eyeon
SOLUTIONS

MASTER DATA MANAGEMENT

AN EYEON SOLUTIONS WHITE PAPER

November 2016

CONTENTS

1
S&OP DATA SCIENCE SCAN

1.1. What is an S&OP data scan?

Good master data has many advantages. Not only the data itself, but also the process around it will bring many advantages! A good data process in place means that organizations can trust the data and convert it quickly into knowledge. One of the many benefits is that redundant data is eliminated because when master data is centrally located and maintained there is a complete coherence and specifics regardless of which end-user or department uses the data. It is easier to make effective data analysis and to gain a better insight and understanding of how the organization is performing. The decision-making process is more effective and faster and the focus will be on analyzing the data instead of discussing the quality.

Non-alignment of the master data has a detrimental effect throughout the whole organization, especially when there is no process in place in order to maintain the alignment. This occurs, for example, when each business unit is responsible and accountable for their local data but without any synchronization between the business units. This creates a problem organizationally at a high level because misalignment of the local data sets makes it difficult to generate reports. Another disadvantage is the amount of room for discussion about how to interpret the data and identify the source(s). The discussions will have a negative effect on business decisions.

The EyeOn data scan identifies how the data is connected and also the main related issues. The scan answers relevant questions, such as: What are the main data sources? Who are my data owners? How can I manage my data? Who are the consumers of the data? How "clean" is my data?

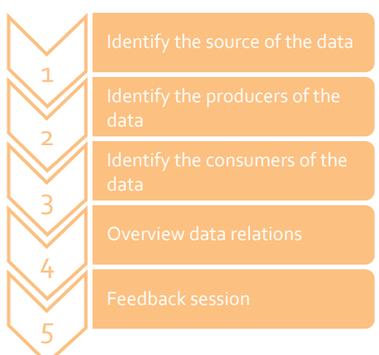

Figure 1: Data scan steps.

Approach

1. Identity the sources of the data
- Process system mapping to understand the data chain
- Agree on scoping
2. Identity the producers of the data
- Collect all information about the creation of data
3. Identity the consumers of the data
- Collect all information about how and who is using the data
4. Provide an overview of data relations
- Generate a data relation overview
5. Feedback session
- Face-to-face session with key stakeholders to share and discuss the findings of the analysis.
- Set directions and priories for further implementation of the data process.

The scan can be executed within 2-3 weeks.

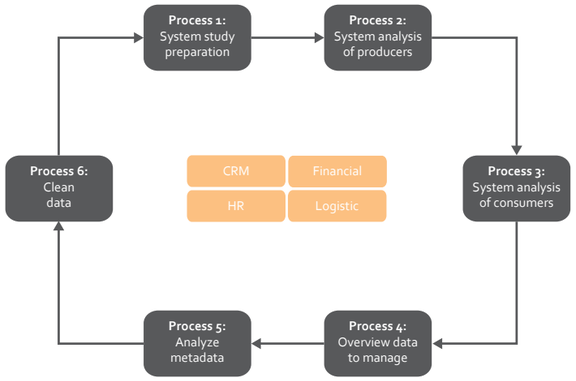

Figure 2: Overview of data scan process.

Results

- Structured insight into the data flow with its producers and consumers.
- Report to share the most impacted data issues and advice on how to handle the issues.

The data scan is entirely based on data derived from the organization. The results are validated with key stakeholders. The result is helping you to build a strong business case for improving your data management process.

1
S&OP DATA SCIENCE SCAN

1.2. Why is an S&OP data science scan needed?

Experience shows that a lot of issues are encountered that are related to data. These range from small problems such as data not being in the right format, to big problems that have a large impact on a project and could be a potential showstopper. Regardless of the problem or issue, each leads to discussions and longer project lead times.

The list below gives the most encountered issues that can occur:

- Duplicate codes with different descriptions
- Different codes with the same descriptions
- Missing mappings
- Different MDM[2] definitions in different systems
- Different interpretations of MDM definitions
- Receiving MDM from multiple parties and which needs to be linked
- Processes that are not standardized
- Processes that are not included
- MDM guidelines are not followed
- Duplicate records
- Timing issues: not all data sources are updated simultaneously with equal information, creating inconsistencies due to timing
- Different granularity (products, time, organization)
- Difference in organizational structure, multiple hierarchies: especially when logistics and finance/market data comes together - often the case in S&OP projects

These issues can also have consequences such as:

- Difficulties in obtaining regulatory compliance
- Reduced operational efficiency
- Reduced agility due to fragmentation
- Reduced profitability due to incorrect pricing
- Invoicing errors
- Inaccurate business reports
- Lack of decision-support

All of these issues lead to inconsistent data, which does not support the decision-making process. When people unknowingly take important decisions based on incorrect data, it will lead to bad decisions that could have a detrimental impact on the business. These issues require considerable time and effort to fix and should ideally be solved at the source systems.

