Test Data Management Bottlenecks

Erik Haahr

Agenda

1. Introduction

2. Trends affecting Test Data Management

3. Dimensions in Test Data Management

4. Decisions belonging to Test Data Management

5. The first draft framework for a Test Data Management Maturity Model revealed
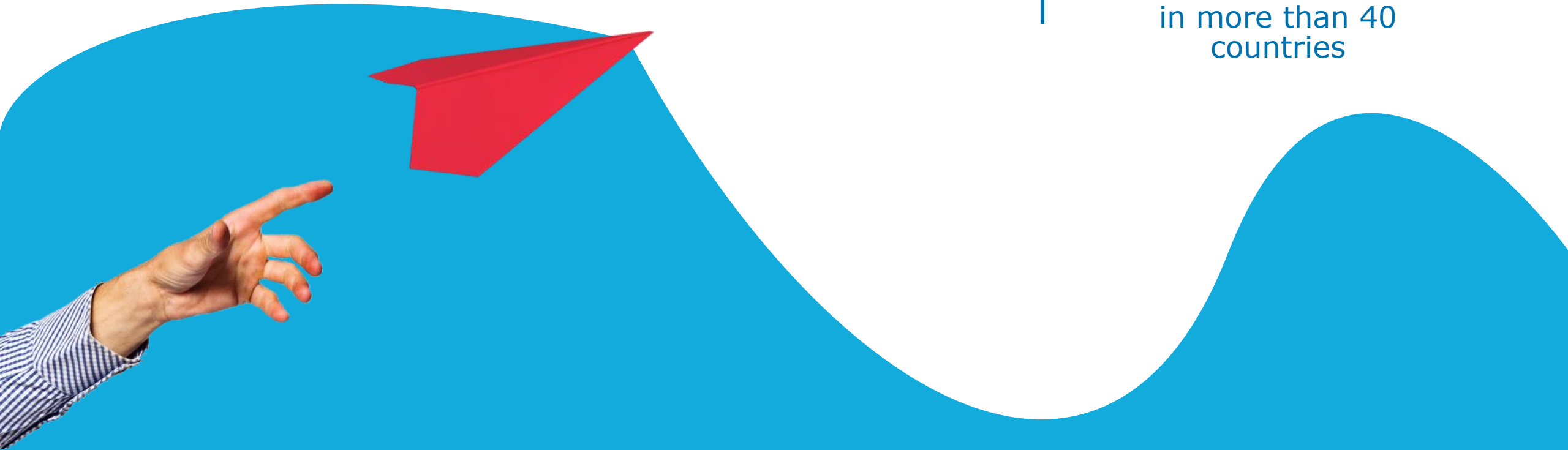
From a start-up ...to a global leader

1967

7 people in Grenoble, France

2019

Over 200,000 employees of 120+ nationalities in more than 40 countries

Who am I…

- **Education:** Electronics and Software Engineer
  Profession: Enterprise Architect

- Experience with **Test Data Management** as well as **Test Automation** in the **Pharma Industry**

- **General Data Management** Experience from working with Big Data and IoT

- Personal: **Married, Trail runner, Grand children**

- Strangest food: **Insects**

Trends affecting Test Data Management

Trends affecting Test Data Management

Changing development paradigms (DevOps)

Trends affecting Test Data Management

Increasing data regulations (GDPR)

Systems decreasing in size and scope (microservices) with individual life-cycles

Trends affecting Test Data Management

Cloud computing

Trends affecting Test Data Management

Test Data Management (TDM) needs to change and adapt

Test Data Management Issues

Test Data Management Issues

Dimensions

- Time
- Security
- Provisioning
- Size
- Consistency
- Agile

Decisions

- Multiple identical test databases
  - Can I reuse an existing Test-DB?
  - Must I create a new instance?
- Test data architecture
  - How can I provision a set of test-DBs for my "project"?
  - Can I use synthetic test data?
  - How can I pseudonymise/mask my test data?
- Production data (GDPR etc.)
  - How can I use test data and comply with GDPR?
- Agile (DevOps/DevSecOps/DataDevOps…)
  - How can I frequently get a fresh data set?
- Sizing (small/big vs. representative)
  - How do I decide the size of data set and select the content?

