50/hour x 10 hours).*

**What do you need of senior management?** Be specific. Here are a few examples:

- approval from the director
- commitment from IT supervisors to find solutions
- shared resources (data sets from other organizations)

**Don't forget the obvious.** Make sure there are no grammatical errors with slides. The data should be accurate *and current*. Arrive 45 minutes prior to the presentation. The first 15 minutes should be used for the setup of the laptop. The next 15 minutes should be used to make sure the presentation can be viewed from the main viewing screen. Handouts should be passed out prior to the arrival of the attendees.

## APPENDIX 2A – SUPPLEMENT FOR USING TABLEAU

The purpose of this appendix is to cover important aspects of Tableau, as it pertains to data quality checks. Specifically, it includes a list of steps that can be used to increase data accuracy and to help identify anomalies with data sets. For those personnel who have not yet used Tableau, or are just now learning about it, please refer to Appendix 2B. Appendix 2B is a list of steps I recommend to get a robust understanding of Tableau capabilities.

### STEP 1

Before using Tableau, the user must get the correct data set. The user should be specific about the following areas:

- time period covered for the data set
- correct fields
- format (like Excel)

Make sure a folder has been set up with an easily recognizable name, and the file has been properly saved. Make a note of the name of the file and folder. With so many files and folders being used these days, it is easy to lose track of where your file is.

### STEP 2

Load the data into Tableau. In the upper left hand corner under “Connect” select the type of file you’ll be uploading (Excel, text file, etc.). Will you be using another file for indexing purposes? If not, proceed to Step 2A.

For example, one file contains ordering information. Another file contains customer names. The unique field both files have in common are the customer ID numbers. If you are referencing another file, you’ll need to make sure a proper “join” is set up. For example, with the two files referenced above, an inner join will ensure the customer ID # for one file is matched with the customer ID # for the second file. Once a “join” is initiated, it will connect both files together into one. Once the join has been created, the data will show in the lower pane. Go through it and check the following:

- names of columns
- any null values (missing information)

Null values can indicate key information is missing, like a date. It can also mean a duplicate entry was made. Take null values as a warning. It’s an opportunity to check for data validity. It is best to delay any further steps until the source of the problem is known. Otherwise, you risk data quality issues, tainting your analysis.

### STEP 2A

Select your file. You may have to click the “Update Now” button in order to see the data.

### STEP 3

Review your data. Now is the time to begin identifying other anomalies with the data. After performing step 2, you may identify additional null values. This may be related to missing information (incomplete data set), or it may indicate a duplicate value from a previous data set. For example, a missing date which creates a null value in Tableau may be related to a customer having a duplicate order for the same date. One way to check the validity of the data is to review the original file and see if these types of relationships can be identified. You may also encounter problems with numerical values. For example, a customer ID number might show up as a value (with commas). In this situation, Tableau has initially captured the value as a “measure.” You can change this by moving to the “Sheet1” tab then move the name of the measure to the dimensions pane (converting it from a measure to a dimension). Now go back to the data source and you’ll see that the value has been corrected (not showing commas). These are the types of things that can cause problems. Let’s say you’re creating a calculation in Tableau. One of your values is a dimension. The calculation treats your value as a dimension. When you convert your value to a measure it changes the calculation. In summary, it pays to take a few extra minutes to review each column to *validate* each value as a measure or dimension.

### STEP 4

Review the names of the columns. If the columns have names such as “Tpv98” it can get confusing, especially if multiple columns have similar names. This can create confusion, which adds to data validation problems. It’s a good practice to rename these fields with titles that are understood. You can either go back to the original file, then rename columns, or you can do it from the Data Source tab. If you do it in Tableau, simply right click the column and select “rename.” Type in a name that makes sense. For example, “Program Code” may be a better name than “code.” By renaming the columns with clear titles you’ll avoid confusion-a symptom of data quality problems.

## APPENDIX 2B – SUGGESTED STEPS TO GET KNOWLEDGE OF TABLEAU

This guide is for personnel who are just now starting to use Tableau. It lists specific steps that should be helpful for gaining a robust understanding of Tableau capabilities.

