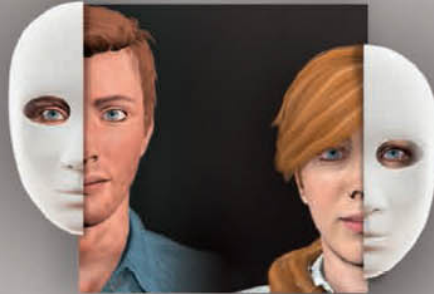



Klaus Franz · Tanja Tremmel · Eckehard Kruse

# Basiswissen



# Testdaten- management

Aus- und Weiterbildung zum  
Test Data Specialist  
Certified Tester Foundation Level  
nach GTB

 dpunkt.verlag

# Inhalt

**Cover**

**Über den Autor**

**Titel**

**Impressum**

**Geleitwort**

**Vorwort**

**Danksagung**

**Inhaltsverzeichnis**

## **1 Einleitung**

- 1.1 Was dieses Buch bewirken soll
- 1.2 Wie das Buch aufgebaut ist
- 1.3 Wie das Buch gelesen werden kann
- 1.4 Wie das Buch zum Lehrplan passt

## **2 Herausforderungen, Risiken, Nutzen**

- 2.1 Herausforderungen für das Testdatenmanagement
  - 2.1.1 Zielgerechte Testdatenbereitstellung
  - 2.1.2 Rechtzeitige Planung und Beauftragung der Testdatenbereitstellung
  - 2.1.3 Verbindung zwischen Testinfrastruktur und Testdatenanforderung
  - 2.1.4 Verknüpfung von Testfall und Testdaten
  - 2.1.5 Auflösen von Wissensmonopolen
  - 2.1.6 Wiederverwendbarkeit von Testdaten
  - 2.1.7 Effiziente Verwaltung und Pflege von Testdaten
  - 2.1.8 Beachtung regulatorischer Vorgaben
  - 2.1.9 Einbindung in den Testprozess
  - 2.1.10 Umgang mit großen und komplexen Datenmengen

- 2.1.11 Exklusivität der Daten und Mehrfachnutzung der Testinfrastruktur
- 2.1.12 Festlegung der Verantwortlichkeiten
- 2.2 Nutzen und Kosten des Testdatenmanagements
  - 2.2.1 Nutzen des Testdatenmanagements
    - 2.2.1.1 Wiederverwendbarkeit und Verfügbarkeit
    - 2.2.1.2 (Datenschutz-)Rechtliche Absicherung
  - 2.2.2 Kosten des Testdatenmanagements
- 2.3 Risiken durch unzureichendes Testdatenmanagement
  - 2.3.1 Risiko Datenschutz
  - 2.3.2 Risiko Datensicherheit
  - 2.3.3 Risiko Projekterfolg
  - 2.3.4 Risiko Testqualität
- 2.4 Ziele des Testdatenmanagements
- 2.5 Kontrollfragen zu Herausforderungen, Risiken, Nutzen

### **3 Regulatorischer Hintergrund**

- 3.1 Gesetzliche Regeln
  - 3.1.1 Gesetze und Vorschriften
    - 3.1.1.1 Bundesdatenschutzgesetz
    - 3.1.1.2 Datenschutz-Grundverordnung
    - 3.1.1.3 Weitere deutsche Gesetze
  - 3.1.2 Aufbewahrungsfristen
  - 3.1.3 Datenschutz im Testdatenmanagement
  - 3.1.4 Datensicherheit im Testdatenmanagement
  - 3.1.5 Auditierbarkeit des Testdatenmanagements
- 3.2 Branchenspezifische Regeln (Compliance)
- 3.3 Unternehmensinterne Regeln (IT-Governance)
- 3.4 Kontrollfragen zum regulatorischen Hintergrund

