

International Association for Data Quality,
Governance and Analytics (IADQ GA)

"Think with Data"

SUPPLY CHAIN DATA AUDIT GUIDELINES

Version 1.0 - March 2021

SPONSERED BY:



VISIT US AT WWW.IADQGA.COM

TABLE OF CONTENTS

International Association for Data Quality, Governance and Analytics	1
VISIT US AT WWW.IADQGA.COM	1
Preface	4
Today's Challenges	4
Introduction	4
Data Strategy	5
Why Is It Critical To Have A Data Audit?	5
Breadth and Depth of the Supply Chain Data Audit	6
Data As An Asset	6
Data as a Building Block	6
Data Efficiency and Effectiveness	7
Data Types	7
Data Formats	7
The 7 "Vs"	8
Data Governance	9
Data Bias	9
Audit Interval	9
Establishing A Supply Chain Data Audit Plan	10
Possible Data Audit Oversight Team Members	10
Data Audit Activities	10
Data Audit Timeline	11
Metrics	11
Identify Critical Data Sets	11
Key Performance Indicators (KPIs)	11
Data Strategy Review	12
Assessment of Data Literacy	12
Data Standards That Could Affect Supply Chain Data	12

TABLE OF CONTENTS

ISO 8000 – Data Quality	12
Key Concepts of ISO 8000	13
Defining Data Use	14
Map Data Flow	14
Map Decision Criteria Based on Data	14
Define Data End of Life Disposition, Retention and/or Permanent Storage	15
Questions of a “Supply Chain Data Audit”	15
Data Legend	17
Data Set Historical Chronology	17
Continuous Data Evaluation	17
Data Audit Action Plan	18
The Purpose of Data Tools	18
Data Cleansing and Organizing Tools	18
Data Quality Attributes	19
Electronic File Transfer (EFT)	19
Data Access & Usage	20
Topics	20
Description	21
Data Usage/Reliability	22
Data Audit Questions:	22
Data Analytics Capabilities	23
Data Analytics Maturity – (Readiness)	23
Training	23
Supply Chain Data Audit Expectations	23
Post Audit Activity	24
Audit Findings	25
Author & Reviewer	26
Special Recognition	27

PREFACE

“Data is shaping the future of humanity. The production distribution and consumption of digital data – the data economy – are driving rapid advances in machine learning, artificial intelligence and automation. Individuals and businesses are using data to reduce search and transaction cost and make informed choices. Data is facilitating scientific and medical research, making societies more productive. It is helping to improve the efficacy of public policy, delivery of public searches, transparency and accountability. Data is helping us track progress on every sustainable development goal in the 2030 agenda and deliver broad based social welfare.”¹

Today’s Challenges

In modern enterprises, supplier data is handled by multiple departments, and manipulated by siloed processes and disparate systems of governance. Within procurement, supply chain professionals are faced with the challenge of ingesting the enormous amount of data they have available to them. We all have a great appetite for data, but it’s not always easy to “digest”. Data often resides in various point solutions, with little to no consistency in how it is updated and evaluated for accuracy. This creates inconsistencies in the data structure that make it difficult to merge data from various sources. This means that reliable data that can be utilized with confidence will never be available in real-time.

What’s more, once you’ve done what you can to establish confidence in your data, the job is only half complete—the user must then make sense of the data and identify actionable insights. Easier said than done.

Having a single source of the truth that can process all that data is critical. Ideally, this single source of truth will eliminate the arduous task of establishing data confidence, so that procurement professionals can turn their attention to interpreting and working with data without worrying about its validity. These interpretations can be aided by assessing data in different ways, for example, by creating visual representations of data. Visualization allows professionals to quickly highlight patterns and outliers, saving a lot of time and making it easier to share insights with your internal stakeholders.

INTRODUCTION

A supply chain data audit is a critical early consideration for any digitization effort. In the majority of digital initiatives, this is not the case, and the importance of data is often neglected. The usual primary focus of any supply chain digitization is on the software tools to be used and the tools interoperability between data sources, its analytical capabilities and its user interface. While these issues are extremely important these should not be the first consideration. Data quality and governance should be the first

¹Source: Data Economy: Radical Transformation or Dystopia? January 2019 U.N. Department of Economic and Social Affairs

consideration to assure success. A good supply chain data audit will focus on at least three major areas: data governance, data quality and data analytic capabilities. There is growing quantitative evidence that the lack of “data quality” and “data trust” are significant factors in contributing to data latency.² A good supply chain data audit will involve internal users of the data, representatives of your supply chain, and senior management—ensuring that all stakeholders have the opportunity to address their confidence and trust levels in your organization’s supplier data.

When preparing for a supply chain data audit, it is critical to first understand the problem you are solving for, and establish a baseline of your data’s current status in terms of accuracy and completeness. When your organization can clearly articulate why you have a need for clean supplier data, it will be that much easier to evaluate if your supply chain data audit was successful.

WHY IS IT CRITICAL TO HAVE A DATA AUDIT?