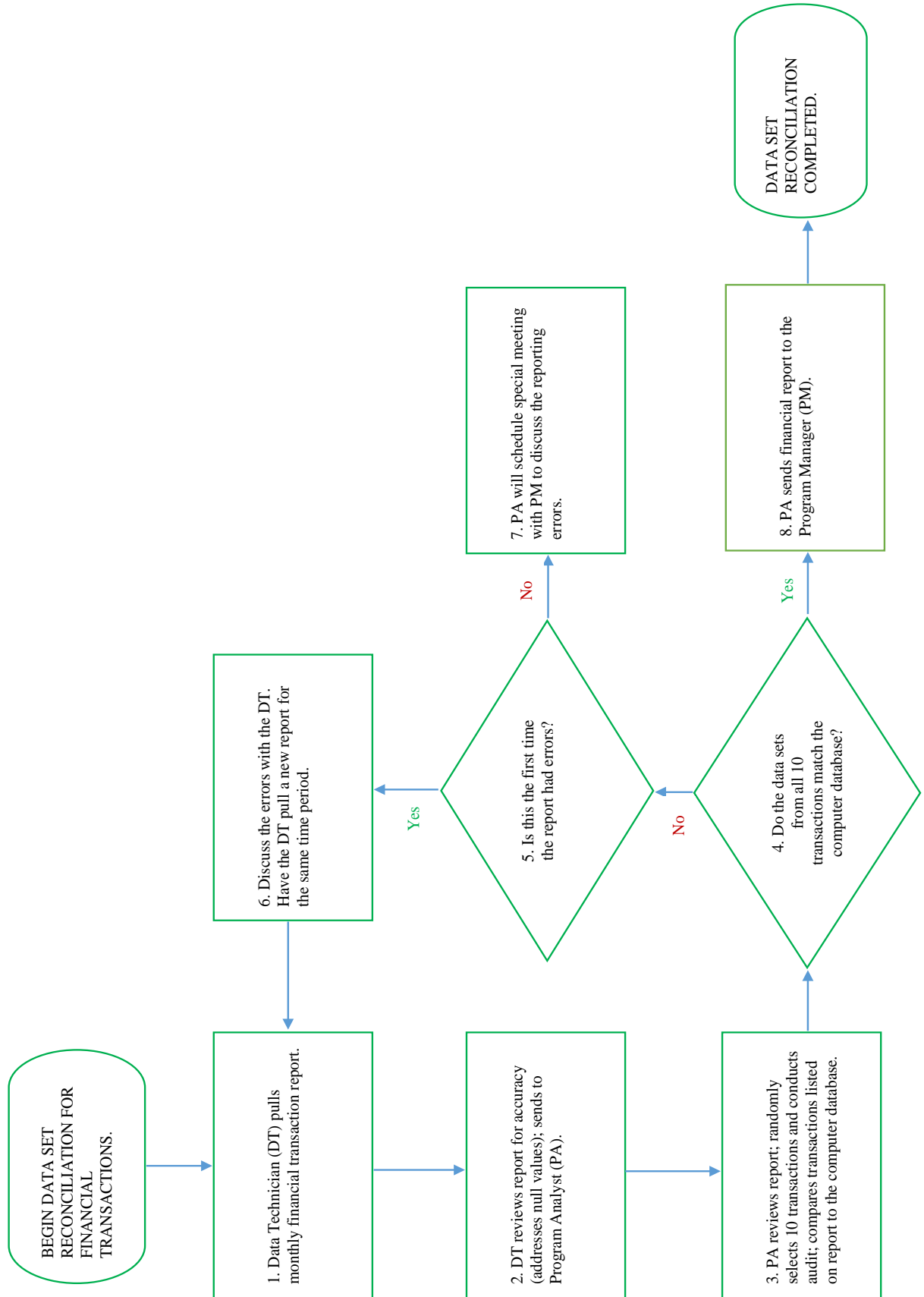
1. Get access to Tableau. The Tableau program may be a little confusing to IT technicians. The easiest thing to do is submit an IT work order and have the main Tableau website handy. The Tableau Reader version is free. The full version must be purchased, but is a good investment if the long-term goal is to use Tableau for data management, analytics, or Quality Assurance (QA) processes.
2. Once loaded, activate the Tableau application. Become familiar with the Tableau menus. There are basically four major parts (at the bottom) of the application:
  - a. **Data Source.** This tab is where users will first add data to the Tableau file. Before jumping into the graphics portion of Tableau (worksheets), the user should review the Data Source tab carefully to check for anomalies, and the data within it. This tab represents the first area of opportunity to identify any data validity issues. Please reference Appendix 2A, steps 2 and 3, for specific information.
  - b. **Worksheet.** This is the primary window for viewing and manipulating the data; filtering data; creating graphics; building relationships between data sets; and creating calculations. This tab represents a secondary means of identifying data quality anomalies. For example, if a calculation is created to view a specific data set, the graphics portion will help depict disparities.
  - c. **Dashboard.** This tab is used to summarize various graphics from worksheets.
  - d. **Story.** A story *usually* displays a relationship between worksheets, dashboards, and data sets. This tab can be used to summarize key concepts about the data.
3. Take the first five free video mini-courses on the Tableau website. They should each be 2-10 minutes long. Take notes. Try the demonstrated features on your Tableau program. If you only have Tableau Reader, you'll be limited to certain tasks. There are various types of free files that can be viewed. Take the time to understand them.
4. Take the next five free video mini-courses and repeat the process above.
5. Visit the blog site and become familiar with how to submit an inquiry. You can review the Question and Answer segment of the blog to get an understanding of the types of problems Tableau users have.
6. Review the Eventbrite.com website. There are several free webinars available. There may be a few free Tableau courses that can be taken. There are similar sites on the internet in which you can view free Tableau training videos.
7. If you live near a major city there should be a Tableau working group. Members of these groups are Tableau users from several organizations, both in the public and private sectors. The meetings usually take place once a quarter and are a good way to network

with Tableau users, face-to-face. The more you talk the Tableau language, the faster you'll learn.

