01-02-2022
ORGANIZATION

Data Health Assessment

For

Organization

Document Number: PDQH-ORG-DHA-001

Document Information			
Number	PDQH-ORG-DHA-001	Version	1
Effective	Date: 01-02-2022	Issued	Date: 01-02-2022
Compiled by	PiLog Cloud		

Amendment History				
Version	Date	Changed Chapter/Topic/Page	Pages	Checked By
01	Date: 01-02-2022	First Release	17	CoE Team

About This Document

This document contains detailed information about the material master data of Organization. An analysis was done on the data obtained from Organization and the details of the analysis were compiled in this report.

This report contains the following information:

The Scope of Analysis
Methodology adapted
Current status of Organization data
The next course of Action
Suggestions and Recommendations

Acronyms and Definitions

Acronym	Definition
ISO	International Organization for Standardization
ERP	Enterprise Resource Planning
FFT	Free Format Text
IT	Information Technology
MDM	Master Data Management
MDRM	Master Data Record Manager
MFR	Manufacturer
OEM	Original Equipment Manufacturer
OTD	Open Technical Dictionary
PPO	PiLog Preferred Ontology
PPR	PiLog Preferred Records
PTNO	Part Number
SPL	Supplier
DQH	Data Quality Hub

Contents

1. PREFACE:

Data is of high quality, if it fits for its intended uses in operations, decision-making, and planning.

Data quality refers to the state of qualitative and quantitative pieces of information. Data is that which provides information about other data.

A Data Health Assessment (DHA) profiles and analyzes the quality and integrity of your master data (e.g., Material, Vendor, Service, Asset, Customer, Business Partner, Equipment's, Location etc.,). A DHA visualizes,'AS-IS' state of your data quality, completeness, consistency, conformity to standards, and duplicates. The report then summarizes the impact that poor data has on your business and recommends the next steps to achieve measurable data and business metric optimization. The analysis process can be conducted across multiple domains that may involve datasets for customers, vendors, materials, and finance types of information

2. PROCESS & METHODOLOGY TO DELIVER SCOP

PiLog utilized proprietary automated processes to generate highly probable data points that indicate the data health and quality of the customer Data Set.

These data quality processes and methodologies have been developed and proven over several years and allow us to readily identify certain characteristics indicating problem areas in both data and systems within an organization

The approach is to determine records that are duplicated and unique, structured, standardized and rich in item property or characteristic values to give an overview of quality

To determine the scope of the underlying root causes and to plan the ways that tools can be used to address data quality issues, it is valuable to understand these common data quality dimensions:

A. Completeness of Master Data:

Data completeness refers to an indication of whether or not all the data necessary to meet the current and future business information demand are available in the data resource.

It deals with determining the data needed to meet the business information demand and ensuring those data are captured and maintained in the data resource so they are available when needed. Various processes include Data Extraction, Data Transformation, Data Loading, Security implementation & Job Control.

B. Redundancy of Master Data:

Data redundancy ensures that data is not duplicated un-necessarily across any part of the system

The online store may have a sales department and a complaints department. When a user buys a new product, this will be store as a transaction, but both the sales department and the complaints department may need this information at some point.

It would be possible for each department to have its own separate database and each time there is a transaction it gets put into each database. This solution is not ideal though, as having to input the same data into multiple places risks user entry errors leading to data consistency problems.

C. Compliance of Master Data:

Data compliance refers to a state of being in accordance with established guidelines or specifications.

The growing number of quality standards and regulations (industry specific or not) has also drawn attention to Master data management. In order to comply with these requirements, companies must meet certain criteria which are directly or indirectly impacted by the quality of data in the systems. There are many compliance risks that companies run from having bad Master data management:

- SOX risks occur in maintaining reporting structures and processing critical master data such as vendor bank accounts, fixed-asset data, contracts and contract conditions
- Petrochemical industry companies that are regulated by refining safety and operational standards and recommended practices may have significant exposure to legal risk and could even lose their operating licenses if their master records are incorrect with respect to product composition, storage locations, recording of ingredients, etc.
- Fiscal liabilities, such as VAT, produce risk. The VAT remittance may be incorrect if the relevant fields in the master data are not appropriately managed, possibly leading to inaccurate VAT percentages on intercompany sales

D. Consistency & Integrity of Master Data:

Data consistency refers to the transparency of the information, or the ability for other to see the changes and trends of data