For most organizations, the “why” is that poorly managed data leads to undesirable business decisions, which can cost significant time and money. According to the MIT Sloan Management review, the cost of bad data can account for 15-25% of revenue for most companies.³

Data is the life blood of all organizations. It’s use, in near real time, can make a critical differentiator for an organization. Organizations are now collecting and storing high volumes of both structure and unstructured data. These two data streams can easily and quickly contaminate decisions based on data if the data is not audited for relevance, completeness, accuracy, voracity, timeliness, velocity, etc.

The purpose of a data audit is to provide a level of confidence in the data prior to its use in both operational and strategic decision making.

DATA STRATEGY

Once you’re clear on why your organization will benefit from a supply chain data audit, you can take stock of your current data management processes across all arms of your organization. Where is your data stored? Who is managing it? What systems and processes will it need to power in order to move your organization towards digital transformation?

² Source: 2020 4th Annual Data Quality and Governance Study (<https://static1.squarespace.com/static/5ea48ae194c41a2df14fa825/t/602c4b5df20e6f748ef287e3/1613515622616/4th+Annual+Data+Quality+%26+Governance+Study+2020.pdf>)

³ “The cost of bad data is an astonishing 15% to 25% of revenue for most companies” (Source: MIT Sloan Management Review – November 27, 2017 – Thomas C. Redman)

You can begin to answer these questions by requesting a copy of the data strategy for the overall company and any additional strategies in place for the different lines of business. It should be confirmed that the strategy is still relevant to meet the needs of the business or if changes need to be made to act as a basis for the data audit. Remember, this is a “data” audit and not an “IT” audit although there will be some “IT” related questions.

By cross referencing your company-wide data strategy with your current vendor master, you will be able to understand the opportunities for improvement you can achieve with a supply chain data audit.

BREADTH AND DEPTH OF THE SUPPLY CHAIN DATA AUDIT

At the Supply Chain Data Audit planning stage, the scope of the data audit must be defined. The term “Supply Chain” has considerable breadth and depth. Often included in the term “Supply Chain” are areas such as sourcing, purchasing, contract administration, accounts payable, incoming inspection, production control/planning, logistics, etc. **It is critical that the breadth and depth of the data audit be documented and agreed to by all participants prior to the actual data audit.**

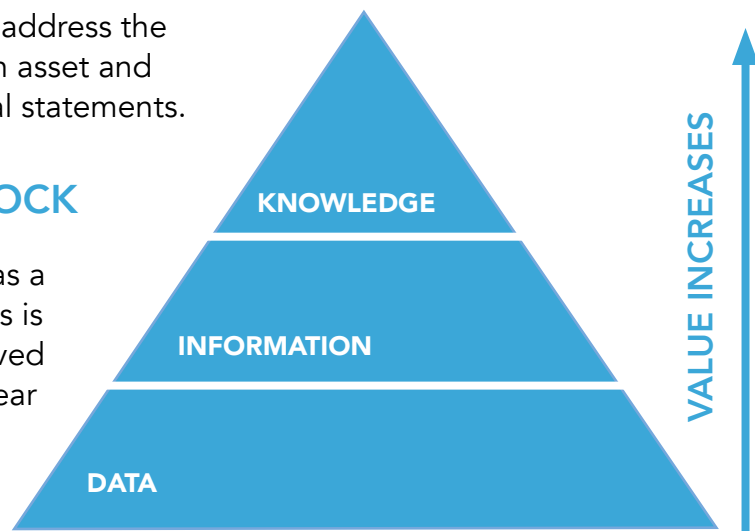
DATA AS AN ASSET

Most enterprises today are heavily dependent on data. Some newer companies are created solely around their data and the use of such data. However, most companies do not have a reliable way of handling data as a critical asset. Data, if addressed in a balance sheet, appears as an intangible asset. The valuation of the data asset varies and, in most cases, there is not a clear understanding of the useful life of the data asset nor is there a good method of depreciation for data assets.

The supply chain data audit must address the current state of viewing data as an asset and their representation on all financial statements.

DATA AS A BUILDING BLOCK

Data is well known to be utilized as a building block for knowledge. This is the desired end state to be achieved through enhanced utilization of near real time trusted data.



(Some examples utilize a fourth tier titled “Wisdom”. For this purpose, we are using the three tier representation.)

DATA EFFICIENCY AND EFFECTIVENESS

The efficient and effective use of data must also be included in a data audit.

Data Efficiency is defined as how efficiently the statistical algorithms and ETL process can be applied over the data.

Data Effectiveness is defined as how well the data can solve the problem within the defined constraints on the resources. For instance, classification of tweets for sentiment analysis can be made effective if we make data effective. Data is made effective through annotations.

DATA TYPES

1. Numerical Data (Quantitative Data)

Numerical data is all data that is represented by numbers

2. Non-Numerical Data (Qualitative Data)

Non-numerical data are usually represented by words. Non – numerical data can be a word, fact, opinion, etc. Non – Numerical data are hard to organize and usually represent an opinion.

3. Big Data

Big data is a term recently coined that means very large data sets. These data sets are often used in training neural nets in artificial intelligence (AI) or for training when using machine learning (ML).

4. Dark Data

Dark data refers to the massive amounts of data that are generated and collected but rarely, if ever analyzed or used.

DATA FORMATS

1. Structured Data

Structured data usually refers to data that is numerical and resides in a database and/or spreadsheet.

