



Collect, Transform, Generate and Test

MetaSuite and HP Quality Center Enterprise, generating Test Data from any data source from any platform, including mainframe



Table of contents

Executive Summary	3
MetaSuite	3
Problem Statement	4
1. Testing Cost	4
2. Storage Cost	4
3. Privacy-related issues.....	4
4. Reputation damage and direct business loss	4
Major Challenges for Preparing and Managing Test Data.....	5
Realistic data are hard to collect and sort	5
Correct test data are difficult to assemble and need high IT skills.....	5
Unmasked sensitive data put the business at risk.....	5
Re-using or adjusting test data is difficult to accomplish	5
Solution Description	6
Step 1 ► COLLECT	6
Step 2 ► TRANSFORM	6
Step 3 ► GENERATE.....	7
Step 4 ► TEST	7
Benefits	7
Use Case 1 – Test Data Wizard.....	8
1. Collect the source data	8
2. Sample	8
3. Transform and mask	9
4. Generate and run the program	9
5. Input for HP Quality Center Enterprise	10
Use Case 2 – Standard MetaSuite	11
1. Collect the source data	11
2. Sample	12
3. Transform and mask	13
4. Generate and run the program	13
5. Input for HP Quality Center Enterprise	13
Summary / Conclusion	13

MetaSuite

MetaSuite is a software solution to rapidly generate or extract, transform and deliver large and complex data volumes.

It is a powerful Data Integration solution that can be used for different objectives such as data acquisition for Business Intelligence and data warehousing, consolidation and delivery of master data in support of Master Data Management, data migrations and conversions, and Test Data Management.

Executive Summary

Uniquely, MetaSuite is capable of integrating **data from any source to any target on any platform** (e.g., Windows, a flavor of Unix or Linux, or a main-frame system like z/OS or BS2000). Using MetaSuite, organizations can fully capitalize on the wealth of data in all their business-critical solutions.

In this document, we will concentrate on the aspect of MetaSuite that complements HP Quality Center Enterprise, namely its rich and extensive **Test Data Management capabilities**, which are one of MetaSuite's major assets.

This powerful solution is designed to automate the generation or extraction of test data, offering powerful sampling and masking functions. Such test data can be generated based on real production data or using substitution data sets.

The main benefits of this solution include reduced development efforts and costs, enhanced reliability and security of your applications and compliancy with privacy regulations.

This White Paper targets all parties technically interested in MetaSuite, be it executives, technical managers, software architects, operations people or developers, and specifically everyone involved in Test Data Management.



Problem Statement

Before delivering an application to the market, it is essential to submit it to extensive testing.

Such testing processes must be cost-effective, compliant, reliable and secure, but they should also be repeatable in order to reduce the application development lead-time and costs.

Therefore, **Test Data Management is an essential step** in the development process of an application. Test Data Management addresses the following issues:



1. Testing Cost

When using a full copy of your production data for testing purposes, it is obvious that the testing process will consume a lot of computer time and resources, as well as an increased management cost for handling such large volumes of data. A smaller, representative sample of the data would be enough for your testing needs.

When working with production data for your testing activities, your company might lose time and money with every disruption of the day-to-day operational activities.

As generating test data is a recurring process, automating this process will not only speed up the process, but will also reduce the costs and the risks of error involved.

2. Storage Cost

Full copies of production data are not only time consuming, both in computer time and lead time, but they also require storage space.

Performing your tests on a representative sample of the data would reduce the required space and the time needed to copy and store the production data.

3. Privacy-related issues

The use of production data also involves privacy-related legal issues. All over the world, governments have privacy laws.

The USA have Sarbanes- Oxley for Data Protection & Integrity, PCI OSS for payment card data security and HIPAA for federal protections regarding personal health information. In Europe, we have the Data Protection Directive.

4. Reputation damage and direct business loss

Violation of the above-mentioned privacy laws can not only lead to huge fines, but also to reputation damage, which is another important reason for implementing test data management.

Major Challenges for Preparing and Managing Test Data

Realistic data are hard to collect and sort

Business application data are typically spread across the organization and stored in different ways. This can turn the extraction of data into a time-consuming and complex process.