Inconsistent or wrong data not only affect the decision-making process, it also impacts the day-to-day operations and can lead, for example, to stock being too high or low. And without the right data, an organization is unable to determine whether or not stock levels are correct. These are just a few examples of how bad data can adversely affect a company's reliability.

2 MDM: Master Data Management, this is a comprehensive method of enabling an enterprise to link all of its critical data to one file, called a master file, that provides a common point of reference.

2 HOW TO TACKLE DATA ISSUES?

In order to tackle the data issues there are three areas requiring attention, namely, the process, the data and the awareness aspect.

2.1. Awareness

One of the most important areas is the awareness of wrong/ incorrect data. In most cases business users know that the data is incorrect, but (higher) management has no idea that this is a real issue or threat. In order to increase awareness, a couple of actions can be taken that will increase the priority given to data science.

- Create a dataflow with the RACI (Responsible, Accountable, Consult, Inform) model
- Give management and other relevant departments a clear vision of data issues
- Be open in all data issues so that people will be receptive to possible solutions
- Explain the issues and risks to management in plain, non-technical language
- Explain the benefits of correct data in terms of resources (time & money).

2.2. Process

Step 1. Identify the sources of data
Different sources, such as ERP systems, data or business warehouses, Excel or other systems can lead to code, description and format inconsistencies of the data (e.g., different date notations).

The first step in the process is to identify and list all source systems to give an overview in the form of a matrix. The second step is to identify all main processes (e.g., HR, Inventory, Finance) and then define the sub-process for each main process. The source systems should be linked to the (sub)processes in order to show clearly which data originates from which system affecting the process. The data availability should also be prioritized to help show which processes will be affected, so that updating of data can be achieved in a stepwise manner, which is the preferred method.

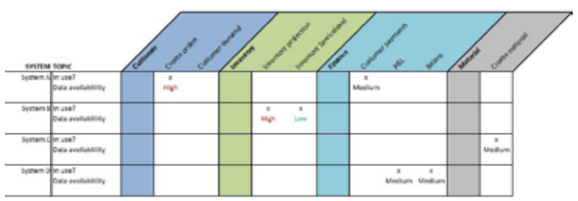

Figure 3: Matrix of processes connected to data sources.

Step 2. Decide which data to manage
When all the data is identified, it must be decided which data needs to be managed. In most cases five different types of data exist:

- Unstructured
  - E-mails, intranet, PDF files, video, speech, etc., without any real structure in the content.
- Transactional
  - This is in most cases related to sales, inventory, invoices, deliveries, etc.
- Metadata
  - Metadata such as report definitions, xml files, log files, configuration files.
- Hierarchical
  - Data storing relationships between other data, e.g., a product or a customer hierarchy.
- Master
  - Data that is critical to the business and which falls under four types:
    - › People (employee / customer)
    - › Things (product / service)
    - › Places (warehouse / divisions)
    - › Concepts (contract / licenses)

Deciding which data to manage can be achieved by describing the way it behaves and interacts with other data. In most cases master data involves transactional data, for instance, when a customer buys a product.

In this example, two types of master data are used, namely, People and Things, and it will result in a transaction such as an order or invoice. Master data can be further described using the CRUD (Create, Read, Update, Destroy) cycle[2], which determines how the master data is used within a given cycle.

	Product
Create	Product manufactured
Read	Periodic inventory catalogues
Update	Package, raw material changes
Destroy	End of Life
Search	System C, …

2 https://msdn.microsoft.com/en-us/library/bb190163.aspx.

2 HOW TO TACKLE DATA ISSUES?

Step 3. Identify producers and consumers of data
Producers and consumers (users) of data are often unknown but need to be identified. The producers are the users or systems that create new data in the different source systems. A similar matrix to the one above can be used to give an overview of which system or user creates data for which process. This exercise gives an overview of all the systems that contain relevant data and will identify unknown systems.