It is when you ensure that the same data that is being used in different parts of the system will always be the same. The data is consistent across the entire system

In many examples, you may find the same data in needed in more than one place. In the online store example, the users address may be part of their signup information, but it will also need to appear on delivery information. Having to input the data directly into every place that it is needed causes problems if that data needs to be changed

If the user moves home, then every place their address appears will need to be manually changed to ensure the data in consistent. If you do not do all of these updates, then you'll have a situation where the address in one part of the system will be different to the address in another part of the system. This is solved by centralizing the data so there is one place that the address is stored. Any part of the system that needs to have the address can then just reference back to that central address location and find to find the data. When you do this you only have the change the data once for that change to spread through the whole system, ensuring that data always remains consistent.

Data integrity refers to the accuracy and reliability of the data being collected

E. Accuracy of Master Data:

Data accuracy is one of the components of data quality. It refers to whether the data values stored for an object are the correct values. To be correct, a data values must be the right value and must be represented in a consistent and unambiguous form

For example, birth date is December 13, 1941. If a personnel database has a BIRTH_DATE data element that expects dates in USA format, a date of 12/13/1941 would be correct. A date of 12/14/1941 would be inaccurate because it is the wrong value. A date of 13/12/1941 would be wrong because it is a European representation instead of a USA representation

There are two characteristics of accuracy: form and content. Form is important because it eliminates ambiguities about the content. The birth date example is ambiguous because the reviewer would not know whether the date was invalid or just erroneously represented. In the case of a date such as 5 February, 1944, the USA representation is 02/05/1944, whereas the European representation is 05/02/1944. You cannot tell the representation from the

value and thus need discipline in creating the date values in order to be accurate. A value is not accurate if the user of the value cannot tell what it is

The concept of accuracy also applies above the data element level. Data elements are never recorded in isolation. They are value attributes of business objects such as personnel records, orders, invoices, payments, and inventory records. The business objects represent real-world objects or events, and each consists of one or more rows of one or more tables connected through keys. Object-level inaccuracies consist of objects that are missing, have missing parts, or that exist but should not

F. Provenance of Master Data:

Data provenance documents the inputs, entities, systems, and processes that influence data of interest, in effect providing a historical record of the data and its origins. The generated evidence supports essential forensic activities such as data-dependency analysis, error/compromise detection and recovery, and auditing and compliance analysis

The provenance of data which is generated by complex transformations such as workflows is of considerable value. From it, one can ascertain the quality of the data based on its ancestral data and derivations, track back sources of errors, allow automated re-enactment of derivations to update a data, and provide attribution of data sources. Provenance is also essential to the business domain where it can be used to drill down to the source of data in a data warehouse, track the creation of intellectual property, and provide an audit trail for regulatory purposes

The use of data provenance is proposed in distributed systems to trace records through a dataflow, replay the dataflow on a subset of its original inputs and debug data flows. To do so, one needs to keep track of the set of inputs to each operator, which were used to derive each of its outputs

G. Uniqueness or Duplication of Master Data:

Duplication of Master Data is a very common area of concern, normally resultant of systems that do not mitigate causes of bad data. Factors involved in capturing Master Data are derived from different procedures; multiple users, locations and languages; data degradation; multiple versions of master data; variation of standards/practices over time and human error

Hence, uniqueness of mater data items is the first check performed on a master data set to determine if processes and systems generate master data that is not duplicated. The lack of unique items or the presence of duplicated items implies that a Governance Structure including master data record management systems solution that forces end users to verify the nonexistence of the item before creation is not in place

Item uniqueness is examined across the following data elements:

i. Material Numbers

ii. Descriptions

iii. Manufacturer Names / Supplier Names and Part Numbers

Uniqueness also helps determine the potential initial data set to be used for cataloguing. If material numbers are not determined to be unique, then descriptions will be used. If descriptions are not determined to be unique then a combination of descriptions and manufacturing names / supplier names and part numbers will be used

H. Structure of Master Data:

Un-Structured data is data that has no relevance. Structuring begins with the development of data requirements for a business function and its related objectives. Master data usage and relevance is reliant upon master data having structure to ensure consistent classification and identification of data to continually eliminate data reconciliation issues. Further, structuring data, complimented with a record Management tool, will contribute to the elimination of item duplication