Accessing and extracting test data usually takes a large portion of the testing effort. In addition, testing teams usually have limited skills for dealing with the complexity of all the data sources. They often have to rely on other people to retrieve the necessary data.

Correct test data are difficult to assemble and need high IT skills

To reduce the amount of test data, you need to make sure that your sample is representative for all of the production data. This means that you need to use statistically relevant and correct sampling formulas.

For your samples, you also have to make sure that you keep the correct links between, for example, customer data, related invoices and payments.

Unmasked sensitive data put the business at risk

Business information stored in the operational data hold a large number of sensitive personal and business information. Due to the privacy laws, such information cannot be shared for testing purposes.

Therefore, it is important that these data can be completely and irreversibly masked, so they can be shared with everyone in the testing environment, in or outside the company.

Re-using or adjusting test data is difficult to accomplish

Manually creating and maintaining test data is an expensive and time-consuming activity. That is why you should have a repeatable process that is able to refresh the test data by executing a pre-defined program against the actual production data or using substitution data sets.



Solution Description

MetaSuite provides a functionality for generating test data in a fast, smooth and secure way. First you collect or describe the definition of your test data and secondly, you create the test data itself.

MetaSuite provides two ways for generating those test data:

- 1. By using the standard MetaSuite functionalities.**
- 2. By using the MetaSuite Test Data Wizard.**

The Test Data Wizard generates test data on a one-to-one basis (one source is mapped to one target). In 80% of the cases this will be sufficient. For more complex generations of test data, whereby you want to add extra fields, combine multiple file definitions, etc., you have to use the standard MetaSuite functionalities.

In both cases, the principle is the same: the data are collected, the sampling and masking parameters are specified, the test data are generated and the output is made available to HP Quality Center Enterprise for testing purposes.

Step 1 ► COLLECT

The first step consists in establishing the correct definitions for your data (using MetaStore Manager).

The source data can be your own production data, but you can also use substitution files and tables such as lists of addresses or names. Data sources from any platform, including main-frame, can be used as input.

MetaSuite MetaStore Manager provides a simple “collect” function that will collect those definitions for you. Once collected, MetaSuite will present these definitions in an internal, easy-to-use, uniform format.

MetaSuite MetaMap Manager lets you specify the mapping rules, and the required sampling and masking parameters.



Step 2 ► TRANSFORM

When using the Test Data Wizard, these actions can be performed in a few simple clicks:

1. Select the data definition

The first step consists in specifying the data definition to be used for selecting the data.

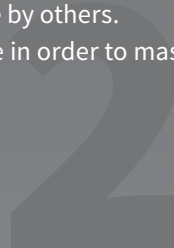
2. Select a sampling method

Once the data definition has been specified, you can select the sampling method, by using pre-defined sampling formulas.

3. Mask or anonymize the data

Using the same wizard, you have the following masking possibilities:

- Select a system-defined masking or anonymization routine.
- Create your own sampling formula as a user-defined function. Once defined, it can be made available for use by others.
- Use the content of a table in order to mask your data.



Step 3 ► GENERATE

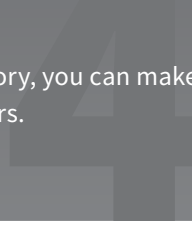
Once you have executed Step 1 (Collect) and Step 2 (Transform), you can generate a program that will run on the platform of your choice, be it Windows, a flavor of Unix or Linux or a mainframe system like z/OS or BS2000.



Step 4 ► TEST

Now you can execute the generated program. As a result, you will obtain a system overview report (files read, records excluded by the sampling process) and the final output data. If the result is not what you want, you can change the sampling routine or the masking and anonymization routines until you obtain the desired result.

Once the results are satisfactory, you can make the program available to others.