### **4 Grundlagen**

4.1 Definitionen zu Testfällen, Testdaten und Testdatenmanagement

4.2 Testdatenarten

4.2.1 Klassifizierung nach Verwendungsart

4.2.2 Klassifizierung nach Erzeugungsart

4.3 Kontrollfragen zu den Grundlagen

## **5 Verfahren zur Erzeugung von Testdaten**

5.1 Erzeugung von Produktivdaten

5.2 Erzeugung synthetischer Daten

5.3 Erzeugung pseudonymisierter Daten

5.4 Erzeugung anonymisierter Daten

5.4.1 Anonymisierungsverfahren

5.4.1.1 Informationsreduktion

5.4.1.2 Informationsveränderung

5.4.2 Aspekte der Datenmaskierung

5.4.2.1 Einfluss der Fachlichkeit

5.4.2.2 Statische und dynamische Datenmaskierung

5.4.2.3 Maskierung unstrukturierter Dokumente

5.4.2.4 Werkzeugunterstützung

5.5 Testdaten im konkreten Projekt

5.6 Kontrollfragen zu den Verfahren der Testdatenerzeugung

## **6 Der Testdatenmanagementprozess im Überblick**

6.1 Testdatenmanagementprozess vs. Testprozess

6.2 Kernelemente des Testdatenmanagementprozesses

6.3 Kontrollfragen zum Testdatenmanagementprozess

## **7 Aktivitäten im Testdatenmanagementprozess**

7.1 Testdatenrichtlinie erstellen

7.2 Testdatenmanagementprozess planen und steuern

7.2.1 Testdatenmanagementprozess planen

7.2.2 Testdatenmanagementprozess steuern

7.3 Testdaten spezifizieren

7.3.1 Inhaltliche Anforderungen an die Testdaten analysieren und spezifizieren

7.3.2 Organisatorische Verfahren festlegen und planen

7.3.3 Einhaltung der regulatorischen Vorgaben sicherstellen

7.4 Testdatenerzeugung konzipieren

7.5 Testdaten bereitstellen

7.6 Testdatenabschlussbericht erstellen

7.7 Testdaten archivieren

7.8 Kontrollfragen zu den Aktivitäten

## **8 Ergebnisartefakte im Testdatenprozess**

8.1 Testdatenrichtlinie

8.1.1 Testdatenmanagement in unterschiedlichen Projektvorgehensmodellen

8.1.2 Schnittstellen zu anderen Prozessen

8.1.3 Metriken des Testdatenmanagements

8.1.4 Die erste Version einer Testdatenrichtlinie

8.2 Testdatenkonzept

8.3 Testdatenspezifikation

8.3.1 Fachliche Aspekte

8.3.2 Rechtliche Aspekte

8.3.3 Organisatorische Aspekte

8.4 Testdatenbereitstellungsroutinen

8.5 Testdatenbereitstellungsplan

8.6 Testdatenbereitstellungsbericht

8.7 Testdatenbewertungsbericht

8.8 Testdatenabschlussbericht

8.9 Kontrollfragen zu den Ergebnisartefakten

## **9 Rollen im Testdatenmanagementprozess**

9.1 Testdatenmanager

9.2 Testdatenmodellierer

9.3 Testdatenrealisierer

9.4 Kontrollfragen zu den Rollen

## **10 Testdatenwerkzeuge**

10.1 Klassifizierung nach Geschäftsmodell

10.2 Klassifikation nach Verwendung

10.3 Eigenschaften von Testdatenwerkzeugen

10.4 Welche Testdatenwerkzeuge gibt es?

10.5 Kontrollfragen zu Testdatenwerkzeugen

## **11 Einbindung in die Organisation**

11.1 Sponsoren finden

11.2 Organisationseinheiten integrieren

11.3 Verantwortlichkeiten festlegen

11.4 Einbindung in die Prozesse

11.4.1 Einbindung in das Testumgebungsmanagement

11.4.2 Einbindung ins Konfigurationsmanagement

11.4.3 Einbindung in die IT

11.5 Einführung des Testdatenmanagements in die Organisation

11.6 Kontrollfragen zur Einbindung in die Organisation

## **12 Beispiele für projektbezogene Testdatenartefakte**

12.1 Das Beispielprojekt RiCoSys

12.2 Testdatenkonzept RiCoSys

12.3 Beispiele für Testdatenspezifikation

12.4 Beispiel für Dokumentation der Testdatenbereitstellungsroutine

12.5 Beispiel für Testdatenabschlussbericht

12.6 Abschlussbetrachtungen zu den Testdatenartefakten

### **13 Beispiel für eine Testdatenrichtlinie**

13.1 Datenschutzbestimmungen für Produktionsdaten

13.2 Datenschutzbestimmungen für Testdaten

13.3 Abschlussbetrachtungen zur Testdatenrichtlinie

### **Anhang**

A Lösungshinweise

B Glossar

C Synonyme

D Abkürzungen

E Quellen

E.1 Literatur

E.2 URLs

### **Index**

## 4 Grundlagen

### *Es war einmal ...*

*ein Tester, der sich um die Beschaffung der Testdaten in einem Projekt kümmerte. Er wollte alle Entwickler, Testautomatisierer und Fachbereichstester glücklich machen und mit Testdaten versorgen. Das tat er auch. Er stellte jedem Team sehr schnell eine komplette Kopie der Produktionsdaten zur Verfügung.*

*Gut gemeint ist nicht immer gut gemacht. Jeder der fünf Programmierer musste einen riesigen Datenbestand in seiner Entwicklungsumgebung verwalten. Die Testautomatisierung war ausgelagert, was zur Folge hatte, dass die externen Kollegen sensible Daten einsehen konnten. Die Fachbereichstester fanden in dem – für ihre Zwecke viel zu großen und unübersichtlichen – Datenbestand nicht die Datenkonstellationen, die sie für ihre Zwecke brauchten.*

*Es haperte an den Grundlagen des Testdatenmanagements!*

*Die bereitgestellten Testdaten waren nicht zielgruppenorientiert, nicht auf die definierten Testfälle zugeschnitten, nicht datenschutzkonform und sie verschwendeten Ressourcen.*

In diesem Kapitel werden die Grundlagen für ein systematisches Testdatenmanagement gelegt. Dazu gehören die notwendigen Definitionen und die Klassifizierungen von Testdaten.

### 4.1 Definitionen zu Testfällen, Testdaten und Testdatenmanagement

Wer testen will, braucht Testfälle. Im Glossar des ISTQB® [URL: ISTQB-Glossar] wird ein Testfall so definiert:

#### **Testfall**

Ein Testfall umfasst folgende Angaben: die für die Ausführung notwendigen Vorbedingungen, die Menge der Eingabewerte (ein Eingabewert je Parameter des Testobjekts), die Menge der vorausgesagten Ergebnisse, sowie die erwarteten Nachbedingungen.

Testfälle werden entwickelt im Hinblick auf ein bestimmtes Ziel bzw. auf eine Testbedingung, wie z.B. einen bestimmten Programmpfad auszuführen oder die Übereinstimmung mit spezifischen Anforderungen zu prüfen (wie Eingaben an das Testobjekt zu übergeben und Sollwerte abzulesen sind).

Nun sollte schon in frühen Projektphasen begonnen werden, Testfälle zu beschreiben, auch wenn noch nicht alle konkreten Informationen, die für ihre

Ausführung benötigt werden, vorliegen. Daher werden Testfälle entsprechend ISTQB® [URL: ISTQB-Glossar] in abstrakte und konkrete Testfälle unterschieden:

#### **Abstrakter Testfall**

Ein Testfall ohne konkrete Ein- und Ausgabewerte für Eingabedaten und vorausgesagte Ergebnisse. Er verwendet logische Operatoren, weil die konkreten noch nicht definiert oder verfügbar sind.