2. Unstructured Data

Unstructured data comprises emails, voice messages, video, text, etc.

THE 7 “VS”

The 7 “Vs” (shown in totality below) are a collection of concepts that aid in defining both structured and unstructured data forms.

1. Volume

Volume is how much data we have – what used to be measured in Gigabytes is now measured in Zettabytes (ZB) or even Yottabytes (YB). The IoT (Internet of Things) is creating exponential growth in data.

2. Velocity

Velocity is the speed in which data is accessible. Real-time data is now the only acceptable speed in most cases.

3. Variety

Variety describes one of the major challenges of data. Data can be structured or unstructured. Structured data is data that is in rows and columns. Unstructured data can include many different types of data from text, XML voice, video to SMS.

4. Variability

Variability is different from variety. **Variability** (also called spread or dispersion) refers to how spread out a set of **data** is compared to another set of data.

5. Veracity

Veracity is about making sure the data is accurate, which requires processes to keep the bad data from accumulating in your systems.

6. Visualization

Visualization is gaining in criticality in today's world. The use of charts and graphs to visualize large amounts of complex data is much more effective in conveying meaning than from spreadsheets and reports chock-full of numbers and formulas.

7. Value

Value is the end game. After addressing volume, velocity, variety, variability, veracity, and visualization – which takes a lot of time, effort and resources – you want to be sure your organization is getting value (value = quality + functionality/ price) from the dataset.

DATA GOVERNANCE

Data governance is an essential initial effort for any organization as they attempt to digitize their supply chain. Without a formal governance plan and a permanently assigned head of supply chain data governance, there is limited probability of a successful and sustainable digitization of your supply chain. Data is far more likely to be misclassified, disorganized and inappropriately interpreted whilst still being used in the decision-making process.

DATA BIAS

All data contains some form of bias. Biases can show up in analytics based on how a question is hypothesized and explored as well as how the data is sampled, organized and analyzed. Bias in data analysis can be from human sources using unrepresentative data sets, leading questions in surveys and/or biased reporting and measurements.

During the audit process, you should attempt to identify any bias in the data sets and document such biases identified for the benefit of the future data users.

Data bias can come about in many ways, including but not limited to:

- Propagating the current state
- Under-represented populations
- Faulty interpretation
- Cognitive bias
- Analytic bias
- Confirmation bias
- Outlier bias
- Timing bias

(NOTE: Biases also exist in algorithms, however this is a topic that is outside of this audit document's scope at this time)

AUDIT INTERVAL

Data audits should occur at a minimal interval of every 12 months. For critical data sets, it is recommended that audits of these data sets occur every 6 months or, in some cases, even more frequently due to the shelf life of the data, data volatility, etc. Audits of critical data sets should be "Zero Base" audits. "Zero Base" audits are audits that are conducted in their entirety and not partial audits examining only a limited number of criteria.

ESTABLISHING A SUPPLY CHAIN DATA AUDIT PLAN

Participation and representation are two critical components to the design and makeup of any audit plan. The creation of a board of audit team members is critical not only to the success of the data audit itself but also to the implementation of any data audit suggestions post-audit. In order to give a better idea of what such a board may look like, a collection of qualifiable members has been listed below.

Possible Data Audit Oversight Team Members

1. Chief Data Officer (CDO) if one exists
2. Data audit sponsor
3. Data end users
4. Data owners
5. Database administrators
6. Data analysts (if applicable)
7. Data scientists (if applicable)
8. Network/Data security
9. Supply chain management
10. Policies, regulations and procedures representative
11. Internal & external business auditors

The audit should have an agreed-to timeline for each step of the data audit processes along with identified leaders and specific deliverables for each step. The audit team should have a designated leader to make any final decisions in the case a committee consensus cannot be achieved.

Data Audit Activities

A documented data audit plan is critical to a data audit's success. The plan must include:

1. Assemble a data audit oversight team
2. Define the business data requirements
3. Define the metrics of success
4. Identify existing data set sources
5. Collect data sets
6. Analyze data sets
7. Investigate data set issues
8. Report format for data set audit findings
9. Identification of any data regulatory requirements:
 - a. EMEA
 - b. California

Data Audit Timeline

Estimated time (in weeks) to complete each activity based on company size:

COMPANY CLASSIFICATION	ACTIVITY									TOTAL WEEKS
	1	2	3	4	5	6	7	8	9	
SMALL / MID-SIZE	1	1	1	2	2	3	2	1	1	14
LARGE	2	3	3	4	4	5	3	2	2	28

(See "Data Audit Activities" section above for key to "Activity" classification numbers)

Metrics

Collect all governance and usage/reliability related metrics. Identify who reports these metrics, their frequency and acceptance by what level of management. What significance is placed on these metrics?

Identify Critical Data Sets

It is impossible for a data audit to review all of an organization's data sets. It is therefore critical that before the data audit begins, senior management and operational management identify those data sets that are critical to the operation and future success of the organization. For a better idea on this identification process, please review the following:

- The data sets should be broken into **at least 2 general categories**. The first category being "operational", and the second category being "strategic".
- The number of data sets included in any data audit **must be reasonable** to assure the audit is completed in a reasonable time—as defined by the impact such an audit could have on the operations and planning of such an organization.