8. If your organization can afford it, sign up for an introductory resident two-day Tableau training course. I suggest you take a course that is taught by an instructor who works for the Tableau organization (to get more detailed information on solutions to problems, and trends in the industry).
9. Before the course, write down a list of questions you have about Tableau and take it to class. In the meantime, take the next five free video mini-courses on the Tableau website.
10. Show up to the Tableau class and take good notes, and ask a lot of questions. You should get a practice manual and thumb drive to take back to work. During the class, network with other users and learn how they use Tableau in their organizations.
11. Take the practice manual and thumb drive back to the office and review the concepts learned in class. Try creating the Tableau view for the "Solution," which are specific step-by-step instructions at the back of the book. When you are comfortable, move on to the "directions," which have less step-by-step instructions. Last, see if you can create the Tableau view with just basic parameters, located in the "specification" of each practice.
12. If you have access to the full version of Tableau this is when you start to develop your own original worksheets, using information specific to your organization. If you get stuck on concepts or steps you can:
  - a. refer back to your practice manual
  - b. create a blog entry (someone will usually respond within a day)
  - c. submit a work order to the Tableau website
  - d. ask another Tableau user for help
13. Take more free video mini-courses.
14. If your organization has money to send you to another Tableau course then go ahead and sign up for the next course in the series. Before you attend the next course, make sure you understand the basic Tableau concepts, and have practiced several times using the practice manual.
15. Continue working on small projects that relate directly to your organization.



## APPENDIX 3 – SAMPLE PROCESS MAP



## APPENDIX 4 – SAMPLE STANDARD OPERATING PROCEDURE (SOP)

### 1.1 Purpose

The purpose of this SOP is to provide step-by-step instructions for conducting periodic audits of financial reports. Specifically, each monthly financial transactions report will undergo an audit. This audit will test data quality by comparing transactions in the report to information located in the Mark V database. These audits are necessary to ensure senior management receives accurate data for briefings that are related to financial transactions.

### 1.2 Scope

This SOP applies to the Data Management Directorate (DMD). The management analyst (MA) and any other designated backup, will use these procedures to conduct audits on a monthly basis. It is understood that when data discrepancies exist, the MA will meet with the data technician (DT), the author of the original report, and work on a plan to reconcile the reporting errors. This is necessary before any reports are turned in to senior management. If the same report (reconciled) has further discrepancies, the MA will meet with the DMD director to discuss further action.

### 1.3 Roles and Responsibilities

#### 1.3.1 **DT**

The DT will provide monthly financial transactions reports to the MA. A new report will be provided to the MA the first working day of the month. It will include all financial transactions for the previous month. The DT will have a designated backup to ensure the monthly reports continue without interruption.

#### 1.3.2 **MA**

The MA will receive the monthly report from the DT on the first working day of the every month. The MA will conduct an audit using the process procedures outlined below. The MA will be given two working days to conduct the audit before turning in the financial transactions report to the DMD director. If the MA finds errors on the initial report he/she will go back to the DT and ask for an adjusted report. If the second report still has errors the MA will meet with the DMD director and await further instructions. It is the responsibility of the MA to have a designated backup, who will use the procedures below if cross coverage becomes necessary.

#### 1.3.3 **DMD director**

The DMD director will review the content of the reports and prepare the information for senior management briefings. The DMD director will meet with the MA to discuss anomalies and determine what actions should be pursued.

## 2. Process

- 2.1 The Mark V database has the capability of producing reports for several financial transactions. The user (in this case the DT) must select a series of dates to include for the financial transactions. The user can then view or print a report for the date range. The initial report will contain 20 fields. The DT will be responsible for copying and pasting the first 10 fields into an Excel spreadsheet. This spreadsheet will be reviewed for accuracy before turning in to the MA.
- 2.2 On the first working day of the month, the DT will prepare the report that contains financial transactions from the previous month. The DT will turn in this report to the MA the same day.
- 2.3 The MA will audit the report within two working days. The MA will review the first five and last five financial transactions in the report. The MA will login to the Mark V database and make sure each of the fields in the ten transactions match the data in the Mark V database. There are three different scenarios below.
- A. The information matches. In this case, forward the report to the DMD director.
  - B. The information does not match. Is this the first time there were errors? If yes, then go back to the DT and reconcile the errors. If not, then proceed to C below.
  - C. If the report had errors more than one time then the MA will meet with the DMD director and await further instructions.
- 2.4 Once the DMD director receives the report he/she will review it and prepare for the senior management briefing.

## 3. Approval

\_\_\_\_\_ *signature* \_\_\_\_\_  
Name of senior manager  
Title of senior manager

\_\_\_\_\_  
Date

## APPENDIX 5 – SAMPLE COMPOSITION OF QA COMMITTEE

### QA Analyst, GS-13 (permanent)

- lead QA committee facilitator
- trains committee members
- librarian for all activities
- performs random quality control checks

### Program Manager, GS-15

- reviews all final drafts
- signs final documents
- summarizes results to senior management
- backup facilitator

### Engineer, GS-14

- regulation/policy review
- backup recorder
- backup Best Practices research

### Management and Program Analyst, GS-13

- data set sampling/reporting
- backup SOP/process map review

### IT Specialist, GS-13

- research Best Practices
- recorder
- backup regulation/policy review

### Management and Program Analyst, GS-12

- SOP/process map review
- backup data set sampling/reporting

## APPENDIX 6 – SAMPLE QA COMMITTEE CHARTER

**Purpose.** The purpose of the QA committee is to review processes related to data collection and quality. This is an effort to strengthen data management processes for various offices within the Customs and Border Protection (CBP) Agency. The committee will act as the CBP's official third party review team for data management processes. The committee could extend its services to other components within the Department of Homeland Security (DHS).

**Objectives.**

1. Pool talents within the QA committee to review Standard Operating Procedures (SOPs), process maps, and data sets with the goal of identifying and documenting weaknesses in data management systems and processes.
2. Implement robust reviews of data management processes.
3. Submit recommendations to improve data collection and reporting efforts.
4. Collaborate with both the public and private industries to identify, and adopt, Best Practices for data management in various programs.