#### **Konkreter Testfall**

Ein Testfall mit konkreten Werten für Eingaben und vorausgesagte Ergebnisse. Logische Operanden der abstrakten Testfälle werden durch konkrete Werte ersetzt.

Für ein System unter Test werden im Rahmen der Testspezifikation in der Regel zuerst abstrakte Testfälle beschrieben. Sie geben Struktur, Beschaffenheit und Umfang der erforderlichen Testdaten vor.

Spätestens zur Durchführung der Tests werden dann konkrete Testfälle benötigt. Diese benennen im Gegensatz zu den abstrakten Testfällen ganz konkret die Testdaten, die für die Testdurchführung erforderlich sind.

#### **Testdaten**

Testdaten sind Daten, die (z.B. in einer Datenbank) vor der Ausführung eines Tests existieren und die die Ausführung der Komponente bzw. des Systems im Test beeinflussen bzw. dadurch beeinflusst werden. [URL: ISTQB-Glossar]

Testdaten sind konkrete Werte von Datenobjekten, die zur Ausführung eines Testobjekts benötigt werden. Sie können z.B. in Dateiform, in Form von Datenbanktabellen oder in Listenform vorliegen. Unter Testdaten ist ein Vorrat von Werten zu verstehen, die physikalisch vorliegen oder in anderer Weise mit konkreten Testfällen in Beziehung stehen können.

Aus abstrakten Testfällen werden konkrete Testfälle mit Testdaten.

In einem abstrakten Testfall werden die Eingabewerte in der Regel durch Eingabeparameter ersetzt. Wenn z.B. ein Testfall verlangt, dass ein gültiges Datum erfasst werden soll, dann wird der abstrakte Eingabeparameter »Gültiges Datum« im konkreten Testfall durch das konkrete Testdatum »17.04.2017« ersetzt.

Ausnahmen bestätigen die Regel. In einigen Situationen – wie beim Testen von Grenzwerten – kann es effektiv sein, bei der Spezifikation von Testfällen sofort konkrete Testfälle zu beschreiben.

Beispielsweise wird man keinen abstrakten Testfall »Erfasse im Feld <Monat> den größten erlaubten Wert« beschreiben, sondern direkt den konkreten »Erfasse den Wert 12 im Feld <Monat>. Der größte erlaubte Wert eines Monats ist nun mal immer 12 und wird sich auch nicht so schnell ändern.

Testdaten können auch als dynamische Objekte vorliegen, z.B. in einer definierten Einheit aus Datenquelle, Datenstruktur und Regelwerk zur Erzeugung. Das Regelwerk selbst ist dann Bestandteil der Testdaten.

*Testdaten als dynamische Objekte*

So könnte der abstrakte Eingabeparameter »vorgestern« im konkreten Fall durch das Regelwerk »heute() - 2« bestimmt werden.

Die Struktur eines Datensatzes hängt von den Datentypen und Inhalten der einzelnen Datenfelder ab.

*Strukturierte Testdaten*

Die konkrete Ausprägung eines einzelnen Datensatzes wird durch die Gesamtheit der Werte der einzelnen Datenfelder repräsentiert. Durch Beziehungen unterschiedlicher Datensätze zueinander können komplexe Datenmodelle entstehen, d.h., Testdaten können auch als strukturierte Daten (z.B. Datenbanken) vorliegen.

Andererseits können Testdaten auch als unstrukturierte Daten benötigt werden, wie z.B. Texte, (Röntgen-)Bilder oder PDF-Dateien.

*Unstrukturierte Testdaten*

Die Quantität der bereitgestellten Testdaten spielt eine wichtige Rolle für einen effizienten Testprozess.

*Testdatenquantität*

Zu wenige Testdaten – bezogen auf eine Menge von Testfällen – bereitzustellen, bedeutet, dass nicht alle Tests durchgeführt werden können. Zu viele Testdaten zu haben, bedeutet, dass viele Testdaten nicht benötigt werden, die aber Aufwand bei der Erstellung und Verwaltung erzeugen und unnötigen Speicherplatz belegen.

Der Testdatenkoeffizient ist ein Maß für die Nutzung von Testdaten:

#### **Testdatenkoeffizient**

Der Testdatenkoeffizient ist das Verhältnis aus vorhandenen Testdaten zu genutzten Testdaten bezogen auf die Testfälle zu einem Testobjekt oder einem System unter Test.

Der Testdatenkoeffizient zeigt auf, wie sparsam man mit Testdaten umgeht. Wenn z.B. 100 GB Testdaten vorhanden sind, aber nur 15 GB davon genutzt werden, ist der Testdatenkoeffizient mit 15% schlecht.

*Testdatensparsamkeit*

Gute Datensparsamkeit beim Testen bedeutet, dass der Testdatenkoeffizient gegen 100% konvergiert.

Regel: Nur die Testdaten bereitstellen, die wirklich benötigt werden.

Entscheidend für die effiziente Nutzung der Testdaten – und somit für den Erfolg der Tests – ist die Qualität der Testdaten.

*Testdatenqualität*

Testdaten werden immer für einen konkreten Testfall mit einem konkreten Testziel spezifiziert und bereitgestellt. Sie müssen neben fachlichen

auch regulatorischen Anforderungen entsprechen, insbesondere denen des Datenschutzes.