Benefits

The integration of MetaSuite with HP Quality Center Enterprise provides the following benefits:

- ✓ Reduced testing and storage costs
- ✓ Secured and anonymized test data
- ✓ Compliancy with privacy regulations
- ✓ Protection against reputation damage and direct business loss
- ✓ Automated sampling and masking process
- ✓ Accelerated application testing
- ✓ No disruption of the day-to-day operational activities
- ✓ Supports all input sources, including mainframe
- ✓ Outputs to a great variety of formats, including XML



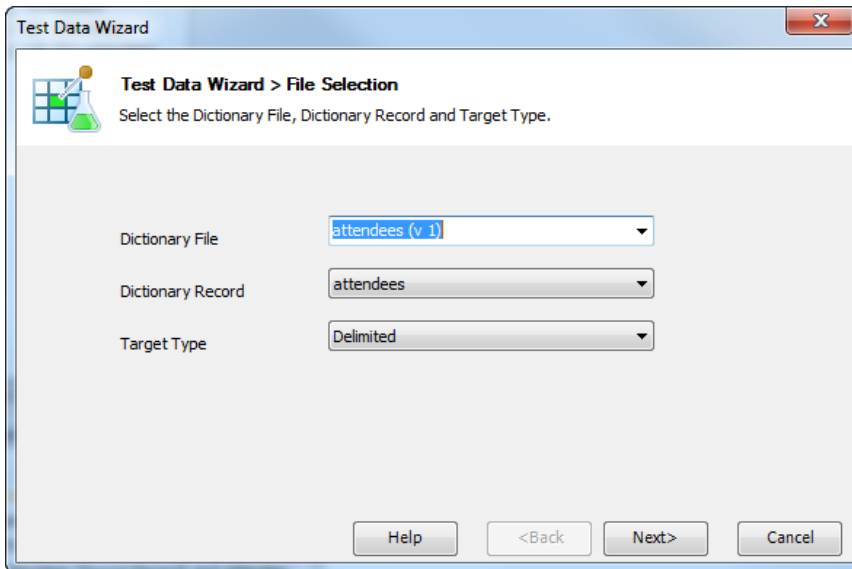
Use Case 1 – Test Data Wizard

In 80% of the cases, the Test Data Wizard will cover your needs for generating test data. For very complex generations of test data, you will have to execute the different steps using the standard MetaSuite components (see Use Case 2 – Standard MetaSuite).

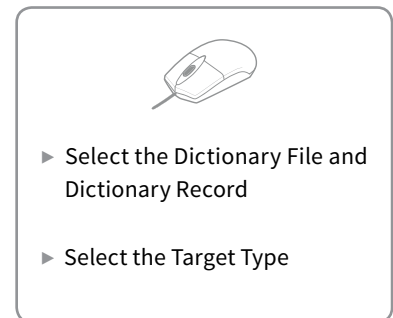
1. Collect the source data

Collect the correct definitions for your source data in the required format.

The source data can be your own production data, but you can also use substitution files and tables such as lists of employees, addresses, names. Data sources from any platform, including mainframe, can be used as input.

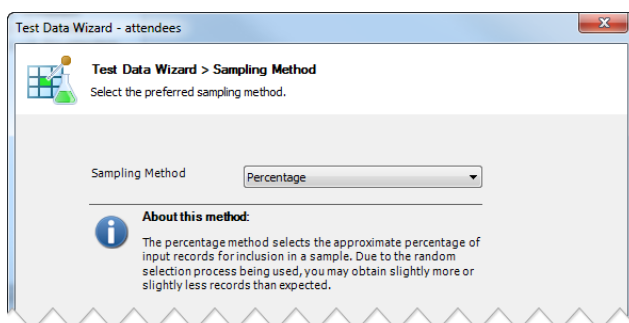


The screenshot shows the 'Test Data Wizard' window at the 'File Selection' step. The title bar reads 'Test Data Wizard'. The main area contains the text 'Test Data Wizard > File Selection' and 'Select the Dictionary File, Dictionary Record and Target Type.' Below this are three dropdown menus: 'Dictionary File' with 'attendees (v 1)' selected, 'Dictionary Record' with 'attendees' selected, and 'Target Type' with 'Delimited' selected. At the bottom are buttons for 'Help', '<Back', 'Next>', and 'Cancel'.