Key Performance Indicators (KPIs)

Identify and document all KPIs for both the data and the data audit. Verify that all KPIs are still required. (KPIs and their purpose should be reviewed monthly given that they are access gateways to your data.)

Data Strategy Review

A review of any existing data strategy should be part of the “Data Audit”. The data strategy can usually be found in the Information Technology (IT) organization. Most data strategies address issues such as data format, data access, data storage, data security, etc. but usually do not include issues regarding the quality of the data, the viability of the data, the application of the data, etc.

Assessment of Data Literacy

Data literacy is the ability to derive meaningful information from data just as literacy in general is the ability to derive information from the written word. Data literacy is typically defined as the ability to read, understand, create and communicate data as information. It also means the ability to create and interpret graphical representation of the data, draw conclusions from the data and recognize when data is being used in misleading or inappropriate ways. The complexity of data analysis, especially in the context of big data, means that data literacy requires some knowledge of mathematics and statistics.

DATA STANDARDS THAT COULD AFFECT SUPPLY CHAIN DATA

There are several data standards organizations that could affect your data architecture, data categories, data definitions etc. One such organization is the International Standards Organization (ISO) (https://en.wikipedia.org/wiki/ISO_8000). Other standards groups exist but are not the focus of this supply chain audit guideline. At a minimum, you should be aware that many data standards exist for various functions and industries, and where and when applicable can be applied at your discretion.

ISO 8000 – DATA QUALITY

ISO 8000 is the global standard for Data Quality and Enterprise Master Data. It describes the features and defines the requirements for standard exchange of “Master Data” among business partners. It establishes the concept of Portability as a requirement for “Enterprise Master Data”, and the concept that true “Enterprise Master Data” is unique to each organization.

“Master Data” is commonly used to manage critical business information about products, services and materials, constituents, clients and counterparties, and for certain immutable transactional and operational records. Applications of this standard have already proven it can significantly reduce procurement costs, promote inventory rationalization, and deliver greater efficiency and cost savings in supply chain management.

ISO 8000 is one of the emerging technology standards that large and complex

organizations are turning to in order to improve business processes and control operational costs. The standard is in the process of being published as a number of separate documents, which ISO calls “parts”.

Key Concepts of ISO 8000

- Master Data

“Master Data” represents the business objectives which are agreed on and shared across the enterprise. It can cover relatively static reference data, transactional, unstructured, analytical, hierarchical and metadata. It is the primary focus of the discipline of “Master Data Management” (MDM). This discipline used to be predominantly taken care of by Information Technology (IT) departments but can equally well be justified as a business function, with IT providing the required technology.

- Quality Identifier

A “Quality Identifier” is an internal product or services identifier or key that is issued and “owned” by an organization and used to resolve a product or service to the minimum ISO 8000 quality data set required to validate the identifier.

- SmartPrefix

An ISO 8000-115 “SmartPrefix” is a unique name or alpha-numeric character string that is used by manufacturers and Distributors to uniquely identify products and replacement parts and link them to corresponding ISO Technical Specifications (ISO/TS).

- Authoritative Legal Entity Identifier (ALEI)

An ISO 8000-116 ALEI is an identifier issued by the administrative agency for the governing body of a nation, state, or community for a physical or juridical person for which they have granted legal status.

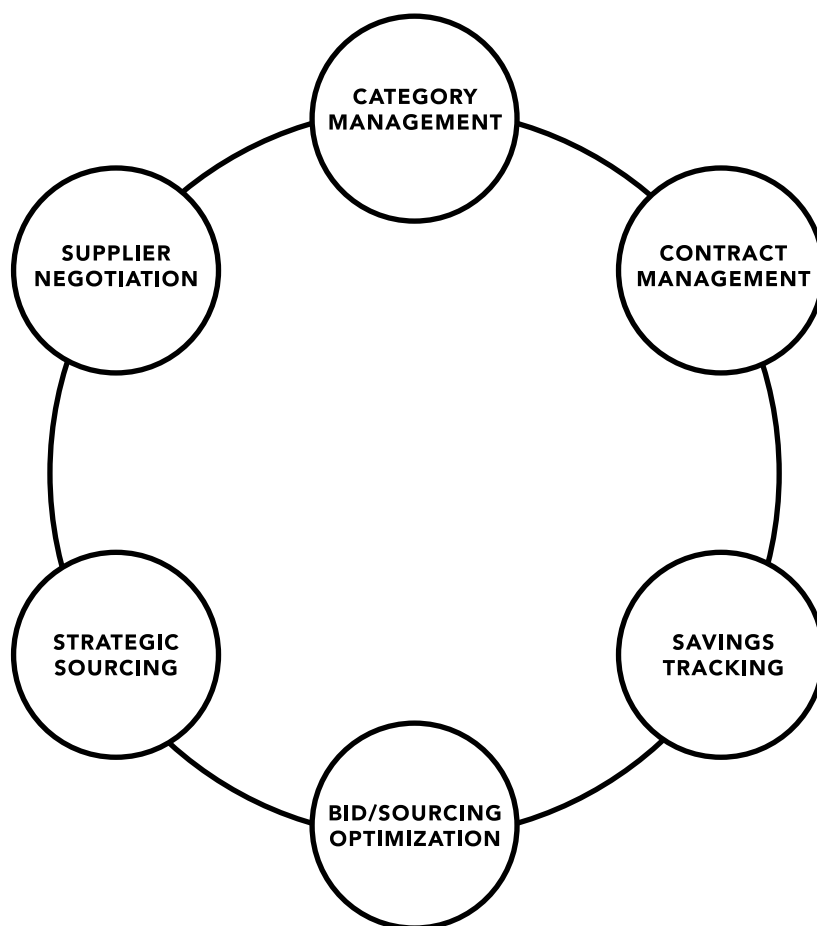
- Authoritative Item Identifier (All)