**Responsibilities.** QA committee members will receive initial training from the QA analyst. Each member will have one primary and one backup role. During the detail, committee members will review records for data management processes from various offices within the CBP; audit relationships between data sets; study process flow; document challenges; and summarize recommendations to improve data management. Each committee group will have access to records for QA reviews, periodic training requirements, and an understanding that all work products can be easily transferred and analyzed, from email, or user access to applications.

**Membership Roles.** The list below represents the roles that each QA committee member will have. Committee members will be of different grade levels and occupational series.

***QA analyst***

Serves as lead QA committee facilitator; trains committee members; serves as librarian for all QA activities; and performs random quality control checks.

***Senior Manager (volunteer #1)***

Reviews all final drafts; signs final documents; serves as backup facilitator; and summarizes results to senior managers.

***Volunteer #2***

Regulation/policy review; backup recorder; and backup Best Practices research.

***Volunteer #3***

Data set sampling/reporting and backup SOP/process map review.

***Volunteer #4***

Research Best Practices; recorder; and backup regulation/policy review.

***Volunteer #5***

SOP/process map review and backup data set sampling/reporting.

**Meetings.** Meetings will be held periodically among QA committee members to discuss work products and milestones.

**Terms.** Each member will serve on a one-year term, while also contributing 20% of duty time towards official QA committee activities.

**Proposed:**

**Approved:**

\_\_\_\_\_  
SIGNATURE  
Name/Title of Senior Manager (GS-15)

\_\_\_\_\_  
SIGNATURE  
Name/Title of Assistant Commissioner

## APPENDIX 7 – SAMPLE CHECKLIST FOR QA COMMITTEE ACTIVITIES

The purpose of this appendix is to provide an outline of all major steps with each QA committee review. The following is a list of suggested steps.

1. Committee members have been selected for a detail, and their roles have been defined.
2. Each committee member has completed the initial training provided by the QA analyst.
3. The program manager (PM) who will serve on the QA committee will meet with senior leadership to determine which organization will undergo a QA review.
4. All committee members meet to discuss the details of the assignment, to include work processes being reviewed, and individual assignments.
5. The recorder sends the committee a draft of the meeting minutes for approval.
6. Committee members shall conduct the following activities:
  - \_\_\_\_regulation/policy review
  - \_\_\_\_data set sampling
  - \_\_\_\_SOP/process map review
  - \_\_\_\_meetings with process owners
  - \_\_\_\_Best Practices research
7. The committee meets to discuss initial findings. Individual assignments are modified based on findings.
8. QA analyst conducts random audits of initial findings.
9. The recorder sends the committee a draft of the meeting minutes for approval.
10. The committee completes modified assignments.
11. The committee meets to decide recommendations.
12. The recorder finalizes meeting minutes, findings, and recommendations.
13. The PM approves all documents in the QA review packet.
14. The QA analyst files all final documents for the QA review.
15. The PM presents the final review with senior leadership.

## APPENDIX 8 – FOLLOW-UP BRIEFING CHECKLIST FOR PROJECT PROPOSAL

It's necessary to conduct a follow-up briefing to make sure stakeholders are aware of completed tasks and future milestones of any data management initiative. The list of items below should be considered with any follow-up briefing.

\_\_\_\_A day before the briefing, make sure handouts have been prepared, to include any reference material, such as website listings, points of contact, etc.

\_\_\_\_Arrive 45 minutes before the briefing. Take 15 minutes for the initial setup to test internet connectivity. Take the remaining time to organize materials and the presentation. Many presenters fail to take the necessary time to setup before a briefing. This can cause attendees to lose interest in the briefing.

\_\_\_\_Cover the background and purpose of the project. Some attendees may be unfamiliar with the project.

\_\_\_\_List everyone who was involved with the project.

\_\_\_\_Cover all completed milestones. It helps to have attendees follow a timeline.

\_\_\_\_List the items that have yet to be done.

\_\_\_\_Summarize what you need from the attendees, including senior management.



## APPENDIX 9 – SAMPLE POSITION DESCRIPTIONS

The sample position descriptions (PDs) in the next several pages were drafted for the purpose of creating a Data Management Branch (DMB). During the initial phases, it is suggested that three positions be created. A Quality Assurance (QA) analyst, GS-13, will be used to evaluate data management policies, procedures, and workflow using QA knowledge, techniques, and tools. This is an effort to strengthen protocols and check-and-balance systems for data management. The QA analyst will be used to train staff and QA committee members on quality assurance. A *Blackboard* training system is available with the Department of Homeland Security (DHS). This system will allow employees to take classes, take tests, turn in homework, and be graded by the QA analyst, similar to a process found in universities. If approved by senior management, the *Blackboard* training system will add depth to the organization's understanding of quality assurance.

A data analytics/management analyst, GS-13, position can be created to ensure there is one dedicated staff member for analyzing metadata and providing analytics to staff and senior management. The incumbent will be used to study meta data (sometimes in large quantities), streamline data gathering processes, train others on analytics software and dashboard set-up, and to bring a modernized knowledge of data management by collaborating with both the public and private sectors. The incumbent will work closely with the QA analyst to synergize both bodies of knowledge.

A program manager, GS-14, will be used to supervise the activities of the DMB. The incumbent will collaborate with other offices within both the CBP and DHS, to assist with data management goals, and will serve as the lead facilitator for data management process improvement. The incumbent will serve as the focal contact for senior management.