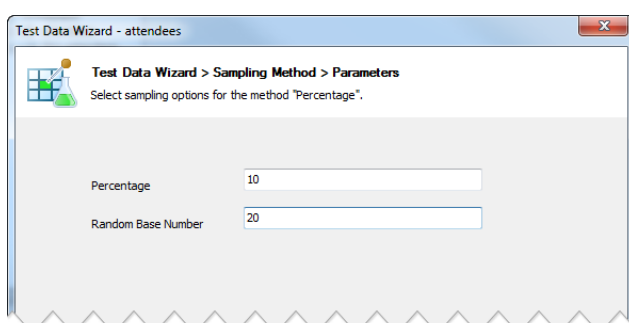
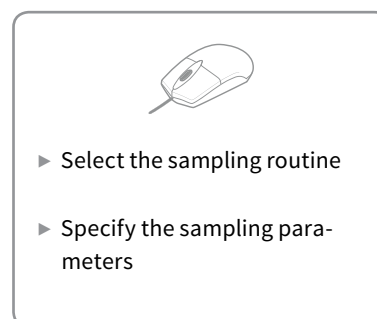


2. Sample

Extract the sample file by selecting one of the pre-defined sample routines.




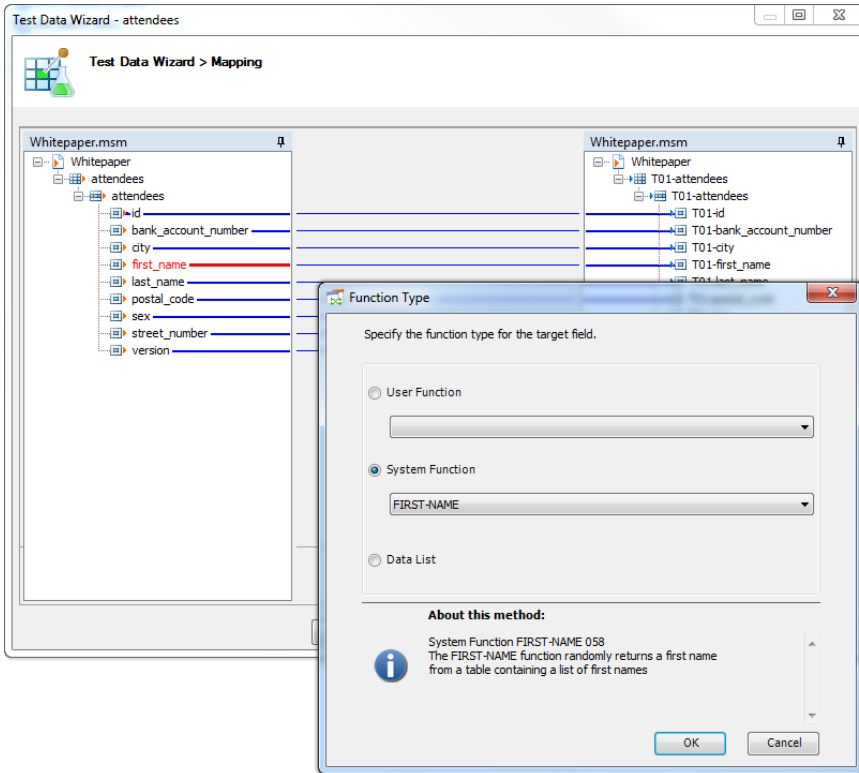
The screenshot shows the 'Test Data Wizard - attendees' window at the 'Sampling Method' step. The title bar reads 'Test Data Wizard - attendees'. The main area contains the text 'Test Data Wizard > Sampling Method' and 'Select the preferred sampling method.' Below this is a dropdown menu for 'Sampling Method' with 'Percentage' selected. An information icon is followed by the text 'About this method: The percentage method selects the approximate percentage of input records for inclusion in a sample. Due to the random selection process being used, you may obtain slightly more or slightly less records than expected.'



The screenshot shows the 'Test Data Wizard - attendees' window at the 'Sampling Method > Parameters' step. The title bar reads 'Test Data Wizard - attendees'. The main area contains the text 'Test Data Wizard > Sampling Method > Parameters' and 'Select sampling options for the method "Percentage".' Below this are two input fields: 'Percentage' with the value '10' and 'Random Base Number' with the value '20'.