An All is a primary key or system identifier for the product or service that is defined in a dictionary and in an ISO 22745 XML specification in standard format that uses property or attribute value pairs for the item characteristics.

- Portability

Data portability is achieved when data is formatted using an agreed and known syntax and when the semantic encoding of the content is explicit.

DEFINING DATA USE



Map Data Flow

The mapping of data flow is a critical component of any data audit. Data flows from many sources within a supply chain process. There is both internal and external data which must be included in the data audit.

The external data flow is often difficult to audit beyond first tier suppliers and trading partners. Key components such as data formats, units of measure, frequency of data exchange, languages, etc all add to the complexity of a thorough data audit.

Map Decision Criteria Based on Data

Decisions are based on the analysis and consideration of relevant data. The audit should review the key decisions that are made both operationally and strategically. It should also map the data and knowledge transfer that occurs during decision making.

The decision criteria used during each activity within the decision-making process should be assessed for relevance.

Define Data End-of-Life Process: Disposition, Retention and/or Permanent Storage

Data retention should be part of any data governance plan. Data retention may be governed by law depending on the organization. Before addressing data retention, you should consult your company's general counsel to be sure your retention strategy is in compliance with any governmental regulation.

SUPPLY CHAIN DATA AUDIT QUESTIONS

Below is a minimum set of questions that should be part of every data audit. Additional questions, based on the type of business, may also be added. These questions should have a numerical range and weighing scale in order to develop a quantitative audit score.

Supply Chain Data Governance Audit Questions

1. Who uses the data?
2. What do they do with the data?
3. How often is the data required?
4. What level of detail is required?
5. What level of accuracy is required?
6. What is the source of the data?
7. What is the format of the data?
8. Where is the data stored?
9. Where is the data backed up?
10. What is the security around the data?
11. Is the data "free" or is there a "fee"?
12. If there is a "fee", what is the cost of the data annually?
13. Is the data Licensed?
14. Are there restrictions on the use of the data?
15. What level of decisions are made with the data? (operational or strategic)
16. Who owns the data?
17. Who is responsible for version control of the data?
18. Who is responsible for cleaning the data?
19. Who is responsible for organizing and categorizing the data?
20. Can the data be used for other purposes?
21. Do you know the lineage of the data?
22. What is the shelf life of the data?
23. Who determines the shelf life of data?
24. How do you prevent bad data from being reused?
25. Is the data 'restricted' or 'proprietary'?

26. Is the decision based on the data 'restricted' or 'proprietary'?
27. What is the criteria for having access to the data?
28. Who controls access to the data?
29. Is the data machine-generated or manually generated?
30. If machine-generated, how often are the sensors recalibrated?
31. Is the data certified by a third party?
32. Is the data estimated or actual numeric data?
33. Is there a separate data acquisition line in a budget?
34. Who determines the budget for data acquisition?
35. What is the cost of data management?
36. Is the data used to train Artificial Intelligence (AI) systems?
37. Is the data used to train Machine Learning (ML) systems?
38. Is the data used to make autonomous business decisions?
39. Is the data used for smart contracts?
40. Is there a master data library of terms?
41. What metrics exist for data quality, efficiency and/or effectiveness?
42. How do you define data quality?
43. What data governance exists?
44. Who sits on the data governance committee?
45. Do your RFQ's contain detailed data requirements, reporting intervals, etc?
46. Are you ISO 8000 certified?
47. If not ISO 8000 certified, do you plan to be? When?
48. Is data retention being addressed in the following areas?
 - a. Procedures
 - b. Approval
 - c. Location
 - d. Access
 - e. Security
49. Are the data sets used efficiently?
50. Are the data sets used effectively?
51. Who is accessing the data? (activity logs)
52. How often are they accessing the data?
53. What are the user account privileges?
54. Who manages software patches?
55. What does the database recovery plan look like?
56. How often is the database recovery plan tested?
57. How often is the database recovery plan reviewed?
58. Who owns the database recovery plan?
59. Have data biases been identified?

DATA LEGEND

The concept of a “data legend” must be part of your supply chain data audit. A data legend is information that accompanies a data file. The data legend contains such information as:

- Source of the data
- Original date of the data
- Version of the data
- Who created the data and their contact information?

Any edits to the data must also be reflected in the data legend. The information regarding edits would include:

1. Date of the change
2. New version reference number
3. List of changes made
4. Individual responsible for the changes and their contact information
5. Reason for the change(s) to the data

The data legend could be incorporated into the file name or appear as a separate section within the data set that is human readable.