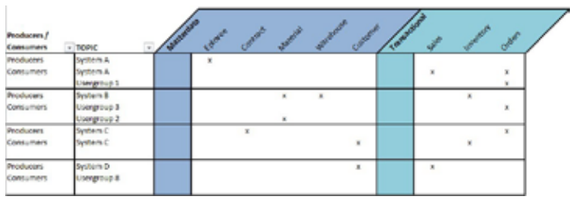

Figure 4: The producers and consumers of data.

Producer alignment is another mandatory step. The alignment must include agreement on data validations, formats and creation processes. This will help to increase the data quality and reliability of data because it must be both correct and consistent in all of the source systems. This alignment is stored in a service-level agreement.

Step 4. Collect and analyze metadata about the data
In order to make sense of the selected data, the entities (e.g. an object such as "Employee") and attributes (e.g. a salary, enlistment date of an employee), as well as the context (the use and meaning of the entities and attributes, and how they are used) must be determined. The data owner must also be identified, which is the most important and often the hardest thing to determine. The data owner is responsible for the quality and accuracy of the data and authorizes the creation and maintenance of the master data.

Step 5. Develop a data model
The data record will depend upon which attributes are used or included and what constraints exist regarding data type, size and/ or values. Also the mapping between the source system and data model must be defined and the way in which data is related must be described. This is a difficult task, but it is important not to make the model unnecessarily complex and therefore unuseful. All this information should be documented in an "interface agreement".

	Orders					
Name	System	Database	Database	Technical data type	Data type		comment
Order date	System A	x	xxxxxx	datetime	date		
Order policy	System A	x	xxxxxx	varchar(20)	text	20 characters	
Modality	System A	x	xxxxxx	varchar(20)	text		
OPID policy	System A	x	xxxxxx	varchar(20)	text		
Shipto country	System B	x	xxxxxx	varchar(200)	text	200 characters	
Shipto location	System B	x	xxxxxx	varchar(200)	text	200 characters	
Account manager	System B	x	xxxxxx	varchar(20)	text	20 characters	account manager code
Sector	System B	x	xxxxxx	varchar(20)	text	20 characters	
Customer	System B	x	xxxxxx	integer	number		customer code
Material group	System C	x	xxxxxx	varchar(20)	text		code of the material group
Material	System C	x	xxxxxx	integer	number		material code
supply chain	System C	x	xxxxxx	varchar(20)	text	20 characters	
Business Unit	System B	x	xxxxxx	varchar(20)	text	20 characters	
Booking date	System A	x	xxxxxx	datetime	date		
Invoice Date	System A	x	xxxxxx	datetime	date		
Handover Date	System A	x	xxxxxx	datetime	date		
Customer Currency	System B	x	xxxxxx	varchar(20)	text	20 characters	currency code (eg, EUR)
Exchange rate	System A	x	xxxxxx	numeric(10,2)	number	2 decimals	currency rate at that moment
Delivered price	System A	x	xxxxxx	numeric(10,2)	number	2 decimals	in currency of the customer
Delivered cost	System A	x	xxxxxx	numeric(10,2)	number	2 decimals	
Delivered gross margin	System A	x	xxxxxx	numeric(10,2)	number	2 decimals	
Delivered net margin	System A	x	xxxxxx	numeric(10,2)	number	2 decimals	
Volume	System A	x	xxxxxx	integer	number		
NSV	System A	x	xxxxxx	numeric(10,2)	number	2 decimals	
Order status	System A	x	xxxxxx	integer	number		status code

Figure 5: An example of an interface agreement on the transactional data orders.

Step 6. Cleaning and transforming the data
By now it is clear which data is master data and which is transactional data. The next chapter deals in greater depth with cleaning, merging and/ or transforming the data before it can be used in reports, etc.

2.3. Data

In the process part we looked at which data should be cleaned, merged and/ or transformed in order to create better quality and consistent data. This is achieved using the ETL (Extract, Transform and Load) method. The first step is to create a rollback scenario in order to return to the current state if required. To have the best manageable result it is recommended that the ETL is carried out in phases.

In the process described, the "system/ process matrix" can be used as input to prioritize the systems for phased planning.