3. Transform and mask


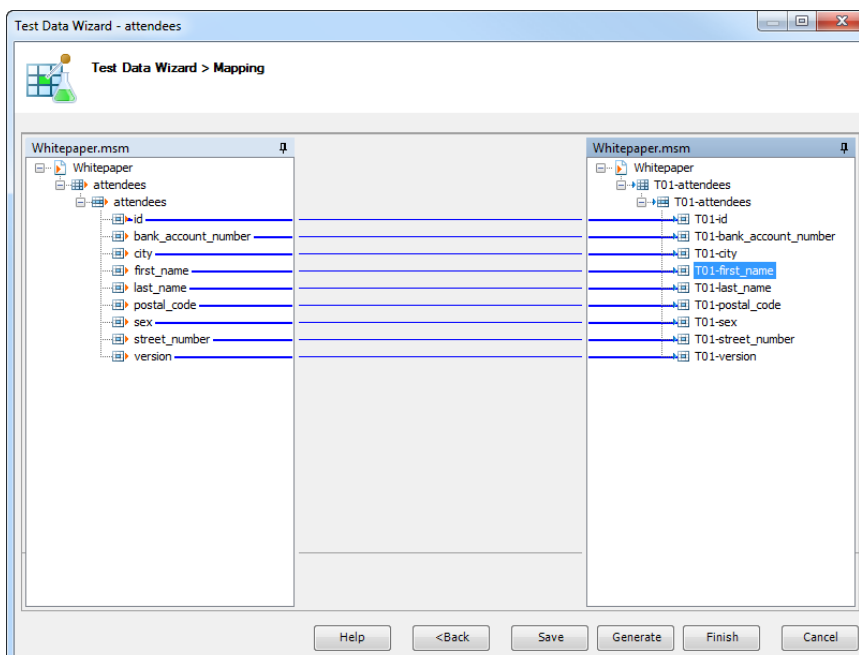
Apply the required masking or anonymization routine.



- ▶ Double-click the target field you want to mask or anonymize
- ▶ Select the required function type