DATA SET HISTORICAL CHRONOLOGY

When using a data set, it is critical that you understand the chronology of the data before you use it. A good tool to use for cleaning data is [OpenRefine 3.3](#). OpenRefine 3.3 is a free tool offered by Google. The tool has an Excel type interface and is extremely easy to use. It also creates a chronology of what was done to the data set to get to its current state.

CONTINUOUS DATA EVALUATION

A continuous data evaluation program must be defined, agreed to and implemented. Reliance solely on a quarterly, semi-annual or annual data audit is not acceptable without continuous data evaluation.

The scope of the continuous data evaluation should include all “critical data” identified from the data audit. (Data considered “critical” can change from audit to audit.). “Critical Data” is data that is used for strategic business decisions and/or across multiple business functions.

DATA AUDIT ACTION PLAN

FILE REFERENCE	FINDINGS	RECCOMENDATIONS	PRIORITY	AREA AFFECTED	RESPONSE	TARGET DATA	RESPONSIBLE OFFICER

THE PURPOSE OF DATA TOOLS

There are a range of data tools that are available to assist you in performing the:

- a. Cleaning of data
- b. Organizing data
- c. Interpreting data
- d. Making predictions based on data

Data Cleansing and Organizing Tools

There are many data cleansing and organizing software tools. Their capabilities, capabilities, and costs vary significantly.

Excel is still the default tool used by many companies both large and small. There are several built-in capabilities that can suffice for an initial data cleaning and organization initiative. (See IADQGA Data Cleaning & Organizing Paper) Excel is universally available and there are plenty of Excel experts in most organizations who currently possess the knowledge or can easily develop the know to utilize Excel as an initial data cleansing and organizing tool. (Excel typically can't be utilized in big data initiatives due to the cell and row limitations within Excel. (Excel maximum number of rows 1,048,576 by 16,384 columns)

Google Sheets is similar to Excel and is offered under G Suite by Google. Google Sheets has a more limited capacity than Excel. Google Sheets has a maximum of 40,000 rows and 18,275 columns. The analytical capabilities of Google Sheets is very similar to Excel.

OpenRefine is another free and powerful tool is Google's OpenRefine, as previously mentioned. OpenRefine is a powerful tool that operates at the browser level and whose interface resembles that of Excel. OpenRefine is more powerful than Excel as a data cleansing and organizing tool.

Data Visualization Tools

Several data visualization tools exist. The two most popular tools are Microsoft's - Power BI and Salesforce's - Tableau.

- **Power BI – (Microsoft)** Interactive data visualization software.
- **Tableau – (Salesforce)** Interactive data visualization software.

Other tools exist and new visualization tools and services are evolving.

DATA QUALITY ATTRIBUTES

ATTRIBUTE	DEFINITION
CONSISTENCY	No matter where you look in the database, you won't see contradictions in the data
ACCURACY	The information your data contains corresponds to reality
COMPLETENESS	All available elements of the data have found their way to the database
AUDITABILITY	Data is accessible and it is possible to trace introduced changes
ORDERLINESS	The data entered has the required format and structure
UNIQUENESS	A data record with specific details appears only once in the database
TIMELINESS	Data represents reality within a reasonable period of time or in accordance with corporate standards

ELECTRONIC FILE TRANSFER (EFT)

Data is constantly being transferred between databases, collected, summarized, etc. The vast majority of these transfers are handled via Electronic File Transfer (EFT). The audit should address the controls established around file transfers and conduct an appropriate number of EFT's to ensure data accuracy.

DATA ACCESS & USAGE

It is critical to review data access logs to understand who accessed the data, when, for how long, how often, etc. The identification of frequent data users can identify individuals to interview during the data audit to determine what can be improved upon to make them more efficient and effective in performing their jobs through the use of data.

Given the critical nature of data, it is important to include in the supply chain data audit a review of access and usage of data. The following questions/topics can be addressed to your database administrator and/or your data engineer. These questions/topics usually can't be addressed by your data analyst.

TOPICS	DEFINITION
ADMIN ACTIVITY	Displays all EFT administrator activity for the specified date range
ADMIN AUDIT LOG (DETAILED)	The displays all EFT administrator activity for the specified data range, with before and after details
EVENT RULES – JUST TRANSFERS	This details all offload and download actions, typically grouped by site sub-grouped by action, sorted by date in reverse chronological order
ACTIVITY BY PERMISSIONS GROUP	This displays the various actions performed by all the groups such as administrator, all users and all guests. It displays date/time, remote IP address, protocol, action, filename, folder KB transfer and the result.
ACTIVITY ALL USERS (DETAILED)	This displays all folders and files created and the delete activity for all users who logged on to EFT during a particular period grouped by username and sorted in reverse chronological order. The report should include time stamp, remote IP address of the user, protocol, action, file name, folder KB transferred and the result.
ACTIVITY ALL USERS (SUMMARY)	This report displays the transfer activity (total number of uploads and downloads) for all users who logged onto EFT during the date range specified, grouped by username, sub-grouped by date, sorted by username, then transfer direction and date in ascending order.
ACTIVITY FILE TRANSFER AS CLIENT SCANNED DATA RESULTS	This report displays all scanned data results
ACTIVITY FILE TRANSFER AS CLIENT	This report displays all file transfers initiated by EFT
ACTIVITY FILE TRANSFERS AS SERVER	This report displays all transfers initiated by remote clients