Time Issues

- Databases from multiple systems
  - Must be derived from production at the same point in time
  - OR
  - Be generated with the same date interval in mind

- Time to restore test environment
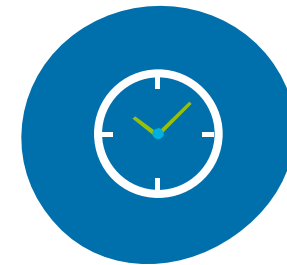
Remedies will be presented under other test data issues.

Security Issues

Relates primarily to data (derived) from production data – not to synthetic data!

- Access Rigths
  - Who can see real and sensitive data in the test environment

- GDPR
  - You will never get consent to use person identifiable data (PID) for the purpose of testing!

  - You will need to identify all attributes that could identify a person even in combination
    - A group of attributes that together identify less than five persons will be considered PID

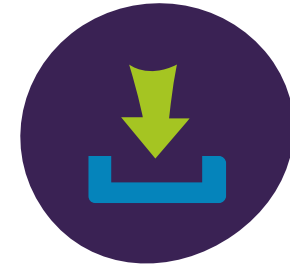- Company confidential data should be treated in the same way as PID!

Want: Mask/encrypt all sensitive data in a way that maintain relations across data sources

Provisioning Issues

- Multiple projects sharing one test database
- Handling multiple copies of the same test database
- Test data from systems with different lifecycles

Want: Possibility to recreate identical test data sets for automatic regression testing simultaneous with additional test data for testing new or modified functionality that is added to original when release is approved

Sizing Issues

- Multiple identical test databases
  - Multiple "projects" in different cycles need their own instance
- Sizing (small/big vs. representative)
  - System Type: Data Entry => Small (synthetic?) data set
  - System Type: Customer Segmentation => Big representative data set
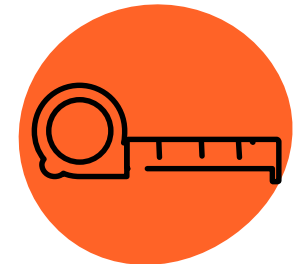
Want: Define the size and automatically create a representative data subset for the chosen objects.

Consistency Issues

- Using PID in key necessitates masking/encrypting data from multiple systems in a single session

- Fresh transaction data from one system doesn't match master data from another systems older test data set

- One project run a test that modifies test data also used by another project
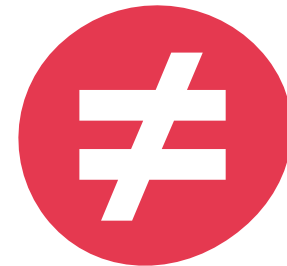
Want: Define dependencies between separate datasets derived from multiple systems.

Agility Issues

- With continous development and integration (CD/CI)
  - You need automatic (and fast) provisioning of the necessary test data set
  - With complete and consistent content
  - After completed test you want to get rid of test data, but keep a log of what was in the test data set

Want: "Fingerprint" logging when generating test datasets.

Test Data
Management Maturity

Test Data Management Maturity

A commonly acknowledged Test Data Management Maturity Model does not exist – yet!

There is a general "Data Management Maturity Model", a "Data Governance Maturity Assessment", a "Data Quality Management Maturity Model" and a "Master Data Management Maturity Model".

- These would be used to guide the development of a Test Data Management Maturity Model

A Test Data Management Maturity Model would probably describe levels of maturity along the six dimensions covered.

A Test Data Management Maturity Assessment would measure your maturity within the six dimensions.

Test Data Management Maturity Assessment Tool

SCOPING

1   **Determine the Scope of the Assessment**

1 a   In this step, a Target Maturity Level should be added on both the Short Term and the Longer Term and the scope for the remainder of the assessment should be confirmed.

> **GO TO the matrix with the suggested Assessment Scope**

2   **Determine the Target Maturity Levels**

2 a   Review the graphical representation of the Agreed Target Maturity Levels for both the Short Term and the Longer Term:

> **GO TO the Target Process Maturity Levels**

3   **Perform the Assessment**

3 a

Fill in the Maturity Assessment for the area in scope.
Start from the Assessment Overview Page and click on the process to be assessed to navigate to its detailed Maturity Assessment sheet.