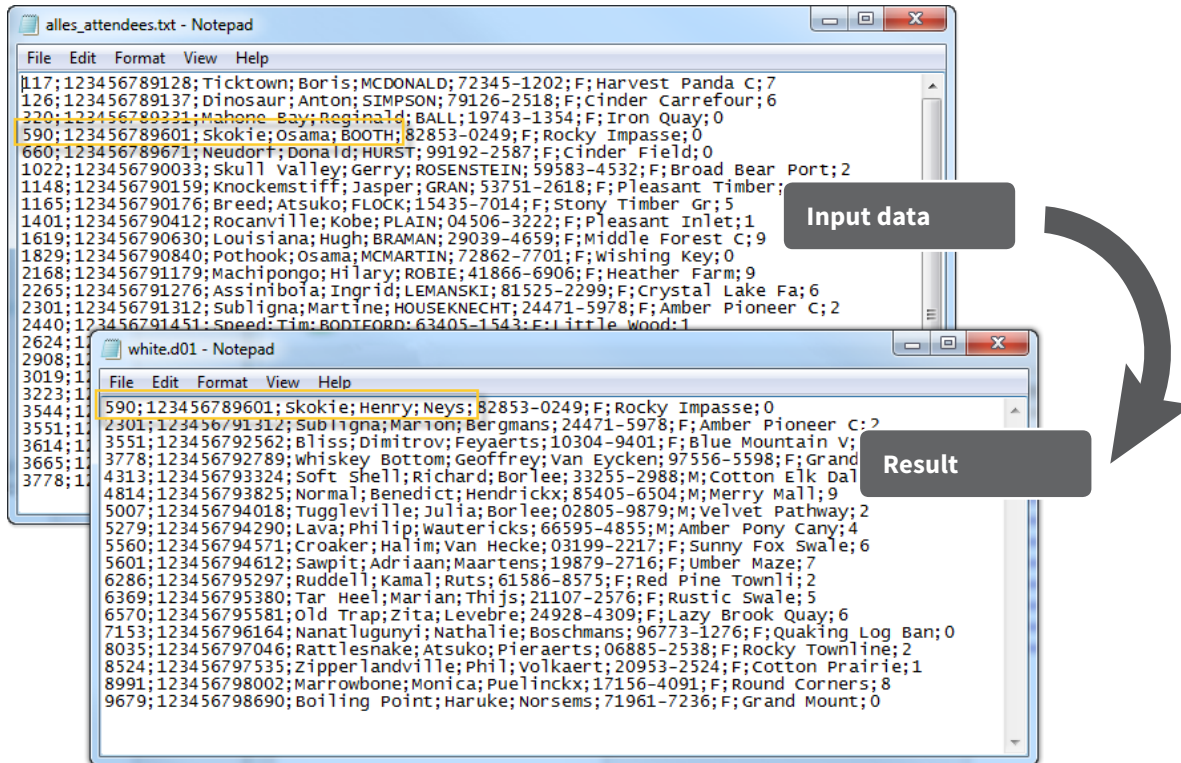
4. Generate and run the program

The program is now ready to be generated.



Click the Generate button to generate and to run the program in order to obtain the final output data.

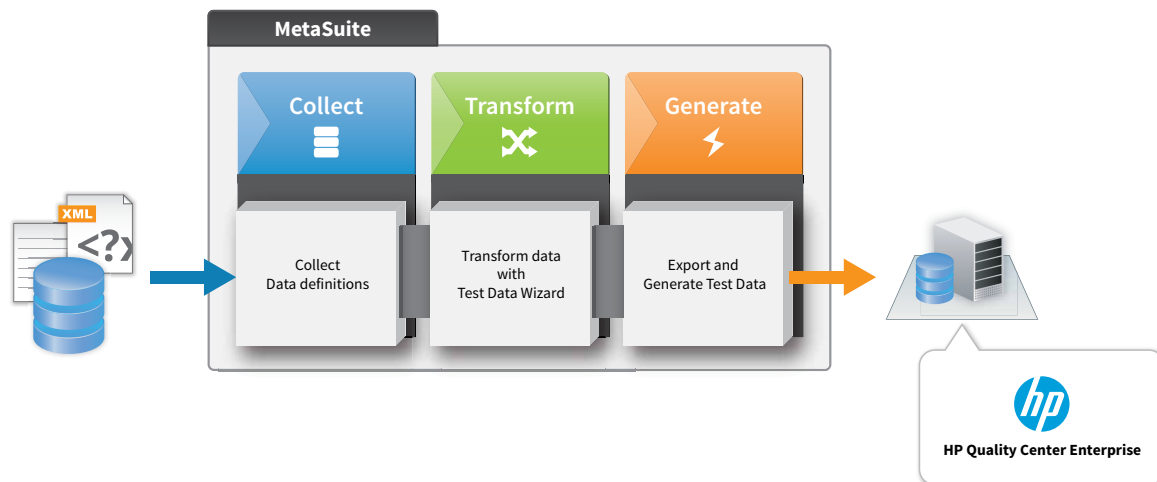
The result:



Notice in the example above that the first and last name have been replaced by the system functions “FIRST-NAME” and “SUR-NAME”, i.e., Osama;BOOTH has become Henry;Neys.

5. Input for HP Quality Center Enterprise

Use the test data in HP Quality Center. The test data generated by MetaSuite can be loaded into the database used by HP Quality Center Enterprise to run the tests. This process is repeatable and allows you to refresh test data with actual values.



Use Case 2 – Standard MetaSuite

For more complex test data, you cannot make use of the Test Data Wizard. For example, if the target file does not have the same structure as the source file, or if many input files have to be joined together, ... In such cases, you need to use the standard MetaSuite functionalities to obtain the test data.