These position descriptions were created using language that came from classified PDs in the Fully Automated System for Classification (FASCLASS)<sup>4</sup>. However, some material is new, as it reflects duties that are specific to a new DMB. These PDs only show duties and do not have factors and other information normally posted to PDs. This is an effort to simplify the understanding of the proposed PDs.

## **QUALITY ASSURANCE ANALYST, GS-13**

Serves as a Quality Assurance (QA) Analyst with responsibility for identifying QA needs; planning and developing policies and technical guidance; implementing and maintaining a Quality Management System; and performing project manager/technical support for various data management processes used by Customs and Border Protection (CBP) agency offices. Provides technical advice on the interpretation and implementation of specific quality and product assurance requirements in all phases of plans, programs, program execution, and problem solving. Specializes in data management QA processes and reviews, to include analysis of tools and techniques used with data mining and analytics. Provides expert consultant services by developing instructions for QA elements. Advises program management officials and individual project teams in the application of QA principles, methods, and procedures. Serves independently as the technical expert regarding requirements of International Organization for Standardization (ISO) 9000 and participates on various teams advising on ISO 9000 requirements. Demonstrates expertise with developing check-and-balance systems for data quality. Serves as a QA liaison to boards, councils, and groups formulating QA policy, regulations, guidance, and practices. Maintains and extends a wide area networking capability throughout government and non-government organizations, to include components within the Department of Homeland Security (DHS). Serves as lead facilitator for QA committees. Establishes liaison with counterparts in a variety of agencies, private industry, and activities related to assigned mission and functional areas. May be called to participate in contract reviews as a QA advisor.

### **QA ACTIVITIES – 50%**

Identifies, analyzes and resolves QA problems. Responsible for development of test requirements for specific QA programs. Responsible for assuring Quality Deficiency Reports (QDRs) are investigated and resolved in a timely manner. Coordinates and conducts internal and external quality audits to assure operations are conducted in accordance with all legal, regulatory, and procedural requirements. Performs Root Cause Analyses (RCAs) of data quality problems. Coordinates and prepares responses to higher level authorities relating to assigned functions, or which have across-the-board policy or procedure implications. Prepares and publishes implementing agreements after coordination with other departments. Monitors or participates in studies and projects conducted by field activities to determine compliance with policies and procedures, and any recommended changes. Conducts staff visits for the purpose of determining the need for new or revised QA policies, procedures, and techniques. Conducts annual reviews of programs that are part of the Federal Managers' Financial Integrity Act (FMFIA). Ensures compliance with management and control processes within the agency as required with process owners and managers. As needed, collects and monitors QA techniques used in tabulating Key Performance Indicators (KPIs) across the CBP. Conducts audits of quality requirements and analyzes audit results, to ensure appropriate quality standards and operational definitions are in place. Uses risk assessments and quality standards to review past performance and plan future activities. Investigates and documents the internal controls of specified aspects of automated or partly automated processes, and assesses compliance with relevant standards. Takes responsibility for the control, update, and distribution of various quality standards, and

advice on their use. Collects, collates, and examines records as part of specified testing strategies for evidence of compliance with management directives, or the identification of abnormal occurrences. Analyzes data sets and procedures and drafts part or all of formal reports commenting on the conformance found to exist. Develops internal procedures and controls. Conducts studies and analyses of current operations to determine how modifications can be accomplished. Prepares and submits reports of findings, conclusions and recommendations. Performs site analysis and investigation. Recommends resolutions to problem areas and furnishes technical advice on adequacy or feasibility of subject matter presented. Participates in Inspector General (IG) reviews, as required. Supports the preparation and delivery of formal audit reports. Monitors and reports on the outputs from QA and audit processes. Controls, updates, and distributes new and revised quality standards, including technical changes.

#### **PLANNING – 15%**

As the QA analyst, independently plans, integrates, reviews and evaluates the substantive and technical activities pertaining to QA programs and data management processes. Coordinates work with various disciplines to continuously review and revise project QA methods and procedures. Identifies needs, plans and initiates requirements, and develops technical guidance and policies covering QA requirements. Based on an analysis of problems identified during quality audits or lessons learned, develops new approaches and methods for accomplishing the work and documents policy and technical guidance. Supports the development of organizational commitment to ongoing quality improvement by ensuring that QA processes are robust and based on Best Practices from various industries. Researches and studies policies and procedures of other large government and private organizations for possible application on new ideas and Best Practices techniques to improve quality, productivity, and cost reduction. Develops criteria and methods for evaluating quality accomplishments and trends. Contributes to implications of emerging technological developments. Formulates and recommends strategy for issues, initiatives, and ideas that are identified, directed, and collected. Maintains active dialog with staff. Serves as a facilitator for cultural change in the attitudes and approaches toward quality. Plans and schedules QA committee activities. Assists in setting administrative and technical priorities for QA teams.

#### **POLICY REVIEW – 15%**

Reviews program and project documents to assure compliance with legal, regulatory and programmatic guidance and advises team members on quality concerning QA requirements and inspection criteria, as well as in work methods to enhance QA posture. Researches, develops, prepares, and interprets all QA policies and procedures and then renders advice to headquarters staff, field installations, and activities. Researches, reviews, and interprets regulatory and directive material issued, or to be issued, by higher authorities. Develops and recommends improvements in QA policies, procedures, instructions, and management to higher authorities. Reviews proposed changes to higher authority regulations or directives for their effect on activities and facilities. Develops position papers, including any change or proposed revision to policies and procedures. Reviews correspondence and other material prepared by other organizational elements for conformance to policies and procedures. Conducts continuing review

and evaluation of the effectiveness of published directives (every three years). Participates with senior management in the development of QA policies.