An ETL method contains several steps:

1. Create rollback scenario
2. Build reference data
3. Extract the current data
4. Validate the current data
5. Transform the data (clean, apply business rules, check the integrity, aggregate/ disaggregate)
6. Store all the adjusted data into a staging system
7. Publish the data from the staging system back to the source

2 HOW TO TACKLE DATA ISSUES?

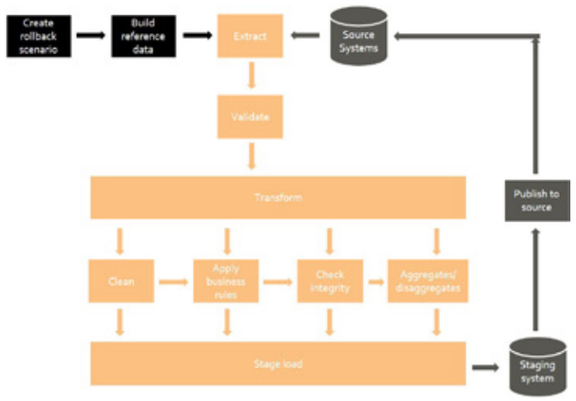

Figure 6: Process of data management.

Step 1. Create rollback scenario
An ETL process is complex and involves considerable data manipulation. In case of an emergency there must be an option to revert to the original state prior to starting the ETL process. Therefore a rollback scenario must be created. In general, these are manual backups of the source systems that are stored at a safe location and are not used in the regular backup schedule.

Step 2. Build reference data
Reference data is a data set that is used by other data fields, but the set has a permissible value. This means that there is a fixed list of values that can be used, for example units of measurement. This kind of reference data needs to be built in order to have system with the same definitions and/or restrictions. The data should be in line with the organization's standards and procedures.

Step 3. Extract
In this step, the appropriate data (which can be multiple tables or a part of the data in a table) is extracted from the source system(s) and is stored in a temporary location to perform the transformations. The temporary location can be a shared folder or a temporary database.

Step 4. Validate
The extracted data needs to be validated before the transformation starts. The validation includes such things as total records extracted, consistency checks, presence check, etc. When the extracted data is not validated, the process stops for the invalidated records. If wrong data is exported, the wrong data will be transformed and the wrong data will be imported back into the systems, which means garbage in = garbage out. Only records that are successfully validated will go to the next step. An outcome of this step is that there is an overview that lists all data within scope that needs to be adjusted in order to pass the validation.

Step 5. Transform
This step contains multiple sub-steps, such as cleaning the data, applying business rules, checking for data integrity and creating aggregates or disaggregates. In the data-cleaning step, all data errors, such as incorrect product descriptions, need to be analyzed and removed. When all data is corrected, business rules must be applied and the integrity of the data checked. Historic data is only used as a reference and can be aggregated in order to increase the performance on data sets that are too large. For example, 10 billion records that are only used for historic reference is too large, but it can depend on different criteria. This means that the data will not be transformed at the lowest level of detail possible, but at a higher level. For example, instead of having the historic revenue on an individual SKU item, it is aggregated to a product group level.

Step 6. Stage load
When all transformations are done, the data should be loaded into a staging database. This database will be a template of the original database in order to align the steps that need to be taken to load the data back into the source systems.

Step 7. Publish
The data that is validated, transformed, cleaned and updated is stored in the "stage load" database (see previous paragraph). The last step is to publish the corrected data back to the source systems in order to use the updated data.

3 THE BENEFITS OF GOOD DATA
4 KEEP CORRECT DATA

3. The benefits of good data

Good master data has many advantages. Also the process around the data will bring many advantages. When there is a robust data process in place, organizations can trust the data and convert into knowledge faster. One of the many benefits is that redundant data is eliminated because when master data is centrally located and maintained there is complete coherence and specifics regardless of which end-user or department uses the data. When the master data is clean, it is easier to carry out effective data analysis and gain a better understanding of the performance of the organization. Another benefit is that good master data makes it easier to embed new applications within the organization. It only requires them to be integrated with the master data application. The return on investment will also be shortened because the data integration will be smoother, the decision making process is faster and more effective, and the focus will be on analyzing the data instead of discussing the quality.

When the master data is not aligned, this has a detrimental effect throughout the whole organization. This is especially the case when there is no process in place to maintain alignment. For example, when each business unit is responsible and accountable for their local data but there is no synchronization between the business units. This creates a problem because it is difficult to create any reports higher in the organization due to the misalignment of the local data sets. Another disadvantage is the amount of room for discussion about how to interpret the data and what the source(s) are. Such discussions will have a negative effect on business decisions.