Lack of structure implies that the business functions within the organization are underperforming in relation to their objectives. Master Data must have uniform structure to ensure the elimination of duplicates or uniqueness, accuracy and relevance. Hence, PiLog will analyze the current data structure used to determine if a deliberate data structure is used to support business objectives

I. Standardization of Master Data:

Standardized data is reliable and has consistency across multiple items. Data Standards, complemented with Data Structure and uniqueness, provide assurance that data viewed

and used is all the data that exists within the master data set, therefore giving the business functions assurance that their business decisions are based on accurate information There are a number of industry standards, such as ISO 8000 that can govern master data and the use of a dataset. This analysis will identify the use of consistent standard rules. Standardization also contributes to elimination of item duplication

J. Data Richness:

Beyond uniqueness, structure and standard, there is a remaining data principal that indicates the overall quality of master data, Data Richness. Data Richness is the amount of properties available to describe an item and its' related relevance. Relevant properties ensure that master data is traceable, verifiable and complete as it is used by the organization. These relevant properties are described as Mandatory Property Values, which ensure the uniqueness and correct usage of that item, in particular to the purchasing of the item

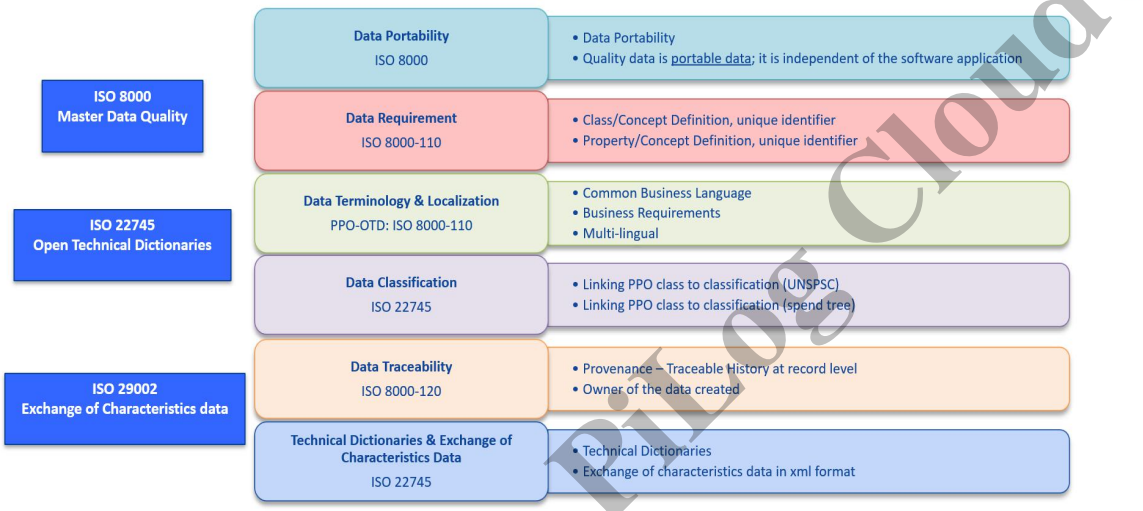
Often organizational data does not contain all the necessary Mandatory Property Values and this leads to the existence of multiple versions or item duplication. By enriching Master Data property values, especially for critical items, duplications are further reduced; uniqueness is improved; and relevance and completeness of items are increased for the purchasing and inventory management cycles

Our Data Assessment is a services engagement backed by our proprietary algorithms that delivers report findings identifying specific data challenges that may be hindering your operational efficiency and ability to achieve successful business outcomes based on the health of your data.

The ISO 8000 standards are high level requirements that do not prescribe any specific syntax or semantics. On the other hand, the ISO 22745 standards are for a specific implementation of the ISO 8000 standards in extensible mark-up language (XML) and are aimed primarily at parts cataloguing and industrial suppliers

Data Harmonization processes & methodologies complies to ISO 8000 & ISO 22745 standards

Data Quality Standards



3.CONSIDERATIONS ON DATA ANALYSIS:

3.1 Consideration on Data Analysis for DHA:

The following considerations or highlights considered for analyzing the Data Health Assessment:

serial_no	Guidelines
1	As some of the material numbers are repeated, the uniqueness of the materials is based on unique descriptions and Record number
2	Description length is analyzed based on number of characters in each Descriptions
3	Same or Duplicate short Descriptions with different material number are found by sorting all the Short descriptions with respect to Material Numbers
4	Same or Duplicate Long Descriptions are found by sorting all the Long descriptions with respect to Material Numbers
5	Inconsistencies in the given data is detected by analyzing the data like Part Number is in different formats as "P/N", "PN" etc.
6	Material of construction for the spare are not given in standardized way, it is given in full form (like Stainless Steel) and also in short form (like SS)
7	The data was analyzed specifically to find out the inconsistencies in Unit of Measure. Itis also in different formats like "Watt, W" etc.
8	Data richness on Reference Data Availability (Part Number, Reference Number, Model Number etc.) is

	Analyzed and counts are given
9	Duplication of Master Data w.r.t to Manufacturer Reference numbers like Part number or Model Numbers are analyzed based on genuineness of Part number with Same Manufacturer etc.
10	Duplicate Part Number with different Part Type, In this criteria Same part number is provided with different flag type (Suppl P/N, OEM Part No, Mnfr Part No).
11	Analysis on Top 10 Commodities and Top 10 UOM are performed by analyzing the Descriptions

A good measure of the quality of Master Data is a Median Grade; when 80% of all the Master Data per Characteristic are equal or above the grade B-

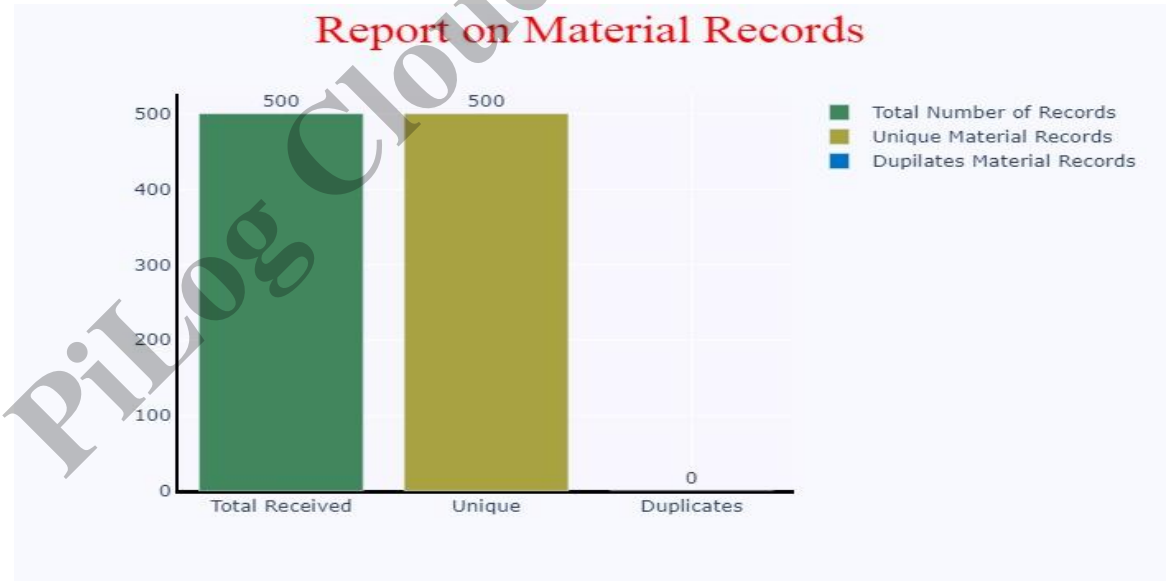
4.DATA HEALTH AND QUALITY ANALYSIS RESULTS

4.1 Detailed Data Analysis on Data Uniqueness

4.1.1 Analysis on Material Records

Report on Total Material Records uniqueness

Criteria	Count
Total Number of Records	500
Unique Material Records	500
Duplicate Material Records	0



4.1.2 Analysis on Descriptions/Texts

Report on Material Descriptions Uniqueness is as below

Criteria	Count	Percentage
Duplicates on Short Description	0	0.0
Duplicates on Long Description	0	0.0
Duplicates on Short and Long Description	0	0.0

4.1.3 Advanced Analysis on Descriptions/Texts

Below is the report generated after exclusion of unwanted text in the source description (E.g.: Special characters, Prefixes)