TRAFFIC - MOST ACTIVE IPS - CONNECTIONS	This report displays the most active IP addresses. These IP addresses show the users who frequently log on to EFT. It displays the data transferred, site name, remote IP address and bytes transferred. (This report can be used to determine "Denial of Service (DoS) attacks against EFT.
TRAFFIC – MOST ACTIVE IPS -DATA TRANSFERRED	This report displays the IP addresses of users who log on to EFT frequently, and the number of connections established by various users. It displays the information on the total bytes transferred, number of connections, remote IP address and the site name.
TRAFFIC – MOST ACTIVE USERS -CONNECTIONS	This report displays the connections established by the most active users.
TRAFFIC- MOST ACTIVE USERS – DATA TRANSFERRED	This report displays the usernames of users who log on to EFT frequently, the number of connections established by various users and number of bytes transferred.
TOP (X) USERS	By number of connections
TOP (X) USERS	By number of bytes transferred
SECURITY – FAILED LOGINS	This report displays the number of users who could not connect to EFT. It displays the username, remote IP address, Protocol used, date, time, remote IP address, port number and result.
TRAFFIC – AVERAGE TRANSFER RATES BY USER	This report displays the average transfer rate for specific users, grouped by username, sub-grouped by date, sorted by username, transfer direction and data in descending order.
WORKSPACES FILES PICKED UP	This shows shared workspace files picked up, organized by send date, sender email, recipient email, file name, file size (KB), subject and expires date.
WORKSPACES FILES SENT	This report lists workspace files sent, organized by send date, sender email, recipient email, file name, files size (KB) subject and expires date.
WORKSPACES FOLDERS SHARED	This report shows which folders have been shared for the given period, grouped by site, organized by date, workspace, path, owner/actor, action/status and participant/permissions.
WORKSPACES FOLDERS UNSHARED	This report shows which folder shares of expired naturally or by their owner, organized by date, workspace, path, owner or actor, status and participant(s)

DATA USAGE/RELIABILITY

Data Usage/Reliability Is the second major component of a supply chain data audit. The data usage/reliability audit component reviews and tests the raw data in the dataset to determine if it is capable or meeting its desired purpose.

Data Audit Questions:

1. What is the source of the data set?
2. What is the date of the data set?
3. Who is the author of the data set?
4. Who uses the dataset?
5. How frequently is the dataset used?
6. How frequently does the data change?
7. What is the dataset used for?
8. Has the dataset been altered or edited?
9. If yes, By whom, when and why?
10. Is the dataset clean of duplicates?
11. Is the dataset clean of blank fields?
12. Have data outliers been removed from the data set?
13. Have the names/descriptions in the dataset been standardized?
14. Segment the datasets into 1) heavily used, 2) frequently used and 3) rarely used
15. Segment the data sets into a cross set of criteria. These criteria are:
 - 1) critical strategic decisions, 2) normal day – to – day decisions and 3) low risk rare decisions

Data Audit Questions - Training

1. Is training mandatory?
2. How frequently are people trained?
3. Is the training at various levels of complexity?
4. What types of courses are offered?
5. Is training offered at normal work hours or after work hours?
6. Is training free to employees?
7. Are certificates required for specific jobs?
8. Who provides the training?
9. How much is budgeted annually for training?
10. Are there enough people available & trained to do the various tasks according to the workflow and expected turnaround times?

DATA ANALYTICS CAPABILITIES

An organization's data analytic capabilities consist of three components. The components are:

1. Technology
2. People
3. Process

Technology assessment addresses hardware, networks, internal consulting and software.

People assessment addresses the number of people, their skills, their ability to be trained in data literacy, data analytics, etc.

Process assessments evaluate work/data/information flow, activities, process owners, process metrics, compliance, etc.

DATA ANALYTICS MATURITY – (READINESS)

Organization data maturity and readiness levels vary widely and need to be assessed as part of any data audit. Shown below is a graphical representation of data analytics maturity level. As assessment of your organization to determine where you are on this line is critical as it should set the level of data analytics your organization can perform.

It is critical that once your level of data analytical maturity is determined this should be part of your final audit report and shared with senior executives to establish their level of expectation from the organization.