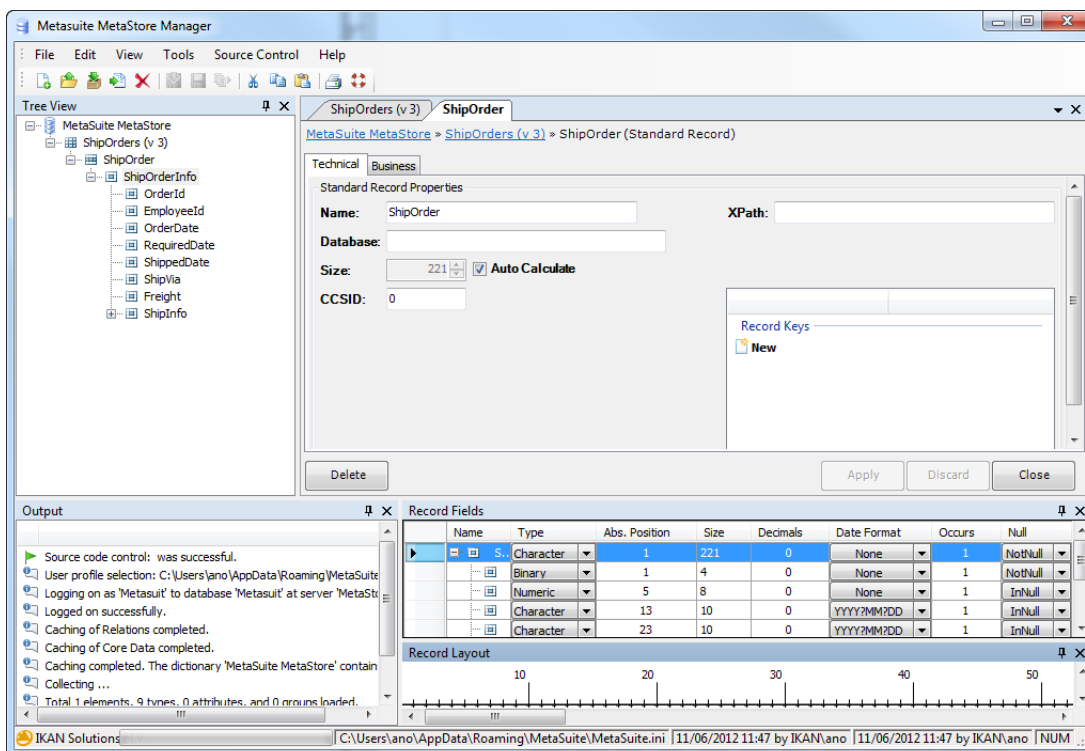
MetaSuite contains two basic components: MetaStore Manager and MetaMap Manager.

- **MetaStore Manager** is used in order to maintain all data definitions, also called metadata.
- **MetaMap Manager** is used in order to create models: logical relationships between the data definitions which can be used both as source or target file. He can also add procedures. Models can be transformed into executable programs by means of the Generate functionality.

The following sections describe the steps in more detail:

1. Collect the source data

MetaStore allows you to collect the data definitions of the source and target structures you want to work on. After collecting the data definitions in the original format, MetaSuite translates these definitions in an internal, uniform format.



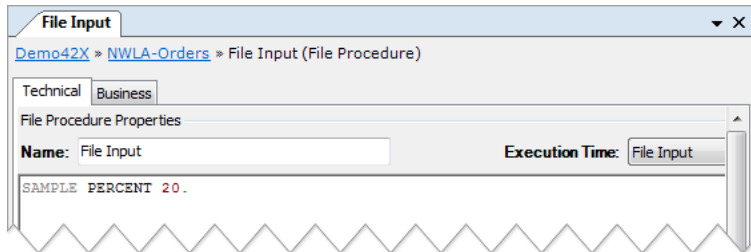
▲ Example of the MetaStore Collect of an XML file.

MetaStore provides a hierarchical view of your data files or database tables and gives you full access to the details of your data. Once you have the data definitions, you can use MetaMap to develop your program and specify the sampling and masking parameters.

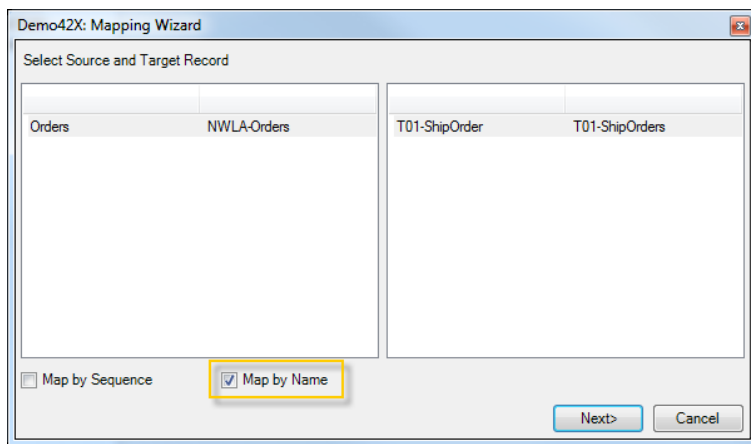
2. Sample

MetaMap, the second component, allows you to define the subset (sampling) and to mask or anonymize the data. By using the look-a-head-parser, you can select one of the predefined sampling routines.