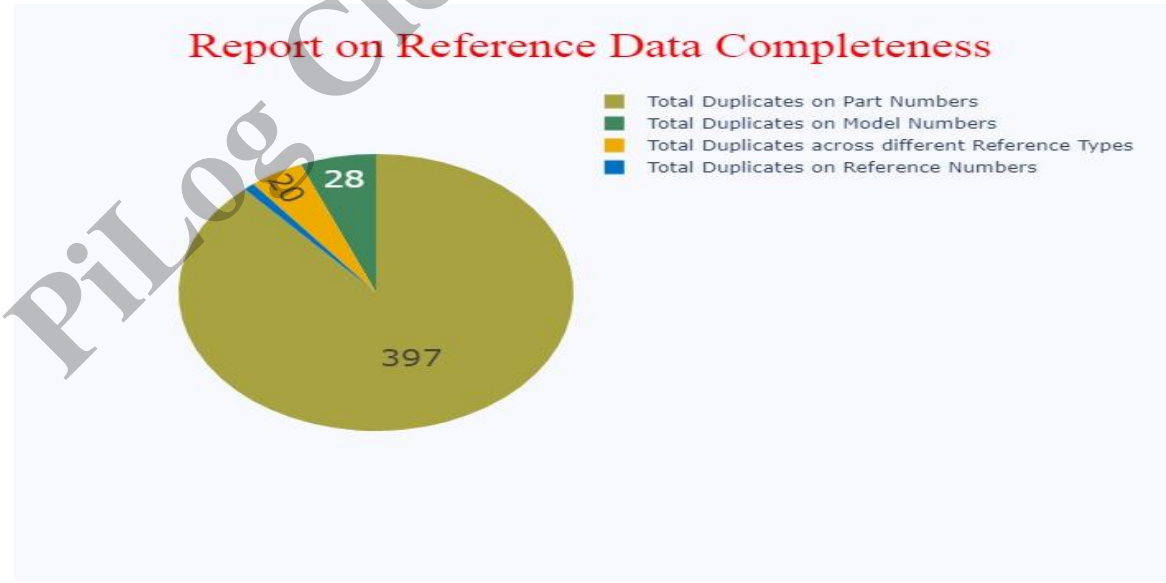
Criteria	Count	Percentage
Duplicates on Short Description	0	0.0
Duplicates on Long Description	0	0.0
Duplicates on Short and Long Description	0	0.0

4.1.4 Analysis on Reference Details

Report on Reference Data Uniqueness is as below

Criteria	Count	Percentage
Total Duplicates on Part Numbers	397	79.4
Total Duplicates on Model Numbers	28	5.6
Total Duplicates on Reference Numbers	4	0.8
Total Duplicates across different Reference Types	20	4.0