**QA COMMITTEE – 15%**

Serves as facilitator and librarian of all QA committee activities, which may take place with CBP only, or extend to other DHS components. Activities include conducting audits, on-site reviews of contractor/government quality, and safety programs/processes. Documents and reviews corrective action requests resulting from on-site reviews. Participates in quality audits and/or quality training programs.

**CONTINUING EDUCATION – 5%**

Updates knowledge of new developments in the functional areas assigned, and skillfully presents innovative ideas to the organization. Attends various training courses and functions with the goal of gaining, and mastering, a modernized understanding of QA techniques and policies, from both the public and private sectors.

**Special Note:** It is advised that the selectee should have a Masters' Degree commensurate with a QA position and have extensive QA experience. This is to ensure the selectee has a broad and robust depth of QA policies, procedures, and techniques.

## **DATA ANALYTICS/MANAGEMENT ANALYST, GS-13**

Serves as a data analytics/management analyst responsible for lifecycle functional requirements for data management across various functional areas of Customs and Border Protection (CBP) agency data mining and analytics initiatives. Incumbent provides authoritative, expert technical guidance and advice covering a full range of analytical techniques, both routine and unique, or complex cases, and ensuring interpretation of the data is consistent using sound data mining processes. These factors are taken into consideration, along with state of the art applications development processes, in order to create suites of research data analysis tools. Once developed, these new tools and techniques allow for metadata configuration and analysis. Responsible for data management, analysis, and integration activities. Develops data analysis products (charts and graphics, etc.) to visually portray assessment results. Leads design workshops to improve functionality and satisfaction of user requirements for analytics software. Demonstrates comprehensive understanding in all phases of data management. Prepares and presents briefings and makes presentations on the assessments progress and results, to include preparation of technical reports and/or documents on progress and results. Incumbent is responsible for understanding, and updating knowledge of, business enterprise regulatory requirements and industry Best Practices for data management, data mining, and analytics techniques. Serves as the subject matter expert in data systems; informatics; data statistical reports; automation initiatives for rapid analysis and studies; testing new data mining algorithms; and training staff in use of tools and data mining techniques. Serves as an expert technical resource in database design and management, and provides input and communication of improvement opportunities to programmer. Provides guidance and technical support to management, program managers and others regarding the requirements, design, programming, testing, and implementation of databases. Acts as liaison to facilitate the sharing and use of content. Enhances full capability of sophisticated applications to extract, synthesize and parse data, analyze, and dashboard large data elements and sets, from many different sources. Incumbent organizes, implements, and executes analytical software applications following appropriate procedures for accomplishment of work.

### **DATA ANALYSIS – 50%**

Develops assessment tools that facilitate data input. Conducts data integrity and validation checks, and builds data mining techniques. Analyzes technical requirements, prepares programming specifications and instructions, prepares design documents, participates in testing, ensures database functionality, prepares compliance documentation, and assists with ensuring system meet security requirements. Develops, manages, implements, and administers content and document management systems for improving the relevancy and consistency of information (or knowledge) shared with users of data management systems. Uses quantitative analytical techniques such as operations analysis, cost benefit analysis, cost effectiveness analysis and systems analysis. Using Tableau and other analytics tools, analyzes large amounts of data from various sources in the management of the information systems integration processes and evaluates the validity of assumptions and the adequacy of analytical data used in predicting requirements. Uses appropriate statistical techniques to identify significant insights and findings. Responsible for extracting data from various databases for analysis and reporting. Uses business objects to extract data and perform analysis. Plans and carries out work, determines approaches

to be used, resolves conflicts, coordinates work with others as required, and interprets policy issues in terms of established objectives of the program. Continually analyzes the production process and develops goals, objectives, tools, policies and procedures for creating knowledge/content products. Identifies potential data deficiencies, provides technical guidance to functional users, and recommends action to resolve deficiencies. Ensures that data problems are resolved by responsible offices. Projected needs are analyzed in terms of policies and objectives for expanding, improving, or integrating data in information systems to the degree to which requirements can be met within existing or planned systems, and the urgency of requirements which cannot be met within such systems. Develops data analysis products and graphics to visually portray assessment results. Serves as an advisor to CBP QA committees.

#### **PLANNING – 50%**

Develops new analytical tools and methods to perform quality and data integrity checks using large query modeling. Develops interactive dashboards and visual analytics systems for leadership. The incumbent is tasked with the responsibility of providing programming support for future Informatics groups. The Informatics groups' primary task, is the design and development of advanced analytical techniques in data management, data mining, analytics, and dashboard development. Utilizing expertise in data processing, develops new machine learning and analytical models for advanced informatics applications. Develops assessment tools that facilitate data input, conducts data integrity and validation checks, and builds data mining techniques. Coordinates data collection efforts of metrics designed to assess how the agency is accomplishing its stated objectives. Works closely with national and local program development teams of federal government databases to develop automation tools to improve work production efficiency, improve data quality, and enhance data usability. Performs pivotal research with regard to on-going data management system initiatives for capturing, cataloguing, indexing, and presenting information. Recommends data management system strategies for reviewing and updating corporate knowledge products and the translation and integration of information and content management policies. Collaborates with other staff to assist with the development of novel clinical evaluation protocols and research projects. Works closely with the QA analyst to audit and reconcile discrepancies with systems and processes related to data management and data quality. Develops algorithms for extracting relevant information from datasets. Participates in, coordinates, or leads the establishment of data collection plans. Leads assessments configuration boards to discuss application issues, user requirements, and set future development priorities. Assists programmers with the development of improvements to existing federal government databases. Analyzes requirements to create databases to support an internal function or customer need. Researches the methods, mechanisms, and improvement opportunities for communicating information, and how that information is shared and understood. Continuously assesses future data management systems content requirements and the evolution of business practices that further define business relationships. Leads design workshops to improve functionality and satisfaction of user requirements.

