

# STeP-IN SUMMIT 2007

## International Conference On Software Testing

### Test Data Management For Enterprise Applications

by  
Naresh Malepaty and Pavan Kumar M.P  
**MindTree Consulting Pvt. Ltd.**

[Naresh\\_malepaty@mindtree.com](mailto:Naresh_malepaty@mindtree.com), [Pavankumar\\_mp@mindtree.com](mailto:Pavankumar_mp@mindtree.com)

Copyright: STeP-IN Forum and Quality Solutions for Information Technology Pvt. Ltd.

Published with permission for restricted use in STeP-IN SUMMIT 2007 in agreement with full copyrights from owner(s) / author(s) of material. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior consent of the owner(s) / author(s). This edition is manufactured in India and is authorized for distribution only during STeP-IN SUMMIT 2007 as per the applicable conditions.

Practices Experience Knowledge Automation

Produced By

**STeP-IN**  
Forum

[www.stepinforum.org](http://www.stepinforum.org)

Hosted By



[www.gsitglobal.com](http://www.gsitglobal.com)

MindTree Consulting Pvt. Ltd.

Test Data Management For Enterprise Applications

White Paper

Authors:

Naresh Malepaty.

Email - [Naresh\\_malepaty@mindtree.com](mailto:Naresh_malepaty@mindtree.com)

Pavan Kumar M.P

Email - [Pavankumar\\_mp@mindtree.com](mailto:Pavankumar_mp@mindtree.com)



MindTree Consulting Private Limited,  
# 42, 27<sup>th</sup> Cross, 11<sup>nd</sup> Stage  
Banashankari  
Bangalore-560 070  
[www.mindtree.com](http://www.mindtree.com)

## Abstract:

Areas that often get overlooked in testing Enterprise application are “Preparation and Management of test Data” and “Appropriate review of test Data”.

Due to the complexity involved in preparation and maintenance of test data, critical defects are identified only at a stage where it is complex and expensive to fix leading to slippages in schedules. The principles and guidelines described in this paper will address the issue and helps to tackle it.

This article delineates best practices that can help you get most out of your Enterprise Application and will also ensure applications robustness and resilience. Test Engineers as well as managers can take the benefit out of it for testing any enterprise applications to prepare and maintain data.

Features addressed in this article are principles of test data management from the perspective of a real time enterprise application, Test Data Classification - Handling seed data and workflow Data, Structured and organized storage of test data and test results, Data acquisition and control for test beds and easy access for future analysis and reporting. This can be applied in Enterprise applications where data is huge and difficult to maintain and data being in different states at different instances or data changing states based on an occurrence of an event.

Defining procedures for access and maintenance of data as addressed in this article and usage of controlled environment will result in success of test data management.

---

## Table of Contents

---

<b>ABSTRACT:</b> .....	<b>1</b>
<b>1.0 INTRODUCTION:</b> .....	<b>3</b>
<b>2.0 PRINCIPLES OF TEST DATA MANAGEMENT:</b> .....	<b>3</b>
2.1 DATA PRINCIPLES .....	3
<b>3.0 TEST DATA CLASSIFICATION:</b> .....	<b>4</b>
<b>4.0 CHALLENGES IN TEST DATA MANAGEMENT:</b> .....	<b>4</b>
<b>5.0 BEST PRACTICES:</b> .....	<b>4</b>
5.1 TEST DATA MATRIX. ....	4
5.2 SEED DATA AND PROCESSED DATA .....	7
<b>6.0 FASTER AND BETTER REVIEWS</b> .....	<b>8</b>
<b>7.0 CONCLUSION</b> .....	<b>8</b>
<b>8.0 DEFINITIONS, ABBREVIATION AND ACRONYMS</b> .....	<b>8</b>
<b>9.0 REFERENCES</b> .....	<b>9</b>
<b>10.0 ACKNOWLEDGEMENTS</b> .....	<b>9</b>
<b>BIOGRAPHY OF THE AUTHORS</b> .....	<b>10</b>

---

## 1.0 Introduction:

This article delineates best practices that can help you get the most out of your Enterprise Applications and will also ensure applications robustness and resilience.