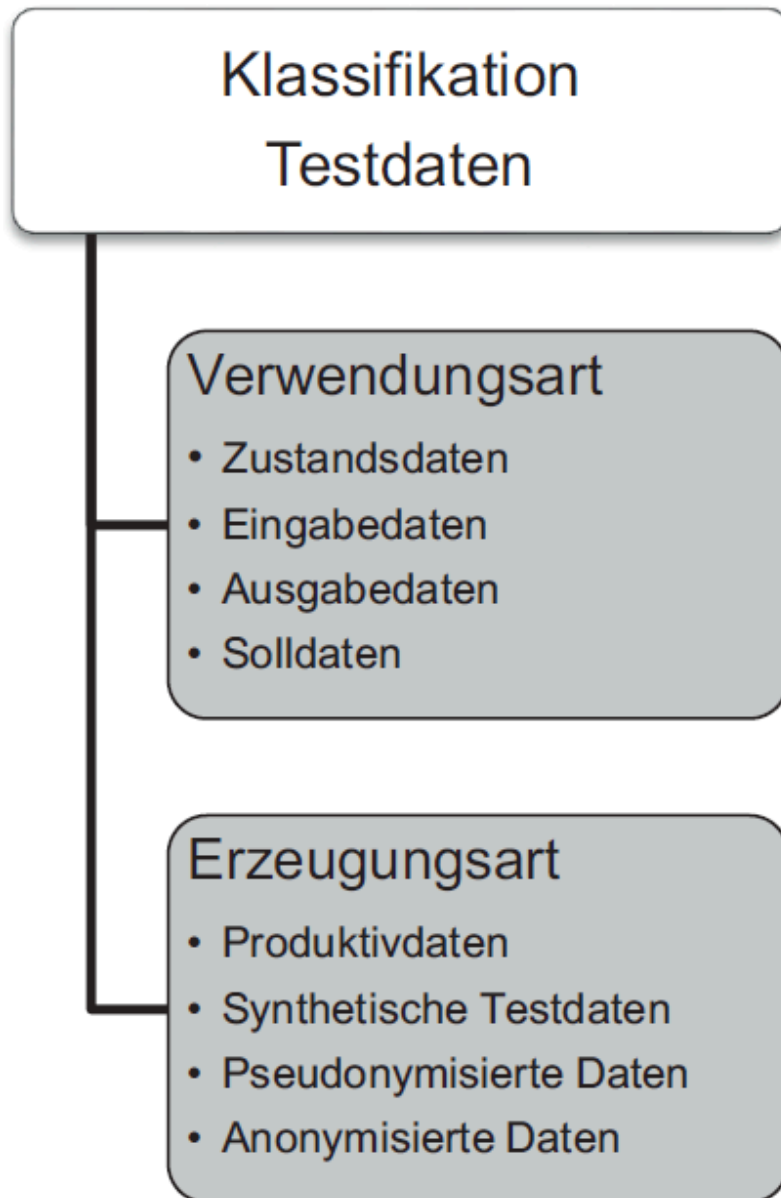
Alle bis hier betrachteten Aspekte bringen uns zu der Definition von Testdatenmanagement:

#### **Testdatenmanagement**

Testdatenmanagement ist die Menge aller methodischen, konzeptionellen, organisatorischen und technischen Maßnahmen und Verfahren zur Behandlung der Ressource Testdaten mit dem Ziel, die Testdaten mit ihrem maximalen Nutzungspotenzial unter Berücksichtigung von Datenschutzerfordernissen in die Testprozesse einzubringen und die effiziente Nutzung der Testdaten zu gewährleisten.

## **4.2 Testdatenarten**

Testdaten müssen spezifiziert werden, damit sie effizient bereitgestellt werden können. Dabei kann eine Klassifizierung der in einem Projekt angeforderten Testdaten hilfreich sein (siehe Abb. 4-1). Die Klassifizierungen nach Verwendungsart – dem Was? – und Erzeugungsart – dem Wie? – werden in den folgenden Abschnitten vorgestellt.



**Abb. 4-1** Klassifikation von Testdaten

#### 4.2.1 Klassifizierung nach Verwendungsart

Personen, die Testdaten benötigen oder bereitstellen müssen, stellen sich (und anderen) folgende fachliche Fragen:

*Fachliche Sicht auf Testdaten*

- Welche Testdaten werden für den Test benötigt?
- Welche dieser Daten werden beim Testen verändert oder »verbraucht«?
- Welche Daten werden beim Testen erzeugt?
- Womit soll das Testergebnis verglichen werden?

Die Klassifizierung nach der Verwendungsart in Zustands-, Eingabe-, Ausgabe- und Solldaten hilft bei der Beantwortung dieser Fragen. Sie stellt eine fachliche Sicht auf Testdaten dar.

#### **Zustandsdaten**

Zustandsdaten sind Daten, die in Datenbanken, Dateien oder Umsystemen vorhanden sind und von Testfällen als wiederverwendbarer Startzustand des Testobjekts vorausgesetzt und benötigt werden.

#### **Eingabedaten**

Eingabedaten sind Daten, die bei der Ausführung eines Testfalls an ein Testobjekt übergeben (eingegeben, eingelesen oder übertragen) werden.

#### **Ausgabedaten**

Ausgabedaten sind Daten, die nach Ausführung eines Testfalls von einem Testobjekt zurückgegeben (ausgegeben, gedruckt, gesendet oder übertragen) werden.

#### **Solldaten**

Solldaten sind Daten, die vor Ausführung eines Testfalls als erwartetes Ergebnis (Testorakel<sup>1</sup>) bereitgestellt werden.

Als Beispiel für die Klassifizierung nach Verwendungsart werden Testdaten betrachtet, die für den Funktionstest der Bestellabwicklung eines neuen Onlineshops für Buntstifte benötigt werden.

#### **Beispiel für Klassifizierung nach Verwendungsart**

Der Onlineshop wurde mit einer neuen Schnittstelle an ein bestehendes Bestellsystem angebunden. Testziel ist die korrekte Abwicklung einer Bestellung mit einem oder mehreren Artikeln.