Data analytic maturity level consists of three basic components. These components are:

1. Data literacy
2. Statistical literacy
3. Ability to make data driven decision making (DDDM)

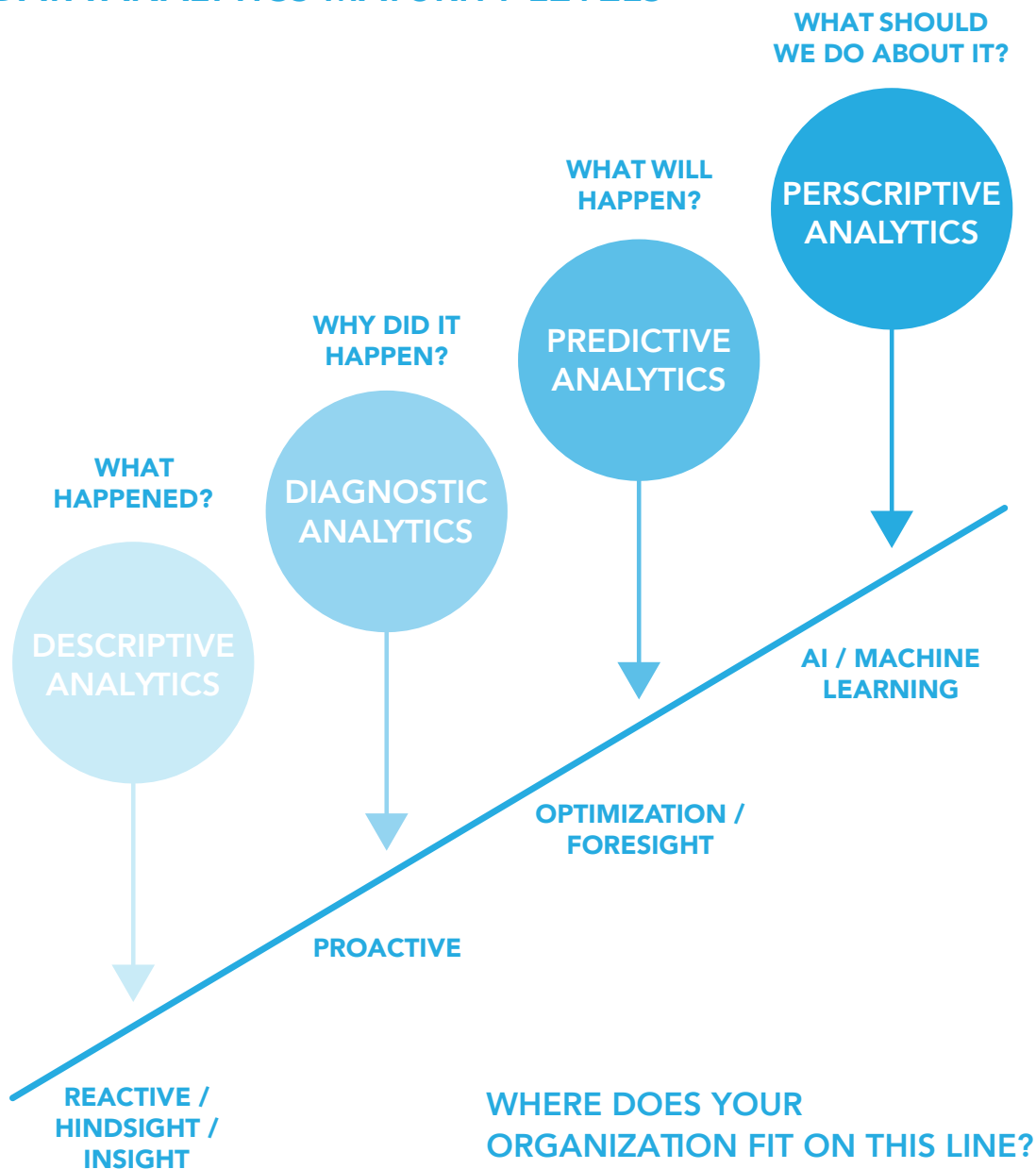
SUPPLY CHAIN DATA AUDIT EXPECTATIONS

The first supply chain data audit will most likely require a level of effort not anticipated during your planning stage. Your expectation regarding process and resources will probably need to be reassessed during the audit. Post audit corrective activity will probably exceed your original plan of time, resources and money.

POST AUDIT ACTIVITY

Undoubtedly there will be plenty of surprises and questions regarding the data sets as a result of your data audit. In reality your work has just begun. Given the critical nature of data in all companies and the increasing volume of data, audits of data should be a continuous ongoing process. The amount of work necessary to maintain such a critical organizational asset is significant and must not be underestimated or supported if you are to achieve the level of data quality and trust required going forward.

DATA ANALYTICS MATURITY LEVELS



AUDIT FINDINGS

The audit findings should be shared with the respective supplier when appropriate to identify any issues and areas of improvement. The data audit will probably be something new for your supplier(s) so you should be prepared to answer questions and assist your suppliers when and wherever possible. While your data audit may inform some changes to your supplier relationship management processes, it is intended to promote trust and transparency and should not be a punitive exercise for your suppliers.



JOSEPH A. YACURA
AUTHOR

M.B.A., M.S., M.Q.M. - Founder –
International Association for Data
Quality, Governance and Analytics

jyacura@iadqga.com



SAVANNAH LONG
REVIEWER

Research Intern – MS Data Analytics
Candidate - Villanova University

SPONSERED BY:



SPECIAL RECOGNITION

I would like to thank the following individuals for their contribution in reviewing and commenting of V1.0 of the “Data Audit Guidelines”.

Canada

Kate Hands
Stephany Lapierre
Shelley Lineham
Jervis Pinto - PhD

United Kingdom

Stylios Kampakis - PhD

United States

Nadiya Parekh – PhD
Kristin Sotak - PhD

South Africa

Nico Coetzee

WWW.IADQGA.COM