## 2.0 Principles of Test Data Management:

Principles are general rules and guidelines, intended to be enduring and seldom amended that in form and support the way in which an organization sets about fulfilling its mission.

Principles may be established at any or all of three levels:

**Enterprise principles** provide a basis for decision making throughout an enterprise, and inform how the organization sets about fulfilling its mission. They are a key element in a successful strategy.

**Information Technology (IT) principles** provide guidance on the use and deployment of all IT resources and assets across the enterprise. They are developed in order to make the information environment as productive and cost-effective as possible.

**Architecture principles** are a subset of IT Principles that relate to Architecture work. They reflect a level of consensus across the enterprise, and embody the spirit and thinking the enterprise architecture.

There are five criteria that distinguish a good set of principles:

- **Understandable:** The underlying tenets can be quickly grasped and understood by individuals throughout the organization. The intention of the principle is clear and unambiguous, so that violations, whether intentional or not, are minimized.
- **Robust:** Enable good quality decisions about architectures and plans to be made, and enforceable policies and standards to be created. Each principle should be sufficiently definitive and precise to support consistent decision making in complex, potentially controversial, situations.
- **Complete:** Every potentially important principle governing the management of information and technology for the organization is defined. The principles cover every situation perceived.
- **Consistent:** Strict adherence to one principle may require a loose interpretation of another principle. The set of principles must be expressed in a way that allows a balance of interpretations. Principles should not be contradictory to the point where adhering to one principle would violate the spirit of another. Every word in a principle statement should be carefully chosen to allow consistent yet flexible interpretation.
- **Stable:** Principles should be enduring, yet able to accommodate changes. An amendment process should be established for adding, removing, or altering principles after they are ratified initially.

### 2.1 Data Principles

One of the prime assets to an enterprise is "*Data*" and this has a value to the enterprise and needs to be managed accordingly.

The factors that needs to be taken care when managing data are, it needs to be

- Accurate / Truthful
- Judicious

For the data to be used proficiently, its user's need to be meticulously trained on its value, define rules for proper sharing and accessing.

The data that is captured in that sort helps in decision making. Individuals that are identified as "Data-Stewards" must have the authority and means to manage the data. A cultural transition is the need of the hour in our thinking from data ownership to data stewardship. Proper care needs to be taken to ensure that the data is not obsolete, incorrect, or inconsistent as this could be passed to Enterprise personnel which would adversely affect decisions across the Enterprise.

### 3.0 Test Data Classification:

Accurate data classification is perhaps the most important and challenging part of data management. The better the classification, the more effective are their results. Data needs to be "Supreme" --- Supreme data means that the value of each sample unit is observed or known. The accuracy of any prediction is directly proportional to the quality and accuracy and completeness of the supplied data. Supreme data, along with the appropriate model choice, usually results in quality predictions. Bad or deficient data will almost always result in dreadful predictions. Data classification assigns a level of business importance, availability, sensitivity, security and regulation requirements to data. The data needs to be grouped under the categories - Mission-critical online data, Business-critical online data, Accessible online data, near line data and Offline data.

### 4.0 Challenges in Test Data Management:

The prime challenges that are faced in managing test data are:

- Covering all the business conditions and validations.
- Maintenance of test data for further reuse.
- Handling of huge amount of data.
- Identifying the Seed Data, Transactional Data and Processed Data.
- Usage of planned data in an appropriate test environment.
- Minimizing the amount of data capture from the user while test execution.



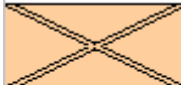
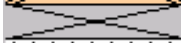

### 5.0 Best Practices:

#### 5.1 Test Data Matrix.

Capture Business Conditions and Business Validations in a matrix format. Plotting conditions in data matrix (Conditions Vs Validations):

- All possible conditions on Y-Axis and
- All possible Validations on X-Axis.