4. Keep correct data with monitoring/reports

After all of the above has been implemented, it must not be forgotten that master data is a continuous process and must be maintained with suitable monitoring/reports in place to avoid data problems.

The following metrics can be applied using KPI (key performance indicators):

- KPI name expressing what is being measured.
- Objective - why is this being measured? What is the impact when not okay?
- Dimension - what data quality (integrity, validity) is related to the KPI?
- Frequency of measurement - how often is the report refreshed?
- Unit of measurement and number of records, percentage, etc.
- The upper and lower threshold
- Who is responsible for the KPI
- Formula - how the KPI is calculated
- Assumptions

Possible KPI's:

- Availability – measure the systems accessed to provide the data for the day-to-day business
- Errors - measure the number of errors when the systems are accessed but the results contain errors or are stopped before giving the complete dataset
- Utilization
- Creation
- Measure how many new records are created and created by whom or what (system or person)
- Changing - measure how many records are changed and by whom or what (system or person)
- Accessing - measure how many records are accessed and by whom or what (system or person)

5 PHASED APPROACH
6 EXAMPLE

5. Phased approach

Implementing complete master data projects is a complex and large task. But partial implementation can also create benefits without having a large and complex project. For example, identifying the sources of the data can be achieved without total implementation. This makes it is easier to collect the right data from the systems that are necessary for daily business. Another example is identifying the producers and consumers of data, which makes it easier to push back incorrect data and request for it to be updated.

6. Example

An organization had a problem with organizing their master-data for reporting purposes. Creation of a standardized process enabled the correctness and completeness of the data to be trusted instead of second guessing every data download. Another advantage of the standardized process is that the data is always ready for use in a specific time period without a lot of manual adjustments.

Information is downloaded according a standardized process that is documented in an interface agreement. This describes which information is coming from which source system, as well as the datatypes of the fields (for example, date time, text, number, etc.) and also the parameters/filters for the dataset. Data download occurs according a specific time, which ensures that data is ready before this time.

When all data is downloaded from the source systems, this will be put into a "staging" database where data is validated according a defined rule-set. For example, when specific data has no values, a default value will be inserted where possible. But also when the data is validated and contains errors, it will be flagged for not creating an output for this individual record. The staging database also checks relationships between the individual files. For example, an inventory file that contains SKU numbers must be matched to the SKU numbers in the "master" file that contains the SKU hierarchy. When all of those rules are applied and there are any errors in the validation (e.g., no value or relationship found) this individual record will be flagged and it creates an error record.

The error records will create an error log file that is sent back to the organization. This file contains all error messages and helps the organization to update the data in their source systems to ensure that future master data is correct. When all records are validated against their respective values, the records are exported to a standardized format that is used for the reporting system.

ABOUT

CONTACT

For additional information on this research, please contact EyeOn.

Jan Veerman
Managing Partner
+31 6 51 88 47 01
jan.veerman@eyeon.nl

Jurgen Maas
Business Consultant
+31 628 91 13 44
jurgen.maas@eyeon.nl

ABOUT EYEON

In striving for success, large companies have to continuously struggle against growing internal complexity. We help our clients manage this complexity by designing, implementing and executing excellent planning processes as a discriminating factor for this success. In order to achieve this, we develop and share knowledge about top level planning and forecasting, with constantly demonstrable return on investment for our clients.

ABOUT INDUSTRY PLANNING AND FORECASTING KNOWLEDGE NETWORKS

EyeOn has many years of experience in setting up and improving planning processes at large multinational organizations in different industries. We actively share this knowledge. EyeOn has specific knowledge networks in High-Tech, FMCG, Process and Life Science. The networks allow you to share experiences and best practices concerning Planning and Forecasting with peer companies in your industry. Next to network events and benchmarking, EyeOn also organizes expert sessions and master classes in various specific domains of supply chain and financial planning

For more information: www.eyeon.nl.

EyeOn Solutions
Croylaan 14
5735 PC Aarle-Rixtel
The Netherlands
T +31 492 38 88 28
mail@eyeon.nl
www.solutions.eyeon.nl

—

EyeOn bv
Croylaan 14
5735 PC Aarle-Rixtel
The Netherlands
T +31 492 38 88 50
mail@eyeon.nl
www.eyeon.nl

—

EyeOn Amsterdam
Herengracht 124-128
1015 BT Amsterdam
The Netherlands
T +31 492 38 88 50
mail@eyeon.nl
www.eyeon.nl