Die Kundendaten, die für die Bestellung ausgewählten Artikel und die Zahlungsangaben sind **Eingabedaten**.

Damit Artikel ausgewählt werden können, muss ein Artikelbestand im Testsystem existieren. Mit diesen **Zustandsdaten**, die beim Test unverändert bleiben, sind die Tests zu den Bestellvorgängen wiederholbar und vergleichbar.

Die in der Datenbank gespeicherten Bestellungen und die an Kunden versendeten und archivierten Bestätigungs-Mails sind **Ausgabedaten**.

Bestellungen können seit vielen Jahren per Post via Bestellbogen oder per Telefon erfolgen. Diese Daten werden von Mitarbeitern des Onlineshops für Buntstifte im Bestellsystem erfasst. Da das Bestellmedium keinen Einfluss auf den Einkauf haben darf, können die auf dem »alten« Weg erzeugten

Bestellungen als **Solldaten** für den Funktionstest des Onlineshops genutzt werden.

Man sieht an diesem Beispiel, dass das Testziel maßgeblich die Verwendungsart von Testdaten beeinflusst. Denn im obigen Beispiel des Onlineshops werden die Datensätze des Artikelbestandes im Test als Zustandsdaten benötigt. Wenn aber z.B. die Verwaltungsfunktionen des Artikelkatalogs getestet werden, sind die Artikeldatensätze die Eingabedaten. Oder für den Test des Abrechnungssystems werden die Bestelldaten als Zustandsdaten benötigt; sie sind in diesem Fall keine Eingabedaten wie im Test des Onlineshops.

Der Verwendungsart von Testdaten wird durch das Testziel bestimmt.

Durch die Klassifikation der benötigten Testdaten nach ihrem Verwendungszweck kann leichter erkannt werden, ob ihre Bereitstellung ggf. optimiert werden kann.

Sollten in unserem Beispiel sowohl der Onlineshop als auch die Verwaltungsfunktionen des Artikelkatalogs neu entwickelt werden, wäre es sinnvoll, zuerst die Verwaltungsfunktionen des Artikelkatalogs zu testen, um mit den erzeugten Testdaten die Onlinebestellungen testen zu können. Das heißt, ein System beliefert das andere mit Testdaten – womit ein eleganter Übergang zum nächsten Abschnitt gelungen ist, in dem die Erzeugungsarten von Testdaten betrachtet werden.

#### 4.2.2 Klassifizierung nach Erzeugungsart

Bei der Unterscheidung von Testdaten nach Erzeugungsart werden die Testdaten aus technischer Sicht betrachtet.

Für die physikalische Bereitstellung von Testdaten müssen Fragen beantwortet werden, wie die angeforderten Testdaten am besten erzeugt werden können:

- Sollen, können, dürfen die Daten aus der Produktionsumgebung kopiert werden?
- Müssen sie dabei verändert werden? Wenn ja, wie und in welcher Intensität?
- Müssen oder sollen die Testdaten synthetisch erzeugt werden?

Die Beantwortung dieser Fragen liefert die Klassifikation von Testdaten nach ihrer Erzeugungsart. Damit hat man eine technische Sicht auf die bereitzustellenden Testdaten und kann sie in Produktivdaten, synthetische, pseudonymisierte und anonymisierte Daten unterteilen.

**Produktivdaten**

Produktivdaten<sup>2</sup> sind reale, ggf. schützenswerte und unveränderte Daten, die aus dem produktiven Einsatz eines Systems stammen oder in produktiven Systemen verwendet werden.

Wenn sich in einem Produktionsabzug personenbezogene Daten befinden, ist dieser zur direkten Nutzung im Test aus Datenschutzaspekten problematisch, aber nicht zwangsweise unzulässig. Allerdings müssen die Zugriffe auf die Produktivdaten streng limitiert und die Verwendung protokolliert werden; Datenschutzvorgaben müssen eingehalten werden (siehe Kap. 3).

### **Synthetische Testdaten**

Synthetische Testdaten sind Daten, die künstlich erzeugt werden und keinen Bezug zur Realität haben. Die Datenquelle darf keinen natürlichen Ursprung haben.

Synthetische Testdaten sind realistisch, haben aber keinen inhaltlichen Bezug zu Produktivdaten, d.h., sie sind nicht real. In Abhängigkeit von Teststufe, Testziel und Testumgebung müssen Varianten und mengenmäßige Verteilung der realen Daten innerhalb des Testsystems abgebildet werden.

Die Erzeugung synthetischer Testdaten bedarf tiefer Kenntnisse des jeweiligen Datenmodells und der sich aus den Geschäftsprozessen ergebenden Datenvarianz. Synthetische Testdaten werden in der Regel mit geeigneten Werkzeugen generiert.

### **Pseudonymisierte Daten**

Pseudonymisierte Daten sind Daten, die durch den Prozess der Pseudonymisierung aus Produktivdaten gewonnen werden.

### **Pseudonymisierung**

Pseudonymisierung ist die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden. (DSGVO Artikel 4, siehe [URL: DSGVO])

Der Vollständigkeit halber ist hier auch die weniger ausführliche Definition von Pseudonymisierung nach dem BDSG, §3 Abs. 6a aufgeführt:

### **Pseudonymisierung nach BDSG**

Pseudonymisieren ist das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren.

Die Datenquelle für pseudonymisierte Daten ist natürlichen Ursprungs und es existiert mindestens eine Zuordnungsregel, die einem Originalwert ein Pseudonym zuordnet. Durch diese Zuordnungsregel, mit der auf die Originaldaten rückgeschlossen werden kann, sind die Daten nicht anonym.