To download complete Reference Details Data click here



4.1.5 Potential Duplicates w.r.t Reference Number

MATERIAL	LONG_DESCRIPTION	Reference Number

4.2 Detailed Data Analysis on Data Completeness

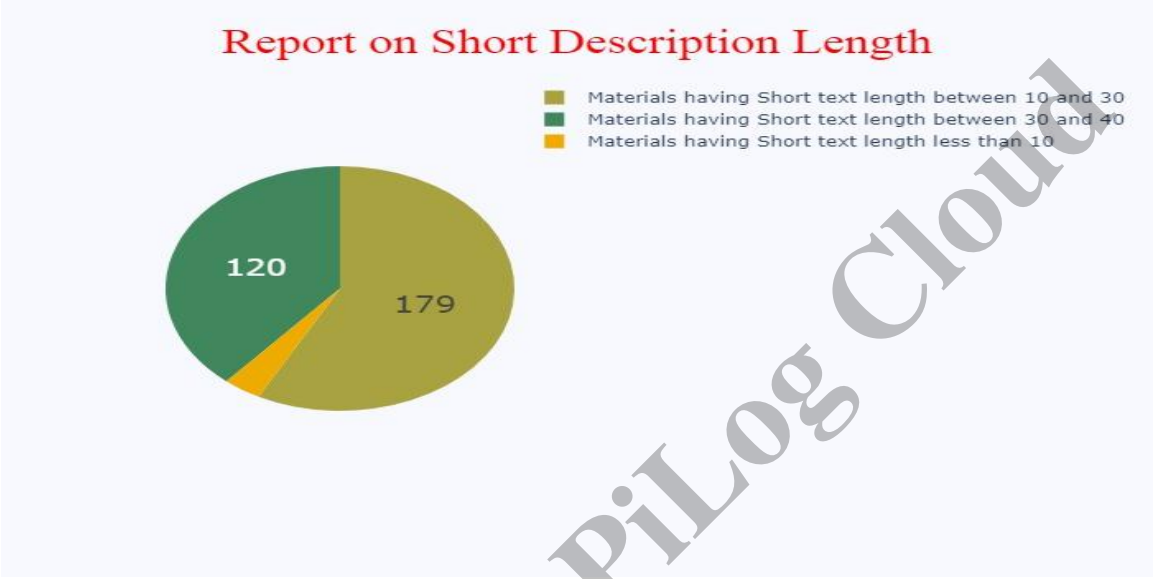
4.2.1 Analysis on Description Length

Below is the Analysis on description lengths of short descriptions

Criteria	Count
Materials having Short text length less than 10	11
Materials having Short text length between 10 and 30	179
Materials having Short text length between 30 and 40	120

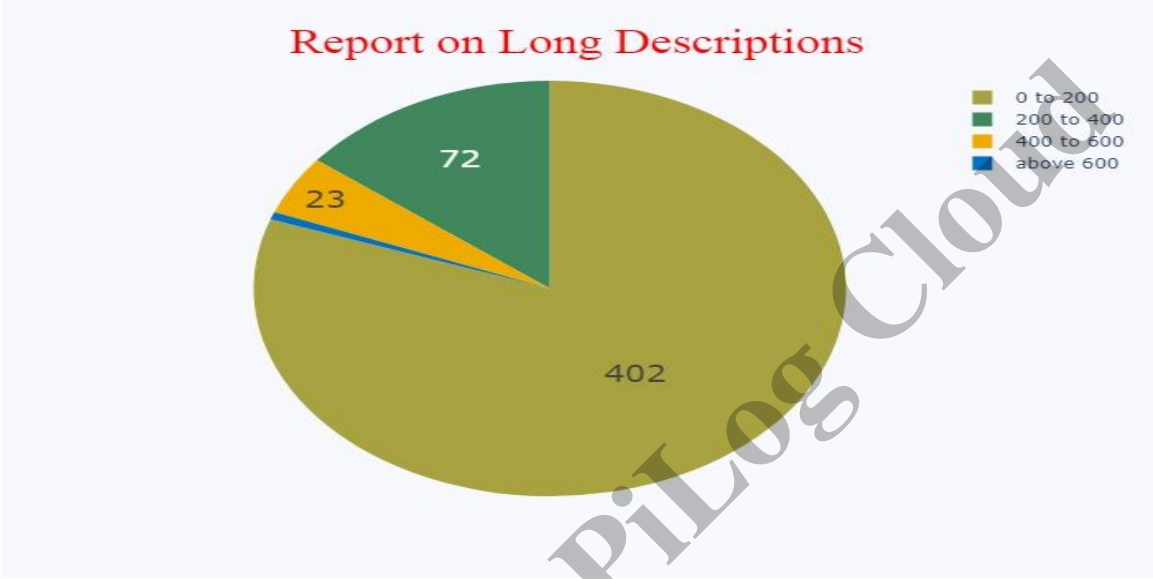
Analysis on Min,Max,Medium Lengths

Criteria	Length
Minimum short text length	11
Medium short text length	120
Maximum Short text length	179