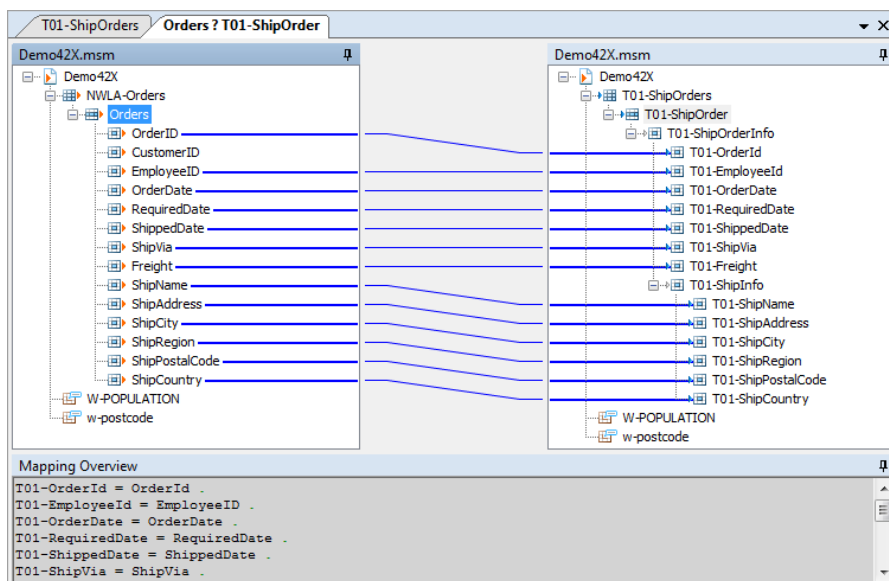
In this example, we take a sample that represents 20% of the source data, chosen at random. Note that several other, more complex, sampling routines are available or that you could easily create your own sampling routine, which later on can be made available to others by defining it as a user-defined routine.



Once you have defined the sample, you just need to map the sample to your target. MetaSuite disposes of an easy-to-use Mapping Wizard to do this. The Mapping Wizard maps by field sequence or by name.



The result:



3. Transform and mask

In the next step, you can mask the data. The same simple principle applies: you use one of MetaSuite's pre-defined masking routines or your own user-defined sampling routines.

In the example below, the numeric OrderId field will be replaced with a random number between 123 and 1234:



4. Generate and run the program

Once all definitions have been selected and all parameters have been defined, you generate an executable program. This program can run on a range of platforms including mainframe, BS2000, Unix, Windows, VMS and I-series. More information on how MetaSuite helps with generating test data can be found in the standard user documentation on our product website: www.metasuite.be/infocenter.

5. Input for HP Quality Center Enterprise

Use the test data in HP Quality Center. The test data generated by MetaSuite can be loaded into the database used by HP Quality Center Enterprise to run the tests.

This process is repeatable and allows you to refresh test data with actual values.

Summary / Conclusion

MetaSuite complements the HP Quality Center Enterprise by providing a test data generation tool which allows creating a representative sample that can be used to effectively test your application in a realistic business-like environment using masked data.

It supports all kinds of sources, including mainframe, and outputs to a variety of formats.

This highly increases the quality of the application testing process while reducing the costs and speeding up the time-to-market.

For More Information

To know more, visit <http://www.metasuite.com>. Contact IKAN Solutions: info@metasuite.com

IKAN Solutions N.V.
Kardinaal Mercierplein 2
2800 Mechelen
Tel. +32 15 797306
info@ikan.be
www.ikan.be

© Copyright 2013 IKAN Solutions N.V.

The IKAN Solutions and MetaSuite logos and names and all other IKAN product or service names are trademarks of IKAN Solutions N.V. All other trademarks are property of their respective owners. No part of this document may be reproduced or transmitted in any form or by any means, electronically or mechanically, for any purpose, without the express written permission of IKAN Solutions N.V.

IKAN