---

Ein beliebtes Beispiel für Pseudonymisierung ist die Bekanntgabe von Prüfungsergebnissen. Ein Prüfungsamt will die Ergebnisse einer schriftlichen Prüfung den Teilnehmern zugänglich machen.

**Beispiel für Pseudonymisierung**

Daher schreiben diese während der Prüfung ein selbstgewähltes Pseudonym auf den Prüfungsbogen. Die Prüfungsergebnisse werden vom Prüfungsamt komplett auf einem Aushang mit den Spalten <Pseudonym> und <Note> veröffentlicht. Da das Prüfungsamt die Zuordnung des Pseudonyms zum jeweiligen Teilnehmer herstellen kann, ist der Aushang pseudonymisiert, aber nicht anonymisiert.

---

Eine Verschärfung der Pseudonymisierung stellt die Anonymisierung von Daten dar. Interessanterweise spielt der Begriff Anonymisierung in der neuen DSGVO keine Rolle! Für die Bereitstellung von Testdaten ist deren Anonymisierung aber sehr wichtig.

**Anonymisierte Daten**

Anonymisierte Testdaten sind Daten, die durch verschiedene Verfahren zur Anonymisierung entstehen, bei denen Produktivdaten manipuliert werden, bis der gewünschte Anonymisierungsgrad erreicht ist.

**Anonymisierung**

Anonymisieren ist das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können. (BDSG §3 Abs. 6, siehe [URL: BDSG])

**Anonymisierungsgrad**

Die Wahrscheinlichkeit, dass ein vermeintlicher Angreifer die produktive Information aufgrund der ihm zugänglichen anonymisierten Informationen errät.

In der Statistik werden drei Anonymisierungsgrade unterschieden (siehe Abb. 4–2 und [URL: Destatis-GI]):

*Anonymisierungsgrad*

**Formale Anonymität**

Ein Datensatz ist formal anonymisiert, wenn direkte Identifikatoren wie Name und Anschrift entfernt wurden, sodass die Datensätze den Merkmalsträgern nicht zugeordnet werden können. Eine Zuordnung ist nur noch möglich, wenn zufällig eindeutige Kombinationen von Merkmalsausprägungen gefunden werden.

**Faktische Anonymität**

Bei einem faktisch anonymisierten Datensatz kann eine Deanononymisierung der Daten nicht mit Sicherheit ausgeschlossen werden. Ein Datensatz gilt als faktisch anonym, sofern ein

Datenempfänger unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft aufbringen muss, um eine Zuordnung der Daten zu den Betroffenen zu erstellen. (BStatG §16, [URL: BStatG])

### **Absolute Anonymität**

Absolut anonymisierte Daten sind durch Anonymisierung oder die Entfernung einzelner Merkmale so weit verändert, dass eine Identifizierung der Auskunftgebenden nicht möglich ist. Auch mit beliebig vielen zusätzlichen Informationen ist keine eindeutige und fehlerfreie Zuordnung möglich.

Betrachten wir als Beispiel für Anonymisierung wieder das Prüfungsamt, das die Ergebnisse einer schriftlichen Prüfung den Teilnehmern zugänglich machen will.

### **Beispiel für Anonymisierung**

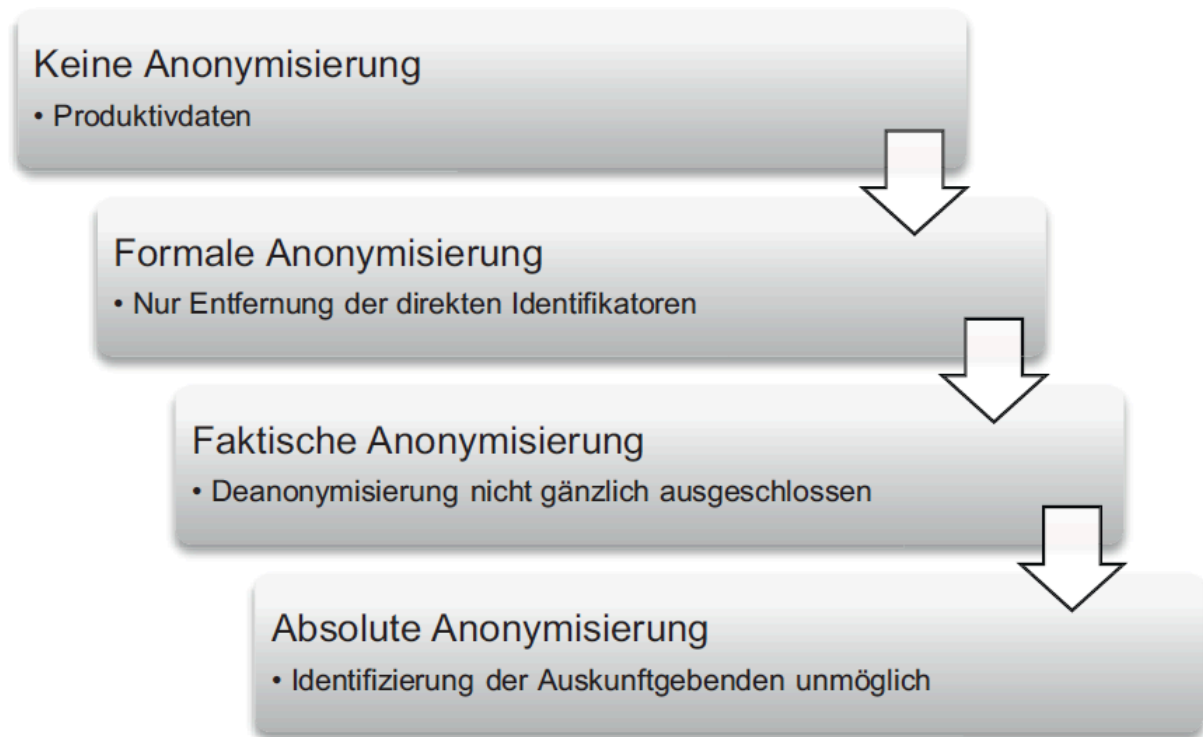
Jeder Prüfling schreibt während der Prüfung ein selbstgewähltes Pseudonym auf eine separate, dafür vorgesehene Seite des Prüfungsbogens.