Some advice on the assessment components:
For every maturity level (0 to 5);
- The CMM statements are split into single sentences.

Sheet tabs: START | Assessment Overview | Target Maturity | Tm | Sc | Pv | Sz | Cn | Ag | Current vs. Target ...

Maturity Assessment Overview

Main Menu

CLICK on the process name to navigate to the its detailed Maturity Assessment-sheet:
INSTRUCTIONS for individual sheets: A relative Weight between 0 and 10 should be allocated for each statement, and then an 'x' is used to indicate which statement is applicable.

| | | Assessment Details | | Maturity per Area | | | | |
| | | | | Target on the... | | Assessment | Gap on the... | |
		In Scope	Status	Short Term	Longer Term	Current	Short Term	Longer Term
Tm	Time	Yes	Open	2	3	2	0	1
Sc	Security	Yes	Open	3	4	1	2	3
Pv	Provisioning	Yes	Open	2	3	2	0	1
Sz	Size	Yes	Open	2	3	1	1	2
Cn	Consistency	Yes	Open	3	4	2	1	2
Ag	Agile	Yes	Open	2	3	2	0	1

Sheets: START | Assessment Overview | Target Maturity | Tm | Sc | Pv | Sz | Cn | Ag | Current vs. Target Maturity | Current Maturity

File | Home | Insert | Page Layout | Formulas | Data | Review | View | Help | Power-user | Tell me what you want to do | Share | Comments

J50

| | Area | Sc | Security | | | | | | | | Assessment Status | Open | | | Sc | | Security | |

Security refers to the policies, practices, and control rules used in the organization to mitigate risks and protect all data assets. By all data, we mean data from all sources (e.g., transactional, social media, machinegenerated data, IoT) and in all content formats (structured, semi-structured and unstructured).

Back to Assessment Overview

Level	Compliance	Contribution	Value
0	0,33	0,00	0,00
1	1,00	0,33	0,33
2	0,88	0,67	0,59
3	0,11	1,00	0,11
4	0,00	1,33	0,00
5	0,00	1,67	0,00

Maturity Level	0	Non-existent

| | | | Maturity Level = | 1,03 |

Nr	Statement	Weight	Not at all	A little	To some degree	Completely	Relative Importance
				Do you agree...			
1	There is no data governance security and privacy policy for the data in structured and semi-structured formats.	5		x			1,65
2	There is no data governance security and privacy policy for the data in unstructured format.	5			x		3,30
3	The IT Security Director/Specialist is not involved in the Data Governance Board	5				x	5,00
4	Your organization is not subject to any confidentiality regulations.	5	x				0,00
5	Your organization very often fails confidentiality audits.	5	x				0,00
6	Sensitive data is never encrypted.	5		x			1,65
7	Unencrypted sensitive data is always used to develop or test systems.	5			x		3,30
8	Administrators, subcontractors or third parties always have access to unencrypted sensitive data.	5		x			1,65
9	Systems are not used to monitor access to sensitive data by users with privileges.	5		x			1,65