For every co-ordinate in the X-Y Axis, we can have the following:

Color Code	Description
	Valid Critical Business case and will be tested. Regression testing will also be done.
	Valid Business case and will be tested.
	Valid Business case but will not be tested. ( Decision to be taken by stake holder).
	Condition not applicable in business.
	Invalid data

That sort-of color coding is chosen as it helps even when one takes a print-out using a B/W printer.

Based on this, Appropriate Test Data needs to be prepared.

Test Data is derived from conditions. (Data should be defined for all the dependant modules and cover all the test conditions).

The prepared Data should be used when a plan is drafted for Automation.

The conditions that need to be met for choosing this Matrix approach are:

- Screens or reports which are "process/logic oriented"
- Minimal data capture from the user
- Largely depends on the data that comes from another module/modules
- Performs action/actions on the data that comes from other module/modules.

Requirement Traceability - To be maintained between the Basic/Alternate flows in Requirements and the Test Conditions.

For every data that is captured, which is intended to be passed as input, needs to have an appropriate output (This details are captured in another sheet and a reference pointer in the form of URL will be keyed-in in the appropriate X-Y Co-Ordinate).

It is good to execute the scenarios that are captured in this test Matrix in a Controlled Environment as we can have total control on the data that is used.

Controlled environment will mainly be using planned data (pre-determined) - This is done in a way expected output is easily determined.

A snap shot of a module where this was used:

S r n	Shipments	Pre Conditions							Special ly Loaded	Pre Condition B/D Compati bly break code	EE B/D "NOT pre cond	EE/P B/D "NOT pre cond	Mixed Loaded			Check (records Permissible - Intra)						
		1. EE B/D , present in B/D Master.	2. EE/aux-EE B/D , present in Stock Master.	3. Not ul within permissible limit.	4. Not Mixed Loaded.	5. B/D Type associated in associated with the aircraft	Emply.	Loaded with hooked shipment					Loaded with lying list shipment	Loaded with hooked shipment	Loaded with lying list shipment	BUP/ Through	Emply.	Loaded with hooked shipment	Loaded with lying list shipment	BUP/ Through	Loaded with PRE/ MAM shipment.	ULD already loaded with Split shipment
		U1-U3	U4-U5	U6-U7	U8	U9-U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20	U21	U22	U23	U24	U25	
1	"Not Ready" Shipments / W/D incomplete																					
2	Loaded Shipments with "UL" status.																					
3	Loaded Shipments with "HM" status on final leg.																					
4	Loaded Shipments with "HM" status on subsequent legs with Airpact NOT allowing dumping.																					
5	Loaded Shipments with "HM" status on subsequent legs with Airpact allowing																					
6	Loaded Shipments with "CS" status and "RCS" Mal status																					
7	Loaded Shipments with "UL" status and "RCS" Mal status																					
8	Lying Shipments with "HM" status and "RCS" Mal status																					
9	Loaded Shipments already loaded in ULD																					
10	Loaded Shipments with "PRE" status on same flight																					
11	Loaded Shipments with "MAM" status on same flight.																					
12	Loaded Shipments with "DEP" status																					
13	Part of a shipment (No of pieces, ul and not less than shipment Pieces, ul and not) with "RCS" status.																					
14	Part of a shipment (No of pieces, ul and not less than shipment Pieces, ul and not) with "PRE/MAM/DEP" status and loaded only																					
15	Split shipment in BULK																					

In the above snap shot, Validations and Conditions are plotted on X-Y Matrix. The critical Business cases have a background of Pink and valid business cases have a background of white. Both of them have an URL link which points to the sheet which has the input and output data.

## **5.2 Seed data and processed data**

(a.) Data that would be used in any enterprise needs to be classified into Seed Data, Transactional Data and Processed Data.

Seed Data - The basic data that would be required for any application. In the case of an e-mail user registration,

- City, State and Country would be the basic seed data.
- Industry, Title and Specialization would be the customized seed data

Transactional Data

- First Name, Last Name and Zip/Postal Code would be the transactional data.

Processed Data - Data that is required to validate the functionality of the application under test.