Die Prüfungsergebnisse werden vom Prüfungsamt komplett auf einem Aushang mit den Spalten <Pseudonym> und <Note> veröffentlicht. Zudem werden die Seiten der Prüfungsbögen mit den Pseudonymen in einem Aktenvernichter der Sicherheitsstufe 4<sup>3</sup> vernichtet. Die Angaben auf dem Notenaushang sind für die Allgemeinheit absolut anonymisiert, da eine Zuordnung zu den jeweiligen Teilnehmern nicht mehr möglich ist. Auch nicht für das Prüfungsamt!

Das etwas flapsige Fazit:

Pseudonymisierte Daten werden anonym, wenn die Zuordnungsregel vernichtet wird.

Für die Ermittlung des Anonymisierungsgrads eines Datenbestandes müssen Informationen vorliegen, die ein vermeintlicher Angreifer nicht zwangsweise hat.



**Abb. 4-2** Anonymisierungsgrade

Die hier definierten Anonymisierungsgrade aus dem Bereich der Statistik können auch auf die Anonymisierung von Produktivdaten zu Testzwecken angewendet werden, wobei absolut anonyme Testdaten wohl nie garantiert werden können. IT-Testprojekte werden immer mit mehr oder weniger faktisch anonymisierten Daten arbeiten (müssen).

Das bedeutet, dass Anonymisierungsgrade auch individuell festgelegt werden können, wie in folgendem Beispiel:

*Beispiel individueller Anonymisierungsgrade*

1. Grad: Maskierung von Identifikatoren  
Namen, Adressen, Geburtsdaten
2. Grad: Maskierung von Schlüsseln  
Sichtbare Schlüssel wie Kostenstellen, Kundennummern, Mitarbeiternummern, Versicherungsnummern, Vertragsnummern
3. Grad: Maskierung von Attributen  
Beschreibende Attribute, die Rückschlüsse auf eine bestimmte Person oder schützenswerte Information ermöglichen
4. Grad: Maskierung von Kombinationen

Beschreibende Attribute, die in ihrer Kombination Rückschlüsse auf eine bestimmte Person oder schützenswerte Information ermöglichen

#### 5. Grad: Vollständige Anonymisierung

Keine Rückschlüsse auf eine bestimmte Person oder schützenswerte Information möglich

### **4.3 Kontrollfragen zu den Grundlagen**

*Mit den folgenden Fragen und Aufgaben kann der Leser überprüfen, ob er den gelesenen Stoff rekapitulieren kann.*

1. Wie sind Testfall, abstrakter und konkreter Testfall definiert?
2. Geben Sie Beispiele für abstrakte und konkrete Testfälle.
3. Geben Sie Beispiele, wie aus abstrakten Testfällen die Anforderungen für die benötigten Testdaten abgeleitet werden können.
4. Wie sind Testdaten und Testdatenmanagement definiert?
5. Mit welchem Maß kann die Datensparsamkeit gemessen werden? Geben Sie ein Beispiel.
6. Wie können Testdaten klassifiziert werden? Nennen Sie Beispiele für die einzelnen Testdatenklassen.
7. Was ist der Unterschied zwischen pseudonymisierten und anonymisierten Daten?
8. Wie sind die Prozesse Pseudonymisierung und Anonymisierung definiert?
9. Können Sie ein Beispiel geben, das den Unterschied zwischen pseudonymisierten und anonymisierten Daten verdeutlicht?
10. Welche Anonymisierungsgrade kennen Sie und wie unterscheiden sich diese?
11. Beschreiben Sie mehrere Projektszenarien, für die jeweils eine andere Erzeugungsart die beste Wahl ist, begründen Sie diese.
12. Im Beispiel in Abschnitt 4.2.1 liefert der Test der Verwaltungsfunktionen des Artikelkatalogs mit dem erzeugten Artikelbestand die Testdaten für den Test der Onlinebestellungen. Welche Erzeugungsart ist das?

# Index

## A

Absolute Anonymität 48

Abstrakter Testfall 40

Abweichungsmanagement 89

Aktivitäten 71

Anforderungsmanagement 88

Anforderungsquellen 95

Anforderungsüberdeckung 19

Anonymisierte Daten 47

Anonymisierung 47, 55

Anonymisierungsgrad 48, 79

Anonymisierungsverfahren 56

Anonymität

- absolute 48

- faktische 48

- formale 48

Archivierung der Testdaten 82

Auditierbarkeit 35

Auditierung 36

Audits 35

Aufbewahrungsfrist 32

Ausgabedaten 44

## **B**

BDSG 28

BDSG-neu 29

Bereitstellung der Testdaten 81

Bereitstellungsstrategie 96

Betroffene Person 29

Bundesdatenschutzgesetz 28

## **C**

CoC 36

Code of Conduct 36

Compliance 36

## **D**

Data Masking 55

Data Redaction 65

Data-Profiling-Werkzeuge 113

Daten

- anonymisierte 47

- personenbeziehbare 66

- personenbezogene 28, 66

- pseudonymisierte 46

- realistische 67

- unrealistische 67

Datenisolation 79, 97

Datenmaskierung 64, 117

- dynamische 64, 117

- statische 63, 117

Datenminimierung 30

Datenschiefstand 12

Datenschutz 20, 32  
Datenschutz-Anpassungs- und -Umsetzungsgesetz EU 29  
Datenschutzbestimmungen 150  
Datenschutz-Grundverordnung 29  
Datenschutzrichtlinien 22  
Datensicherheit 34  
Datensparsamkeit 16, 29, 67  
Datenvermeidung 29  
Deanonymisierung 55  
DSAnpUG-EU 29  
DSGVO 29  
Dynamic Data Masking 64  
Dynamische Datenmaskierung 64, 117