Instructions: A relative Weight between 0 and 10 should be allocated for each statement, and then an 'x' is used to indicate which statement is applicable.
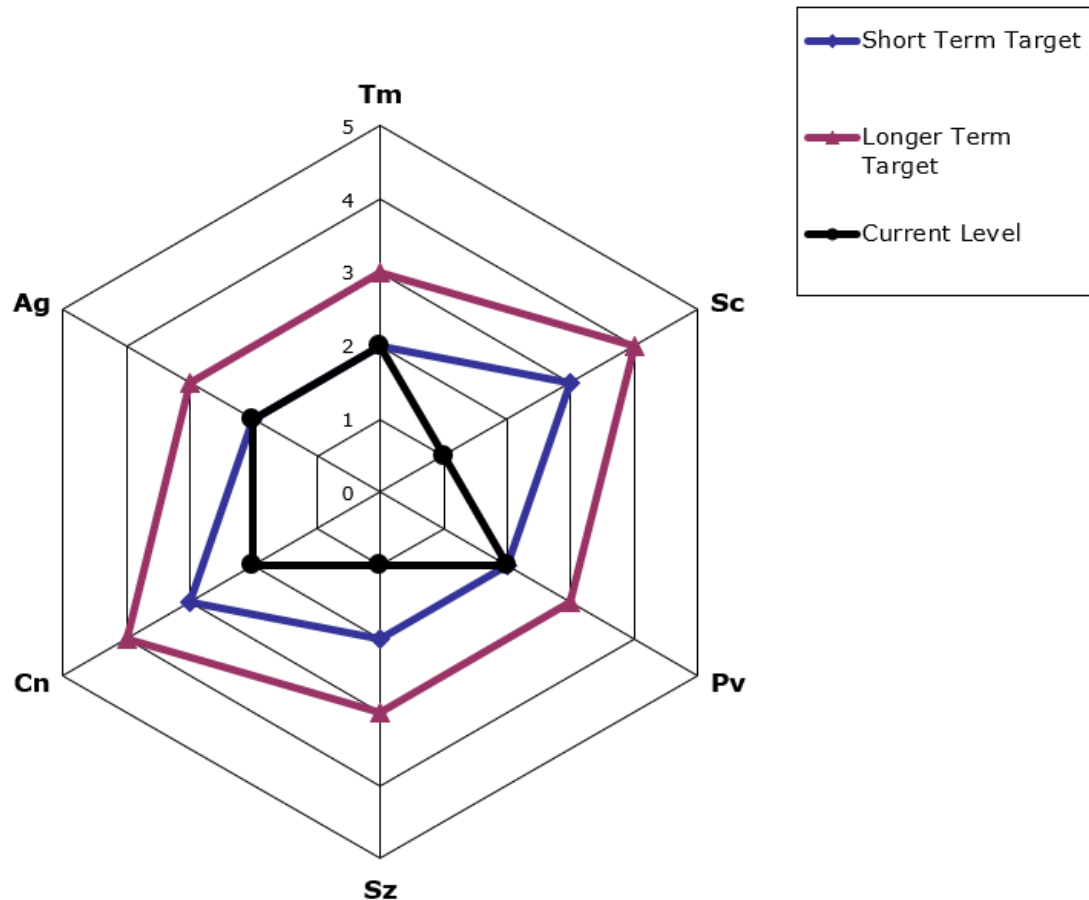
Total Weight	45

Maturity Level	1	Initial/Ad Hoc

Nr	Statement	Weight		Do you agree...

START | Assessment Overview | Target Maturity | Tm | Sc | Pv | Sz | Cn | Ag | Current vs. Target Maturity | Current Maturity

Current vs. Short and Long Term Target Maturity Level

Legend:
- Short Term Target
- Longer Term Target
- Current Level

Radar chart axes: Tm, Sc, Pv, Sz, Cn, Ag

Summary

- Multiple trends are pushing Test Data Management

- We have to change and adapt the way we work with test data

- A Test Data Management Maturity Model can guide improvements

sogetilabs
Part of Capgemini

Erik Haahr
Managing Consultant

erik.Haahr@capgemini.com
Mob.: +45 52 18 93 64
www.twitter.com/ErikHaahr
www.linkedin.com/in/erikhaahr/

Capgemini Danmark A/S

Delta Park 40
DK-2665 Vallensbæk Strand
Tel.: +45 70 11 22 00
www.capgemini.com

About Capgemini

A global leader in consulting, technology services and digital transformation, Capgemini is at the forefront of innovation to address the entire breadth of clients' opportunities in the evolving world of cloud, digital and platforms. Building on its strong 50-year heritage and deep industry-specific expertise, Capgemini enables organizations to realize their business ambitions through an array of services from strategy to operations. Capgemini is driven by the conviction that the business value of technology comes from and through people. It is a multicultural company of 211,300 team members in over 40 countries. The Group reported 2018 global revenues of EUR 13.2 billion.

Learn more about us at

www.capgemini.com

People matter, results count.