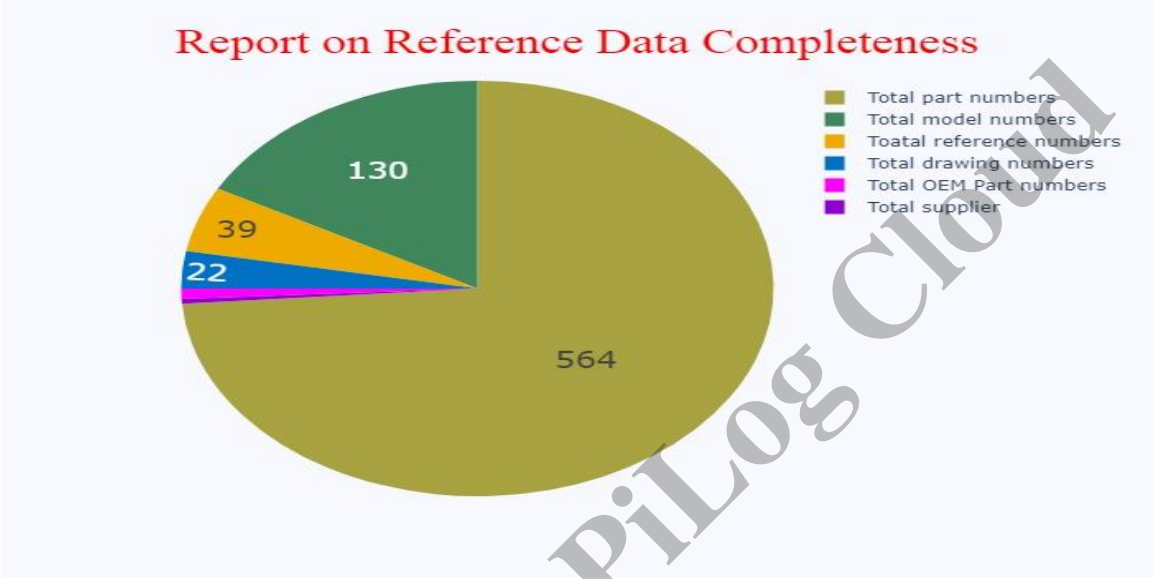
Below is the Analysis on description lengths of long descriptions

Criteria	Count
Long Descriptions with character length ranges between 0 and 200	402
Long Descriptions with character length ranges between 200 and 400	72
Long Descriptions with character length ranges between 400 and 600	23
Long Descriptions with character length ranges above 600	3



4.2.2 Report on Reference Data Completeness

Criteria	Total Number of Records	Percentage
Total part numbers	564	112.8
Total model numbers	130	26.0
Toatal reference numbers	39	7.8
Total drawing numbers	22	4.4
Total supplier	3	0.6
Total OEM Part numbers	6	1.2



4.3 Detailed Data Analysis on Data Consistency

Standard Format	Variant Format	Number of Materials Linked
MANUFACTURER PART NO	DESIGNATION	1
DRAWING	DRAWING NO	2
DRAWING	DWG	16
MANUFACTURER PART NO	MANUFACTURER PART NO	196
MODEL/MACHINE NO	MODEL	125
MODEL/MACHINE NO	MODEL NO	40
OEM PART NO	OEM	53
OEM PART NO	OEM PART NO	2
MANUFACTURER PART NO	P/N	12
MANUFACTURER PART NO	PART NO	198
MANUFACTURER PART NO	PN	116
REFERENCE NO	REF	45
REFERENCE NO	REFERENCE	1
SERIAL NO	S/N	26
SERIAL NO	SERIAL NO	2
SUPPLIER PART NO	SUPPL	14
SUPPLIER PART NO	SUPPLY	3

To download complete Inconsistencies Prefixes Data click here

4.3.1 Detailed Data Analysis on Data Consistency

Standard UOM	Variant UOM	Number of Materials Linked
Current	A	61
Current	AMP	1
Current	AMPERE	1
Voltage	V	34
Voltage	VOLT	1

To download complete UOM Prefixes Data click here

4.3.2 Inconsistencies in the Description:

To download complete Inconsistencies Data click here

4.3.3 Non-standardized UOMs in Description

To download complete Non-Standardized UOMs Data click here

4.4 Analysis on Top 10 UOMs

UOMs are listed below:

UOM	Top10 Count
EA	500

5. Key Recommendations

- Establish a Master Data vision that recognizes Master Data as an asset that reduces cost, risk and exposure through efficient use of unique and reliable master data within the organization, related to external partners and customers
- The organization should convene a working group (e.g., data stewards) representing all relevant stakeholders to determine targets, set thresholds, and define the quality dimensions that are most important

- Periodic assessments should be conducted to determine if acceptable thresholds and targets are being met, and metrics should be updated accordingly
- Implement data standards to drive standard data in accordance with data requirements
- Implement data workflows to manage data approval criteria where the engineers review the data quality
- Implement advanced Master Data governance tools to control the implementation of data standards, work flows, enrichment of data, data lifecycle and uniqueness of data
- Set up analytics for contracting and spend analysis
- Establish Business Rules to align the organization on Cataloguing Standards such as language, measurements, units of measure, presentation of data, etc. using ISO 8000 standards
- Create standardized Short and Long Descriptions for all items to ensure purchasing accuracy
- Harmonize data with a final potential duplicate resolution process across and languages to ensure excess stock is used and not reordered or duplicated between the sites