## **PROGRAM MANAGER, GS-14**

Serves as the senior business analyst in the Data Management Branch (DMB) and senior advisor to leadership on informatics and Quality Assurance (QA) methodologies to achieve work goals and objectives. Responsible for performing and leading a wide range of executive support functions and management analyses concerning data mining, analysis, and QA processes. Responsible for developing and executing studies in support of business operations and support functions across the agency. Works with the executive management team on strategic planning and business optimization. Responsible for performing the following functions: coordination on data collection and data analysis directly related to Customs and Border Protection (CBP) agency offices with responsibility to collect, maintain, review, analyze, evaluate, report, and make recommendations on a wide variety of past, current, and projected data in various CBP offices. Provides objective advice and analyses to senior management on a wide range of challenging and sensitive issues. The incumbent is responsible for policies that assure effective development of programs regarding CBP-wide integration and strategic initiatives in support of senior-level decisions relative to data management, QA, and analytics.

### **PLANNING – 50%**

Leads the establishment of data collection plans. Leads efforts for development for assessment tools that facilitate data input, conducts data integrity and validation checks, and builds data mining techniques. Develops profiles, studies, strategic assessments, and analyses of data management information. Coordinates closely with multiple data collection agencies and consolidates data across various data sources. Consults with offices in the performance of data collection, analysis, validation, reporting, and QA processes. Directs professional and highly technical and complex informatics programs. Provides leadership and technical expertise in the design and development of new and innovative informatics initiatives, advanced analytics, and state-of-the-art applications to collect and analyze voluminous sets of data. Applies comprehensive professional knowledge to facilitate research within the field of informatics, recommending improvements in study design and data collection. Leads efforts for evaluating and refining processes; develops, implements, and refines decision support systems; and assists in the procurement, customization, implementation, management, evaluation and continuous improvement of information systems in support of new and innovative research protocols. This includes how they relate to QA policies and techniques. Leads the design and development of voluminous databases, applying expertise in the latest techniques to perform data mining and data modeling/analysis. Working closely with senior leadership to market branch capabilities, create new business opportunities, improves operations, and optimizes organizational efficiency. Develops, promotes, implements, and exploits data management capabilities, tools, techniques and practices throughout the CBP agency. Leads, plans, and conducts process improvement events to optimize processes and increase capacity. These events may require project management, change management, group facilitation, process mapping, data mining, and statistical analyses. Leads policy formulation, resource planning, fact-finding assignments and working groups. Plans and conducts training in various phases of data management.

#### **REPORTING/ANALYSIS – 25%**

Performs an integrated systems analysis to produce desired strategic or operational results. Develops reports and provides recommendations on strategic assessments, to include their development and refinement to senior leadership. Closely monitors new/incoming data to ensure completeness and comprehensiveness. Analyzes and organizes information for presentation to senior leadership. Factors influencing presentation of information include appropriate format, style requirements, historical data, and consistency with other information, for the intended audiences. Collects, interprets, and analyzes complex data from multiple sources to provide statistical insights. Provides analytical output to studies, reports, and other analyses as required. Analyzes past and current trends as applied to data management. Supports leadership with data subject matter expertise in data mining, analysis, and QA activities. Identifies problematic issues and immediately reports discrepancies and provides appropriate recommendations as required. Plans, organizes, and develops briefings as required for senior management. Collects, analyzes, and assembles data from a variety of sources to assist in preparation of daily, weekly, and monthly reports as required. Effectively assesses information and knowledge needs of CBP organizations. Provides technical, administrative, and professional expertise required for business metrics analysis using advanced quantitative analysis techniques and databases as appropriate. Coordinates with other entities in developing data management rules and maintaining current, accurate reports of requirements as needed. Develops data packages for data calls.

#### **SUPERVISORY DUTIES – 25%**

Incumbent performs the full range of supervisory duties for the DMB. Develops performance standards; plans work to be accomplished by subordinates; sets and adjusts short-term priorities; prepares schedules for completion of work; assigns work to subordinates based on priorities, selective consideration of the difficulty and requirements of assignments, and the capabilities of employees; and evaluates work performance of subordinates and gives advice, counsel, or instruction to employees on both work and administrative matters. As required, interviews candidates for positions within the unit; recommends appointment, promotion, or reassignment to such positions; hears and resolves complaints from employees, referring group grievances and more serious unresolved complaints to a higher level supervisor or manager; effects minor disciplinary measures, such as warnings and reprimands, recommending other action in more serious cases; identifies and arranges developmental and training needs of employees; finds ways to improve production or increase the quality of the work directed; and develops performance standards, counsels employees on performance, and writes formal performance appraisals. Reviews job descriptions for accuracy and makes necessary changes. Determines staffing requirements/mix based on analysis of demand and skill sets availability. The incumbent provides leadership and technical expertise that is characterized by substantial and continuous responsibility for informatics and QA programs. Provides technical guidance, leadership, and administrative oversight to professional, technical/support staff, and federal employees/contractors to achieve the informatics and QA vision and objectives. Leads the branch to deliver high quality analyses and innovative solutions to business challenges. Periodically reviews position structure of the business analytics organization and work assignments for immediate subordinate positions for propriety and to assure that positions are established to provide optimum balance between economy, efficiency, and skills utilization. As