- Users data with their customized options

(b.) Two separate test environments need to be created:

**Adhoc Environment** - Here, one is free to perform all adhoc operations without giving any emphasis on data. Basically this is used to check the field validations that would exist in an application.

**Controlled Environment** - Here, the data is controlled (Only planned test records will be made available). Here in, in place of the seed data being exhaustive we can only have limited (or) required data.

(c.) Population of data into Controlled Environment:

- Appropriate scripts needs to be created to populate seed data into the database
- Automated scripts need to be made to make transactional data using seed data and required parameters (These are stored in CSV files)
- Automated scripts need to be developed which will use the transactional data and migrate it to processed data.

(d.) Maintenance of test data in .CSV files

- Firstly, the data that would be required needs to be stored in excel (xls) file
- One sheet should have all the formulae that would transform the base data into new form.
- This formulae sheet is used to get the new set of data for the next cycle of testing.
- All the xls sheets can be converted to csv using freely downloadable converters.

(e.) Some tips on formulae sheet

- All the dates captured in the data should have a reference with the reference date in the formulae sheet. This helps in updating all the required date fields in the data sheet by changing the reference date.
- Usage of Fill Series option (When creating the data for the first time)
- Use numbers in the suffix (Like User1, User2....).
- Need to be done only for mandatory fields and some optional fields (if necessary).

(f.) Effective use of Pivot table and filter option in excel.

- This helps in all the search conditions for the application
- For any particular condition, the number of records that would come out on a search should be equal to the number of records in the excel sheet.

## 6.0 Faster and better reviews

Test Data review has to be thorough in order to ensure that effective and adequate testing is done. Often, this review is ignored owing to time constraints and other factors.

(a.) Checks that needs to be done while reviewing the test data:

- Understanding of the requirement is correct.
- The data is correct and represent every possible condition.
- Seed data, transactional data and processed data are identified perfectly and documented precisely.
- The test data covers adequate test scenarios.

(b.) Suggested Methodology for Test data Review

During the Requirements Understanding phase, review of requirements is an activity that should be undertaken with utmost care and such review should be done systematically to ensure the clarity, correctness and testability of the requirements.

During the Test preparation phase, identification of scenarios for which test data needs to be prepared should be done. Here are a few suggestions for preparation of test data:

1. Adopt/Create a template for requirement and ensure that this template is used by all the testers in the team. Based on the needs of the project, additional items can be added to the template when ever required or when new requirements are added.
2. Educate the team about the test data gathering and using various methodologies.
3. Educate the team about the use of data matrix where ever it can be applied.

When the above suggestions are implemented, it would be easy to review during the Test Preparation phase.

## 7.0 Conclusion

By knowing the value data has, in an enterprise application, getting the personal trained on its importance, ways of capturing, defining procedures for access and maintenance of data and usage of Controlled Environment, one can completely maintain the data successfully.

## 8.0 Definitions, Abbreviation and Acronyms

Acronym	Description
Data-Stewards	Accountable to manage data; Ensures data quality; Drives the forum on process changes

## 9.0 References

{List down all the sources - web sites/ books/ other papers from which information has been used in the paper}

Item	Description

## 10.0 Acknowledgements

{Acknowledge person(s) who have provided inputs/ review feedbacks or helped in writing the paper}

Name	Description

## Biography of the authors

### Author 1:



**Naresh Malepaty** carries a total experience of 11 years, in which 9+ years is in software testing. He is experienced in test management, automation and performance. He is experienced in design / development of automation frameworks / test suites using automation tools like WinRunner, SilkTest and QARun / Performance tools like Silk Performer and Load Runner. He has worked on different domains like Supply Chain Management, Air Cargo Logistics, Travel and Leisure, US Law and US Forestry. He is currently associated with MindTree Consulting Private Limited, Bangalore.

### Author 2:



**Pavan Kumar M.P** has 2 years of experience in software testing that includes testing of Web & Client Server applications. He has worked on different domains like Air Cargo Logistics, SCM and online shopping. He has experience on tools like QARun. He is currently associated with MindTree Consulting Private Limited, Bangalore.