## **E**

Effektives Testdatenmanagement 123  
Eingabedaten 44  
Eins-zu-eins-Kopie 63  
Ergebnisartefakte 71  
Erzeugungsart 45  
ETL-Prozess 66  
ETL-Werkzeuge 114

## **F**

Faktische Anonymität 48  
Formale Anonymität 48  
Fundamentaler Testprozess 70

## **G**

Generalisierung 57  
Geodaten 31

Golden Copy 63

## **I**

Informationsreduktion 56

Informationsveränderung 56

Integrität 34

IT-Governance 37

IT-Sicherheitsrichtlinie 35

## **K**

Klassifizierung 43

Kompetenzstufen 107

Konfigurationsmanagement 88, 125

Konfigurationsmanagementwerkzeuge 116

Konkreter Testfall 40

Kontrollierbarkeit 35

Konzeption der Testdatenerzeugung 80

Kostenbasierte Metriken 90

Kriterien Datensicherheit 34

Kryptografie 60

## **L**

Lorem ipsum 58

Löschen 58

Löschfrist 83

## **M**

Marktortprinzip 30

Maskierung 57

Maskierungsregeln 61

Metadaten 53

Metriken 89

Modellbasiertes Testen 17

## **N**

Nachverfolgbarkeit 79, 91, 97

Nulling 58

Nummernkreise 97

## **O**

Organisatorische Einbindung 123

## **P**

Payment Card Industry Data Security Standard 36

PCI DSS 36

Personenbeziehbare Daten 66

Personenbezogene Daten 28–29, 66

Perturbation 58

Planung der Testdatenmanagementaktivitäten 75

Produktionsabzug 52

Produktivdaten 45, 52

Profiling 116

Pseudonymisieren 47

Pseudonymisierte Daten 46, 54

Pseudonymisierung 46, 54

## **R**

Realistische Daten 67

Reduktion 56

Regressionstest 19

Regulatorische Vorgaben 16, 79, 97

Releasemanagement 89

Reproduzierbarkeitsaufwand 90

Revisionssicherheit 20, 128

Risiko Datenschutz 22

Risiko Datensicherheit 23

Risiko Projekterfolg 23

Risiko Testqualität 24

Rollen 71

## **S**

Sarbanes-Oxley Act 37

Schlechte Testdaten 22

Shuffling 58

Solldaten 44

SOX 37

Spezifikation der Testdaten 76

SPICE 36

Static Data Masking 63

Statische Datenmaskierung 63, 117

Steuerung des Testdatenmanagementprozesses 76

Strukturierte Testdaten 41

Substitution 57

SWIFT 12

Synthetische Testdaten 46, 53

## **T**

TDaaS 113

Teildatenbezug 52

Test Data as a Service 113

Testausführungswerkzeuge 116

Testautomationsexperten 99

Testdaten 41

    schlechte 22

- strukturierte 41
- synthetische 46
- unstrukturierte 42
- Testdatenabschlussbericht 82, 103
- Testdatenanalysewerkzeuge 113
- Testdatenarchivierung 82
- Testdatenbereitstellung 116
- Testdatenbereitstellungsbericht 82, 102
- Testdatenbereitstellungsplan 81, 102
- Testdatenbereitstellungsroutine 81, 98
- Testdatenbestände 127
- Testdatenbewertungsbericht 82, 103
- Testdateneditoren und -generatoren 114
- Testdatenkoeffizient 42
- Testdatenkonzept 74, 91–92
- Testdatenmanagement 42, 122
  - ungeeignetes 22
- Testdatenmanagementprozess 70
- Testdatenmanager 106
- Testdatenmetamodell 54
- Testdatenmodellierer 108
- Testdatenqualität 42
- Testdatenquantität 42
- Testdatenrealisierer 109
- Testdatenrichtlinie 74, 86, 89, 127, 157
- Testdatensparsamkeit 42
- Testdatenspezifikation 77, 94–95
- Testdatenverteilung 90

Testdatenverwaltung 118  
Testdatenwerkzeuge 112  
Testentwurfswerkzeuge 113  
Testfall 40  
    abstrakt 40  
    konkret 40  
Testinfrastruktur 12  
Testkomparatoren 116  
Testmanagementsuiten 113  
Testprozess 69  
TestSPICE 36  
Testumgebungsmanagement 89, 124  
Testwerkzeuge 112  
Traceability 91

## **U**

Überdeckungsgrade 90  
Ungeeignetes Testdatenmanagement 22  
Unrealistische Daten 67  
Unstrukturierte Dokumente 65  
Unstrukturierte Testdaten 42  
Unternehmensinterne Regeln 37

## **V**

Varianzbildung 59  
Verantwortlichkeiten 18, 123  
Verfügbarkeit 19, 35  
Versionierung 126  
Vertraulichkeit 34  
Vertraulichkeitserklärung 33

Verwendungsart 43, 96

## **W**

Wiederholbarkeit 78, 96

Wiederverwendbarkeit 14, 19

Wissensmonopole 14

## **Z**

Zeitreisen 78

Ziele des Testdatenmanagements 25

Zielformate 79, 97

Zielgerichtete Testdatenbereitstellung 11

Zugriffsmanagement 80

ZÜRS Geo 118

Zustandsdaten 43