deemed necessary, recommends ways to improve productivity or conserve manpower resources by realigning resources or work assignments. Develops/directs development of performance standards, evaluates employee performance and prepares or reviews performance appraisals. Approves/disapproves leave requests. Receives and attempts to resolve complaints, to include group grievances and more serious complaints from subordinates. Effects disciplinary actions in more serious cases. Determines, counsels and provides for employee training and career development.

## APPENDIX 10 – SAMPLE COURSE SYLLABUS DATA MANAGEMENT AND DATA QUALITY

Syllabus | Data Management and Data Quality | Course Number: TBD

### Course Description.

This course is an introductory overview of data management. It covers data quality; conducting reviews of data management reporting; and the creation of Quality Assurance (QA) processes to ensure accurate data reporting.

### Rationale for Course.

There is an increased demand for analytics (data mining large volumes of data). Technological improvements have led to the introduction of powerful analytical software applications, such as Tableau, making data mining projects easier to manage. However, these projects may encounter problems with data quality, specifically the absence of check-and-balance systems needed for data accuracy.

### Who Should Take This Course?

- any federal employee who has experienced data quality issues
- those who have active data mining projects
- staff who prepare or present slides for metrics, KPIs, and Balanced Scorecard data

### Objectives.

This course aims to:

- familiarize staff with common data quality problems and techniques that can be used to correct these problems
- have attendees understand basic QA methods that can be used in developing sound check-and-balance systems
- develop an appreciation of how data quality techniques can be used to produce more accurate metrics and presentations

### Learning Outcomes.

By the end of the course, staff should be able to:

- articulate the need and importance of data quality within data management processes and decision-making
- understand how analytics software can be used to pinpoint data quality issues and streamline reporting for KPIs, benchmarks, targets, metrics, and similar reporting tools
- use Microsoft Excel and other applications to determine deficiencies in data quality

### Exam

An exam will be given at the end of the course to test employees' knowledge of learning concepts.

## **Concepts and Techniques Covered.**

Module 1: Introduction to Data Quality Issues

Module 2: What Does Quality Assurance (QA) Mean?

Module 3: QA Methods for Uncovering Data Quality Issues

Module 4: Development of Check-And-Balance Systems

Module 5: Identifying Data Quality Anomalies

Module 6: Data Set Sampling

Module 7: Error Resolution Processes

Module 8: Risk Management

Module 9: Establishing Standards

Module 10: Developing Standard Operating Procedures

Module 11: Overview of Third Party Reviews

Module 12: Creation of QA Committees

Module 13: Training Employees on Data Processes and Cross-Coverage

Module 14: Best Practices Research

## APPENDIX 11 - FOOTNOTES

<sup>1</sup> In the case of data quality, a third party can be used to give an objective assessment of data. A third party can be an individual, or group. A third party must be independent (outside) of the work unit or process being reviewed.

<sup>2</sup> There are several governing bodies commonly referred to as accreditation organizations. They help develop industry standards; advise organizations on recommended practices, through accreditation reviews; and monitor management processes in both the public and private sectors. A few are listed below:

International Organization for Standardization. “ISO/TS 8000-1: 2011, Data Quality – Part 1: Overview.” [iso.org/standard/50798.html](http://iso.org/standard/50798.html). ISO. Dec 2011. Web.

Information Quality International – the International Association for Information and Data Quality. [iqint.org](http://iqint.org). IQ International. Web.

The Joint Commission. [jointcommission.org](http://jointcommission.org). Web.

The Commission on Accreditation for Law Enforcement Agencies, Inc., (CALEA). [calea.org/content/commission](http://calea.org/content/commission). Web.

<sup>3</sup> References for QA/data quality committees:

American Association of Collegiate Registrars and Admission Offices (AACRAO). “Data Quality Practices.” [aacrao.org](http://aacrao.org). AACRAO. Jun 2015: Page 1. Web.

WatchBlog. “The DATA Act Reporting Deadline Is Here-Will Agencies Show You the Money?” [blog.gao.gov](http://blog.gao.gov). WatchBlog, 8 May 2017. Web.

xBRL US. “Data Quality Committee.” [xbrl.us/data-quality/committee](http://xbrl.us/data-quality/committee). xBRL US, 2017. Web.

United States Department of the Interior. “Data Quality Management Guide, Version 4.5.” [nps.gov](http://nps.gov), Aug 2008: Page 1-4. Web.

Economics & Statistics Administration, United States Department of Commerce. “Commerce Data Advisory Council.” [esa.doc.gov/cdac](http://esa.doc.gov/cdac). Web.

<sup>4</sup> Fully Automated System For Classification. “Search by Position Data.” [acpol2.army.mil/fasclass/search\\_fs/search\\_fasclass.asp](http://acpol2.army.mil/fasclass/search_fs/search_fasclass.asp). Web.